2009

MOLECULAR MODELING STUDIES OF HEPARIN AND HEPARIN MIMETICS BINDING TO COAGULATION PROTEINS

CHANDRAVEL KRISHNASAMY
Virginia Commonwealth University

MOLECULAR MODELING STUDIES OF HEPARIN AND HEPARIN MIMETICS

BINDING TO COAGULATION PROTEINS

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

by

CHANDRAVEL KRISHNASAMY
MS in Pharmacy, Shri G. S. Institute of Technology and Sciences, Indore, INDIA, 1999
BS in Pharmacy, The Tamil Nadu Dr. MGR Medical University, Chennai, INDIA, 1997

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

Virginia Commonwealth University
Richmond, Virginia
August 2009

Acknowledgement

I give all glory and honor to my God JESUS CHRIST in whom I trust who brought me so far. I can do everything through Christ who strengthens me. Philippians 4:13.

I would like to thank my advisor Umesh R. Desai for being a wonderful teacher. His continuous support not only helped me to achieve my academic goals, but also in my profession as a pharmacist. His willingness to accommodate the needs of individual students still amazes me. I am grateful to have someone like Dr. Desai who believed in my research potential and bringing that to fruition, although at times the path was difficult.

I would also like to thank my committee members Drs. Glen E. Kellogg, Yan Zhang, Thomas F. Huff and Phillip M. Gerk. This committee has greatly shaped my research work. The modeling course taught by Dr. Kellogg has helped me a lot in performing modeling experiments for my thesis. I thank Dr. Lemont B. Kier for teaching the QSAR course and helping me to write the paper on the application of molecular connectivity.

I certainly cannot ignore the steadfast love and compassion of Dr. Philip D. Mosier, to help every student in the department including me. At times, when I felt like giving up, he was always there like an angel to get me through and guide me.

I would like to thank Drs. Gunnarson, Monien and Liang of Dr. Desai's group postdoctoral scholars for their contribution in establishing much of the foundation for the projects that I continued.

I would also like to thank Dr. Arjun Raghuraman, a graduate from our group whose help is unforgettable. Jay N. Thakkar, a friend and colleague will be always remembered by my children Nikhil and Shiny as Jay Annan (Brother), who made my life in the beginning of my stay at the Berkshire apartment during my graduate program and in the lab much easier. I would like to thank the former graduate students of our group especially Juniad Afridi, Mohammed Rahman (Mo) and Brian Henry.

Finally, I would like to thank my parents, brother and sister. I am also thankful for my wife Lucksha, who sacrificed a great deal of her comfort and dreams for my sake. I am blessed to have my son Nikhil, a wonderful loving precious one, who stayed in India for three years without me when I came here to do my Ph.D. program. I am also blessed by my daughter Shiny, who makes my days enjoyable with lots of love and fun.

I would also like to thank many others, without their help and prayers it would not have been possible for me to finish this challenging task. I thank God for all those people who were and are part of my journey.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

MOLECULAR MODELING STUDIES OF HEPARIN AND HEPARIN MIMETICS BINDING TO COAGULATION PROTEINS

By Chandravel Krishnasamy, PhD

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

Virginia Commonwealth University, 2009

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

Heparin, a glycosaminoglycan (GAG), is a complex biopolymer of varying chain length and consisting of uronic acid and glucosamine residues, which are sulfated at various positions. The interaction of heparin with antithrombin is the basis for anticoagulation therapy. Heparin accelerates the antithrombin mediated inhibition of factor Xa and thrombin by a conformational activation mechansism and bridging mechanism, respectively. The sequence specific pentasaccharide DEFGH in full length heparin is the most important fragment for high affinity and activation of antithrombin, without which the heparin is incapable of binding to antithrombin. Although heparin is a commonly used

anticoagulant, it suffers from serious side effects including bleeding complications, heparin-induced thrombocytopenia, and intra- and inter-patient dose response variability.

Desai and co-workers have shown that it is possible to replace the GAG skeleton by small, non-saccharide sulfated molecules as antithrombin activators. However, the designed molecules were found to be weak activators of antithrombin due to their binding to the extended heparin-binding site (EHBS), instead of the pentasaccharide-binding site (PBS), of antithrombin.

To design better non-saccharide antithrombin activators, a virtual screening-based approach was employed. Combinatorial virtual screening of 24576 molecules based on tetrahydroisoquinoline core scaffold resulted in 92 hits that were predicted to bind preferentially in the PBS of activated antithrombin with good affinity. The work resulted in a predicted pharmacophore consisting of a 5,6-disulfated bicyclic tetrahydroisoquinoline and a 2′,5′-disulfated unicyclic phenyl ring connected by a 4- to 5-carbon linker. The work has led to several hypotheses, which are being tested in the laboratory through synthesis and biochemical evaluation.

To understand the mechanism of heparin binding to thrombin in greater detail, structural biology and molecular modeling approaches were used. More specifically, the nature of the heparin binding to thrombin was studied with a special focus on understanding the specificity of recognition. Comparative analysis was performed with heparin–antithrombin interaction to assess similarities and differences between the two heparin binding systems. In antithrombin, three important amino acids are involved in heparin pentasaccharide binding, while in thrombin, at least seven basic amino acids are

predicted to be involved. For biological systems, one would expect greater specificity with more interacting points. However, the heparin–thrombin system interestingly displays a lack of specificity. The molecular basis for this lack of specificity is not clear.

A study of antithrombin and thrombin crystal structures with regard to surface exposure, flexibility, and geometry of basic amino acids present in the respective heparin binding site provides the basis for the specificity of recognition (or lack thereof) in the two systems. Interestingly, analysis of thrombin exosite-II showed that Arg101, Arg165 and Arg233 are spatially conserved and form a local asymmetric center. Using *in-silico* docking techniques, selected tetrasaccharide sequences were found to specifically recognize this triad of amino acids indicating the possibility of specific recognition of thrombin. This hypothesis led to the design of a putative lead sequence that is 50% smaller in size and contains 62.5% fewer charges in comparison to the literature reported known exosite II sequence. The design of novel putative 'specific' exosite II sequence challenges the idea that the thrombin–heparin interaction is completely non-specific and gives rise to novel opportunities of designing specific thrombin exosite-II ligands.

CHAPTER 1

INTRODUCTION

Drug design is the process of finding drugs for a particular disease or condition by design, unlike the traditional trial-and-error method or serendipitous discovery. Design, by definition, implies that there is a strategy involved to confront a problem. If design involves rational thoughts and reasoning, then such a process is termed rational drug design. Rational drug design that utilizes the known three-dimensional geometry of a biological target is known as structure-based drug design (SBDD).[1] In contrast, drug design directed only by the known geometry of a potent ligand is referred to as ligand-based drug design (LBDD).[1,2]

The fundamental principle that forms the basis for structure-based design is structural and chemical complementarity between the target and its ligand. The three-dimensional geometry of a target is used either to design completely new ligands or to choose existing ligands by screening collections of compounds. Identifying completely new chemical classes of molecules that are chemically distinct from previously characterized leads for a biological target is termed *de novo* drug design.[1] However, modifying a ligand to improve its physiochemical, pharmacological and drug-like properties is known as lead optimization.

Development of a drug that is potent, efficient, and orally bioavailable, and has minimal or no side effect is a lengthy process involving major investments including an enormous amount of time (usually about 10 to 12 years) and billions of dollars. Any effort or tools that reduce the time and money spent in drug discovery can significantly influence the way drugs are discovered, and is considered an utmost need for pharmaceutical interests and for the benefit of suffering patients.

1.1. Computer-Aided Drug Design (CADD)[1]

The important role of CADD in the drug discovery process is to accelerate the identification of new lead compounds and lead optimization for a biological target/process (Figure 1). The pipeline of drug discovery from idea to market consists of the following general basic steps: a) disease selection, b) target selection, c) lead compound identification, d) lead optimization and e) preclinical and clinical studies.

In practice, these steps are repeated and revisited at several stages based on the results in each step. The compounds for testing can be obtained from natural sources such as plants, animals, microorganisms and by chemical synthesis. Not all compounds become lead compounds with optimal properties that can go on to become a drug candidate, and many fail owing to the absence of activity, poor pharmacokinetic disposition, unacceptable toxicity, insufficient efficiency, or even complexity of synthesis.

Extensive genome decoding of various organisms, including man, proteomic investigations, discoveries of molecular mechanisms of many diseases, and advances in protein chemistry have led to a dramatic increase in the number of new potential targets.

The advancement in these areas of science and the results they produce could be exploited to find new lead compounds by employing computer-based theoretical and experimental approaches.

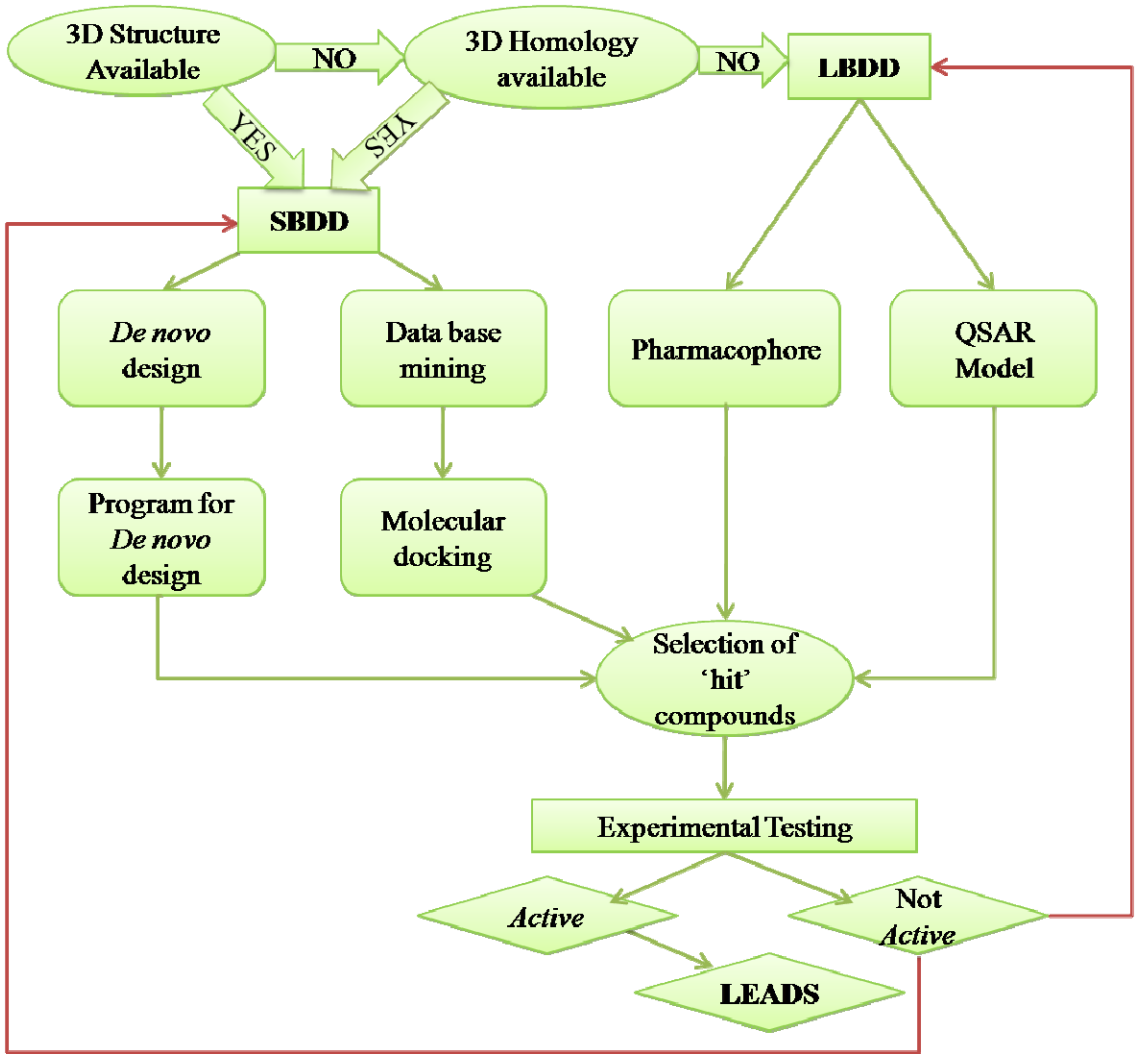


Figure 1. Pathways of computer-aided drug design (simplified). LBDD: Ligand-Based Drug Design, SBDD: Structure-Based Drug Design, QSAR: Quantitative Structure-Activity Relationship. Adapted from Veselovsky, A. V.; Ivanov, A. S. Strategy of computer-aided drug design. *Current Drug Targets – Infectious Disorders* **2003,** *3*, 33-40.

1.2. Hardware and Software Requirements

Modeling of the interactions of macromolecules with ligands requires various methods of calculations and structural visualization. Such studies often require multiprocessor computer systems under UNIX management. Two widely used commercial software packages for CADD are SYBYL (http://www.tripos.com) and Insight II/Discovery Studio (http://www.accelrys/insight). These software packages have many modules that are designed to facilitate various aspects of the drug design process. Many other freeware and shareware molecular modeling programs are also available for molecular modeling purposes. It is crucial to have access to databases with structures of macromolecules and small compounds. The structures of proteins are assembled in Protein Data Bank (PDB; http://www.rcsb.org/pdb) as PDB files from which protein coordinates can be downloaded. Small molecules are available in the Cambridge Structural Database (CSD; http://www.ccdc.cam.ac.uk/prods/csd/csd.html). Primary sequence data is provided by the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov), and the European Molecular Biology Laboratory (EMBL; http://www.ebi.ac.uk/embl/) and the Swiss Institute of Bioinformatics (http://www.expasy.org). Many other databases of low-molecular weight compounds are also available, but databases with commercially available compounds are preferable for practice, because these compounds can be quickly acquired for experimental testing (ZINC; http://zinc.docking.org).

1.3. Ligand-Based Drug Design

Ligand-based drug design is applied when the three-dimensional structure of the macromolecular target is unknown. These methods are based on analysis of sets of ligands with known biological activity. They include the design of pharmacophore models,[2] analysis of quantitative structure-activity relationships ("classic" QSAR)[3] and its relative 3D-QSAR (which takes into account spatial structure of compounds),[4] quantitative structure-property relationships (QSPR),[5] among others. Methods of LBDD can be used for lead compound discovery and lead optimization of previously known ligands.

Pharmacophore models consist of a set of points in space with the specific physiochemical properties and corresponding distances between them that define the elements necessary for the binding of a given group of ligands with a target. Pharmacophore elements include positively and negatively charged atoms, cyclic groups, aromatic rings, and hydrogen bond donor or acceptor atoms. Such a model is designed by alignment of a set of known ligands to reveal commonly-shared features of different molecules in one molecular space area.

1.3.1. Classic QSAR

Quantitative structure-activity relationships are mathematical relationships linking the chemical structure and pharmacological activity in a quantitative manner for a series of compounds. It is generally accepted that Crum-Brown and Fraser were the first to link physiological action θ and "chemical constitution" C using the expression shown below.[6]

$$\theta = f(C) \qquad \text{Eq. (1)}$$

Meyer and Overton correlated biological activity with oil/water partition coefficients of a series of narcotic substances, which added credence to the above equation.[7,8] One of the most widely utilized, successful empirical approaches using descriptors is Linear Free Energy Relationships (LFER) as embodied in the early work of Burkhard, the equations of Louis Hammett and the contributions of Corwin Hansch.[9-12]

Corwin Hansch considered drug action to be a result of two independent processes: i) Transport of the drug from the site of application to the site of action and ii) Non-covalent interactions of the drug with its binding site receptor. Since neither very polar nor very lipophilic compounds have a good chance to permeate several lipid and aqueous phases, he formulated a nonlinear lipophilicity relationship for the transport. Later, lipophilicity terms and electronic parameters, molar refractivity and steric terms were combined in a linear free energy related model to describe the ligand-receptor interaction.

From the inception of QSAR, various descriptors have been developed. The following are important categories of descriptors that have been combined to generate the linear and nonlinear equations that are common to QSAR.

Electronic Descriptors: Hammett's electronic descriptor σ is described by the following equation:

$$\sigma = \log {}^{K_X}/_{K_H} \qquad \text{Eq. (2)}$$

K_X and K_H represent the ionization constants in water at 25 °C of substituted and unsubstituted benzoic acids, respectively. Sigma (σ) is defined as a substituent constant such that positive values of σ represent electron-withdrawing substituents and negative values of σ pertain to electron-releasing substituents. It was subsequently modified as:

$$\rho\sigma = \log {}^{K_X}/_{K_H} \qquad \text{Eq. (3)}$$

The proportionality constant rho (ρ) is defined as a reaction constant and is specific for each reaction under its unique constraints. Thus, ρ is a measure of the susceptibility of a given reaction to the electronic effects of substituents. Numerous other Hammett-based electronic parameters are available.

Hydrophobicity: Hydrophobicity may be defined as the tendency of organic molecules to shy away from water and partition into a less polar phase, which constitutes their own bulk.[13] The classical model of hydrophobic interactions was delineated by Kauzman to assess van der Waals attractions between the nonpolar parts of two molecules immersed in water.[14] The "squeezing out" of water molecules in the vicinity of the mutually bound apolar surfaces provides the driving force for hydrophobic interactions. Thus, the gain in entropy is primarily governed by the repulsion of hydrophobic solutes from solvent water and the limited but critical capacity of water to maintain its network of hydrogen bonds.[15] A hydrophobe immersed in water decreases the entropy of the water by forcing the water molecules to form a 'clathrate' cage structure around the hydrophobic portion. The decrease in entropy is thermodynamically unfavorable.

Steric Descriptors: Steric descriptors represent the bulkiness of molecules. The first steric effect in physical organic chemistry to be numerically defined was Taft's E_S parameter.[16] It was defined as the ratio of the log of acidic hydrolysis of aliphatic esters of the type RCOOR' and CH_3COOR'.

$$E_s = \log {}^{k_R}/_{k_H} \qquad \text{Eq. (4)}$$

Other often-utilized steric descriptors in SAR studies include Charton's steric parameter,[17,18] Hancock's steric parameter,[19] molar refractivity,[20] Verloop's STERIMOL parameters[21] and Hopfinger's molecular shape analysis.[22]

Hydrogen Bonding Descriptors: The importance of hydrogen bonding, a fundamental chemical property in biological systems, is highlighted by its critical role in defining the structure of biomacromolecules like proteins and nucleic acids. Hydrogen bonding ability is classified into hydrogen bond donor and acceptor capacities. The importance of hydrogen bonding is unequivocally recognized in ligand-receptor interactions. Most often, the hydrogen bonding capacity of a molecule is encoded as the number of hydrogen bond donors, acceptors or donatable hydrogens.

Dipole Moments: Dipole moments, μ, have been successfully used to correlate biological activity where whole-molecule parameters are critical. A large number of group dipole moments for aromatic and aliphatic substituents have been assessed by Lien *et al.*[23,24]

Quantum Chemical Indices: Use of quantum chemical descriptors is constantly increasing because of the developments in high-speed computing, as well as the accuracy

and precision of computed values.[25] Direct derivation of electronic descriptors can be obtained from the molecular wave function. The most common approach utilized for solution of the electronic Schrödinger equation is the Hartree-Fock self consistent field (SCF) method.[26] Thus, each electron is assumed to move in the average field of all the other electrons. Two different approaches are used to solve the molecular Schrödinger equations: *ab initio* and semi-empirical.

ab initio calculations include all electrons and all one- and two-electron integrals. Computational time is proportional to a high exponential of the number of electrons in the molecule (N^4 or N^5). Thus *ab initio* calculations are limited by the types of atoms and size of molecules.[27]

In semi-empirical methods, only valence electrons are explicitly included; some integrals are neglected and others are approximated.[28] CNDO (neglect of diatomic and single atom atomic orbital overlap), MNDO, AM1 and PM3 (neglect of diatomic overlap only) are commonly used to methods calculate molecular descriptors such as atomic charges (qX), molecular orbital energies (E_{HOMO}, E_{LUMO}, E_{LUMO}–E_{HOMO}), superdelocalizabilities (s), molecular polarizability (α), dipole moments (μ) and energies (E_T). For a detailed description of all parameters, see reference 25.

Topological Indices: In general, topology refers to the way in which atoms are connected to each other in a molecule thus representing the shape of the molecule. The best-known topological parameters are the molecular connectivity indices which are strictly based on molecular structure.

1.3.2. 3D-QSAR

In 3D-QSAR, the 3D structure of the compounds is considered instead of only 2D structure. 3D-QSAR has some advantages over classical 2D-QSAR. For example, more heterogeneous sets of compounds can be included than in classic QSAR. Molecular fields are calculated instead of substituent constants, and contour maps can be prepared to show the effect of certain properties in specific regions. CoMFA (Comparative Molecular Field Analysis)[4] and CoMSIA (Comparative Molecular Similarity Index Analysis)[3] are well-known examples of 3D-QSAR.

The ligand-based drug design methodologies like classic QSAR and 3D-QSAR have been used for its predictive power more often than design purposes, in part due to the difficulties involved in translating the descriptor information into meaningful structural interpretation. In recent years, SBDD has been employed more extensively than LBDD since the former provides detailed and easily-interpretable information about the target of interest. However, LBDD can be employed to model ADME-Tox (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties.

1.4. Molecular Docking and Scoring

The docking process can be defined as the search for the correct binding geometry of the ligand (binding mode or pose) in a target binding site. In general, the docking process consists of two interrelated components namely; identification of the binding mode and the prediction of binding affinity.[29]

The first part of the docking process is the sampling of the ligand (and sometimes the receptor) conformational space and placement of the ligand into the receptor binding site. The second part is strictly a ranking process of the different docked solutions (poses) using the scoring function, which in turn may be correlated to binding affinity.

The docking process involves the translation of theoretical model of the enzyme inhibition or ligand recognition into a computational model. For example, for an enzyme E and inhibitor I, docking aims at predicting the correct structure of the complex [E+I] = [EI] under equilibrium conditions.

$$[\mathrm{E}]_{\mathrm{aq}} + [\mathrm{I}]_{\mathrm{aq}} \leftrightarrow [\mathrm{EI}]_{\mathrm{aq}} \qquad \text{Eq. (5)}$$

$$\mathrm{K_A} = \mathrm{K_I}^{-1} = \frac{[\mathrm{EI}]}{[\mathrm{E}][\mathrm{I}]} \qquad \text{Eq. (6)}$$

$$\Delta G = -RTlnK_A \qquad \text{Eq. (7)}$$

Calculation of the free energy of binding (ΔG) is related to binding affinity K_A by equations 6 and 7. Although prediction of the correct binding geometry of the [E+I] complex does not require information about K_{A}, prediction of biological activity requires this information. Therefore, docking process must consider the following important factors: steric, electrostatic, hydrogen bonding, inhibitor strain (if flexible) and enzyme strain. In addition, when considering the equilibrium shown in Equation 5, the following factors are also important: desolvation, rotational entropy and translational entropy.

1.4.1. Molecular Docking Algorithms

Protein-ligand docking is a geometric search process. The output from the docking algorithm includes a list of protein-ligand complexes rank-ordered by a given scoring function.[30] Several docking algorithms have been developed that place rigid or flexible ligands in mostly rigid-but recently also somewhat flexible protein binding sites (e.g., DOCK,[31-34] MOE-Dock,[35]AutoDock,[36] FlexX,[37-39] Hammerhead,[40] GOLD,[41] FLOG,[42] Glide,[43] PRO_LEADS[44]). Some algorithms allow partial protein flexibility. Docking algorithms are complemented by scoring functions that are designed to identify the correct binding mode. Various functions have been developed to measure the protein-ligand binding affinity in the docking algorithms. Since many of the functions are not strictly related to binding free energies, functions designed to rank different protein-ligand complexes are referred to as scoring functions.[45-48]

A. Rigid Docking

In rigid docking, the docking problem is simplified by neglecting the conformational degrees of freedom of the ligand molecule. The algorithms based on this approximation can be applied to docking of small or rigid molecules, molecule fragments or conformational ensembles of molecules.

a. Clique Search-Based Approaches

The docking of two rigid molecules can be implied as a problem of matching characteristic features of the molecules in space.[49] A distance compatibility graph is used

to search for matching features. The algorithm for rigid-body docking in the DOCK program is based on the idea of searching for distance-compatible matches. Since its first introduction in 1982, the DOCK software has been extended in several directions. Furthermore, several scoring functions are now applied in combination with the DOCK algorithm.[50-54] Algorithms based on clique-search based approaches also include LUDI,[54,55] CLIX[56] and ADAM[57] but differ in the features used for matching and the way in which they are represented.

b. Geometric Hashing[30,58]

Hashing is a computer science technique which allows fast access to data. The geometric hashing algorithm has two phases: the preprocessing phase, in which the geometric hash table is constructed from a single ligand or a set of ligands to be docked and the features are stored. In the recognition phase, the protein features are used to vote for hash entries of the ligand. A vote means that there is a protein feature that matches a ligand feature. Geometric hashing is a time-efficient method. Pose clustering is another approach used in molecular docking, which is primarily based on pattern recognition but also uses the hashing scheme to match features.[59,60]

B. Flexible Ligand Docking

Rigid docking is very limited in that the conformation of the bound ligand must be known before the docking process begins in order to obtain accurate results. Since most drug-like molecules have at least a few rotatable bonds or even flexible ring systems, it is

important to incorporate ligand flexibility into any generally applicable docking algorithm. Algorithms that treat ligand flexibility can be classified into essentially three categories: a) systematic methods (incremental construction, conformational search, databases); b) random or stochastic methods (Monte Carlo, genetic algorithms, tabu search) and c) simulation methods (molecular dynamics, energy minimization).

I) Systematic methods

a. Conformational Search

In principle, every conformer of a set of flexible ligands could be evaluated with rigid-body docking algorithms. However, as the size of the conformational ensemble increases computing time also increases exponentially. Thus a balance between computing time and the desire to cover all of conformational space is required. The Flexibase and FLOG docking algorithms are based on conformation ensembles and use libraries of pre-generated conformations.[61-63]

b. Fragmentation

Fragmentation is one of the most popular approaches for handling ligand flexibility. The ligand is divided into smaller fragments which can be treated as conformationally rigid or by a small conformational ensemble. There are two ways of handling fragments in the active site, namely “place-and-join” and “incremental construction”.

In place-and-join strategy[64-67] all (or a subset of) fragments are placed in the receptor site independently and reconnected in favorable orientations until they constitute a complete ligand.

Incremental construction is the preferred fragmental molecular docking approach, in which a fragment is first placed in the receptor site and then the remaining fragments are added to the orientations of the first one. The first incremental construction based docking algorithm was developed by Leach and Kuntz[68] for the DOCK program. FlexX is a fully-automated incremental construction algorithm-based molecular docking approach developed for virtual screening.[69-72] Glide[43] and Hammerhead[40] are two other approaches that are based on incremental construction.

II) Random or Stochastic Methods

a. Genetic Algorithms

The genetic algorithm[73] is a general purpose optimization method that adapts the principles of biological competition and population dynamics. Application of the genetic algorithm in molecular docking includes a linear representation of a ligand and receptor conformation called chromosomes and a fitness function to decide which individuals survive and produce the offspring. The chromosomes that correspond to the best intermediate solutions are subjected to crossover and mutation operations analogous to gene recombination and mutation to produce the next generation of ligand/receptor conformations.

For docking applications, the genetic algorithm solution is an ensemble of possible ligand conformations. Jones et al. developed one of the first genetic algorithms for molecular docking and their ideas are implemented in the GOLD docking program.[74,75]

b. Monte Carlo Algorithms

In Monte Carlo (MC) simulation, the local movements of the atoms are performed randomly. The two major components of the MC algorithm are the description of the degrees of freedom and the energy evaluation. The degrees of freedom should be described such that high-energy states are avoided. The energy evaluation is the most time-consuming part of the process and must be made as efficient as possible. Often energy potentials are precalculated on a grid to speed up this step. QXP, ICM, and PRODOCK are examples for MC-based docking algorithms.[76-78]

c. Tabu Search

Tabu search starts with an initial random structure and new structures are created by random moves. During the optimization iterations, a list (the tabu list) is maintained containing the best and most recently visited configurations. Moves resulting in configurations close to one in the tabu list are rejected except if they are better than the best-scoring one. This technique improves the sampling properties by avoiding revisitation of previously sampled configurations. Tabu search is the underlying method of docking in PRO_LEADS.[79]

III) Simulation methods

Simulation approaches begin their calculation with a starting conformation and move locally to conformation with lower energy instead of trying to list the population of a discrete low-energy subspace of the problem. Simulated annealing and molecular dynamics are two important simulation techniques for solving the docking problem.

a. Simulated Annealing[80-83]

The simulation process starts with an initial configuration *A* of a ligand in an active site with energy *E(A)* or score value. This initial configuration is scored. Then, it generates a new configuration *B* with energy *E(B)* and also scored. If a new solution scores better than the previous one, it is immediately accepted. These steps are repeated until the desired number of configurations is obtained. The AutoDock program for protein-ligand docking developed by Goodsell et al. is based on the simulated annealing technique.[80]

b. Molecular Dynamics

Molecular Dynamics simulations can in principle be used to solve molecular docking problems.[84-89] However, the limitation with this methodology is that it is quite time consuming. It follows a path from a starting orientation to low-energy configurations. In addition, several simulations with different starting orientations must be performed to get a statistically significant model. Though this method is not very popular for virtual screening, it is a valuable method for analyzing small sets of ligands.

1.4.2. Molecular Docking: Scoring Functions

Scoring functions in molecular docking process have a twofold function: to direct the docking and to predict the binding affinity of the final poses. The evaluation and ranking of predicted ligand conformations (pose) is a vital part of the structure-based virtual screening.

Binding free energy is a collective term of many factors and can be written as:

$$\Delta G^{\circ}{}_{bind} = \Delta G_{solv}^{\circ\, complex} - \Delta G_{solv}^{\circ\, prot} - \Delta G_{solv}^{\circ\, lig} + \Delta G^{\circ}{}_{int} - T\Delta S^{\circ} + \Delta\lambda \qquad \text{Eq. (8)}$$

where, $\Delta G^{\circ}{}_{bind}$ is total binding free energy, $\Delta G_{solv}^{\circ\, complex}$ is solvation energy of the protein−ligand complex, $\Delta G_{solv}^{\circ\, prot}$ is solvation energy of the protein, $\Delta G_{solv}^{\circ\, lig}$ is solvation energy of the ligand, $\Delta G^{\circ}{}_{int}$ is interaction energy of the protein−ligand, $T\Delta S^{\circ}$ is change in the entropy for protein−ligand interaction and $\Delta\lambda$ is conformational change in protein−ligand complex.[90]

Accurate and reliable prediction of binding free energy by computational methods is a very challenging task. For example, the entropy, a phenomenon that is very influential in the physiological condition where the protein-ligand interaction takes place, can only be crudely estimated. Another problem is that the fast estimation of binding free energy implemented in docking programs makes various assumptions and simplifications contributing to inaccuracy in the result. We will briefly discuss the different scoring functions here in this chapter.

Essentially, four types or classes of scoring functions are currently applied: force field-based, semi-empirical, empirical and knowledge-based scoring functions.

Scoring functions derived from a force field use non-bonded interaction terms to calculate the score, sometimes in combination with solvation terms. Semi-empirical approaches are those in which molecular mechanics terms are supplemented by additional parameters or terms that are empirically derived from observation. Empirical scoring functions employ multivariate regression methods to fit coefficients of physically motivated structural functions by using a training set of protein-ligand complexes with measured binding constants. Knowledge-based scoring use statistical atom pair potentials derived from structural databases as the score.

A. Force Field Scoring

Standard force field scoring functions were originally based on the idea of using only enthalpic gas-phase contributions to estimate the binding free energy. The main advantage of force field scoring is that when used on a precomputed grid, it is fast and transferable. The disadvantage is that force field scores evaluate only parts of the relevant energies, namely, potential energies and ignoring some fundamental contributions such as hydrophobic interactions, solvation, and entropic effects. Though there are many force fields used, the well-known and widely-applied molecular mechanics (MM) force fields include AMBER,[91,92] CHARMM,[93] TRIPOS,[94] MM2,[95] MM3[96] and MM4.[97]

Energy calculations performed by these methods are essentially the sum of electrostatic and van der Waals potentials, plus internal (i.e., intramolecular) distance, angle and torsion contributions:[90]

$$\sum_{total} = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_e\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos(n\varphi - \gamma)\right]$$

$$+ \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon R_{ij}}\right] \qquad \text{Eq. (9)}$$

Where, K_r is spring constant, r is the distance between two atoms, r_{eq} is the equilibrium bond length, K_e is angle bending force constant, θ is bond angle and θ_{eq} is equilibrium bond angle, V_n is the torsional barrier, n is periodicity, φ is the torsion angle, γ is phase shift (offset), A_{ij} is repulsive force constant, B_{ij} is attractive force constant, R_{ij} is distance between atoms i and j, q_i is charge (partial charge) of atom i, q_j is charge (partial charge) of atom j and ε is dielectric constant.

Most force field scoring functions only consider a single protein conformation, which makes it possible to omit the calculation of internal protein energy, which greatly simplifies scoring.

B. Semi-empirical Scoring Functions

Semi-empirical approaches imply the use of empirical or empirically-calibrated energetic terms for calculating interactions not commonly computed by molecular mechanics. Although this strategy allows the inclusion of contributions for fundamental

biological interactions, e.g., hydrogen bonding or solvent effects, semi-empirical models partially lose the universal applicability typical of MM force fields. ICM, GOLD, AutoDock, SDOCK, are some of the docking algorithms that includes semi-empirical scoring functions.

C. Empirical Scoring Functions

Empirical scoring functions estimate the binding free energy by summing interaction terms derived from weighted structural parameters, as first proposed by Böhm. The design of empirical scoring functions is based on the idea that binding energies can be approximated by a sum of individual uncorrelated terms. The weights are obtained by fitting the scoring function to experimental binding constants of a training set of protein–ligand complexes.

The functional forms of empirical scoring functions are often simpler than force-field scoring function and are simple to evaluate, but they are based on approximations similar to force field functions. The main drawback of empirical scoring functions is that it is unclear whether they are able to predict the binding affinity of ligands structurally different from those used in the training set.[98,99]

D. Knowledge-Based Scoring Functions

Knowledge-based scoring functions are designed to reproduce experimental structures rather than binding energies. Knowledge-based scoring functions represent the binding affinity as a sum of protein−ligand atom-pair interactions. Popular

implementations of such functions include Potential of Mean Force (PMF),[100-102] DrugScore[103] and SMoG.[104] The major advantage of knowledge-based scoring functions is their computational simplicity, which permits screening of large compound databases.

E. Consensus Scoring

Given the fact that currently no scoring function performs consistently better than the others for multiple targets,[105] two strategies have emerged: identifying the best performing function for each target or merge several scoring functions in a consensus approach. Consensus scoring could reduce the number of false positives identified by individual scoring functions. The concept of consensus scoring was first introduced by Charifson[106] based on the assumption that the combination of different functions would overcome inherent individual weaknesses and lead to better and more general scoring performance. A very good example for implementation of consensus scoring is X-CSCORE[107] which combines GOLD-like, DOCK-like, ChemScore, PMF and FlexX scoring functions. A limitation of consensus scoring arises when terms in different scoring functions are significantly correlated, where calculation errors are amplified rather than balanced.

1.5. Molecular Docking: GOLD Docking Protocol[108]

GOLD (Genetic Optimization for Ligand Docking) is a genetic algorithm for docking flexible ligands into protein binding sites. GOLD was written by Gareth Jones (University of Sheffield, UK) in DTI LINK (the Department of Trade and Industry)

collaboration with GlaxoWellcome and the Cambridge Crystallographic Data Centre (CCDC).

1.5.1. Preparing the Protein Structure for Docking

Protein coordinates extracted from PDB (Protein Data Bank) files usually need to be pre-processed prior to docking. All hydrogen atoms are added including those necessary to define the correct ionization and tautomeric states of residues such as Asp, Glu and His. All bond types are checked. The sequence order, name of the sequence and atoms are correctly typed and named. Any unusual bonds such as disulphide bridges are checked to see if they have CONECT records. If a metal ion is present, all bonds between the ion and coordinating protein or water atoms are deleted (GOLD finds them automatically).

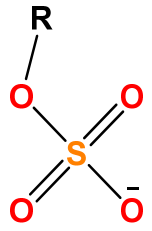
The corrected protein file is saved in MOL2 format. GOLD assigns atom types from the information about element types and bond orders in the input structure file, so it is important that these are correct.

1.5.2. Preparing the Ligands for Docking

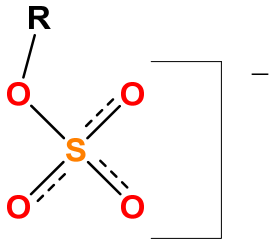
Structure building tools are available in many modeling software packages including SYBYL. Ligands can be drawn and energy-minimized before docking. In order to predict correct protein−ligand binding modes for the ligand, all hydrogen atoms, including those necessary to define the correct ionization and tautomeric states are added and that all bond types are checked for its correctness.

For groups which can be drawn in more than one way (i.e., have more than one canonical form), such as nitro, carboxylate, sulfate and amidinium a special attention must be paid in defining their atom type and bond type (Figure 2).

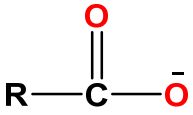
The starting geometry of the ligand should be reasonably low in energy, since GOLD will not alter bond lengths or angles, or will rotate *rigid* bonds such as amide linkages, double bonds and certain bonds to trigonal nitrogens. However, GOLD optimizes the values of torsion angles around rotatable bonds. The corrected ligand is saved as a MOL2 file.



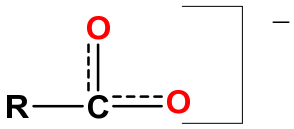
A. Standard sulfate representation



B. Sulfate representation for GOLD docking (SYBYL atom type o.co2)



C. Standard carboxylate representation



D. Carboxylate representation for GOLD docking (SYBYL atom type o.co2)

Figure 2. Standard molecular representation of sulfate and carboxylate groups (A and C). In GOLD, the bond types of all three terminal S-O and two C-O bonds are treated as aromatic bonds and the terminal oxygens are treated equally as charged carboxylate oxygens (B and D). This redefinition of sulfate group is important for docking purposes in order for GOLD to correctly recognize the sulfate group. In reality, each of the terminal oxygens are (partially) charged, not just one as shown in the standard representation (A and C).

The precise geometric positions of rotatable (e.g., hydroxyl and amino) hydrogen atoms do not matter, as they are optimized during the GOLD run. GOLD deduces hydrogen-bonding abilities from the presence or absence of donatable hydrogen atoms. For example, the protonation state of a carboxylic acid group can be controlled by adding or removing the ionizable hydrogen atom. GOLD ignores assigned atom charges, both formal and partial. It deduces whether an atom is charged by counting the bond orders of the bonds that it forms and comparing the result with the atom's normal valence.

Since ring conformations and the torsion angles around *rigid* bonds such as amide linkages, double bonds and certain bonds to trigonal nitrogens are normally fixed at their starting values during docking, a good practice is to perform a few cycles of molecular-mechanics minimization to take the ligand close to its local potential-energy minimum.

GOLD does not alter stereochemistry. Care must be taken to ensure that the ligand possesses the correct stereochemistry. In cases where we are not sure about the stereochemistry or if it is undefined, alternate stereoisomers must also be generated to make comparisons between fitness scores for dockings of different stereoisomers.

1.5.3. Ligand Flexibility: Fixing Rotatable Bonds

Although GOLD is designed to dock flexible ligands into protein binding sites, it also can be useful to fix the geometry of part or all of the ligand e.g., in order to study the possible binding of a pre-determined ligand geometry. This also restricts the search space, therefore convergence may be higher. This can be used to greatly simplify the docking of large molecules with inflexible regions (i.e., the rigid backbone hypothesis).[109]

1.6. Molecular Docking of Sulfated Molecules

1.6.1. Molecular Docking Sulfated Glycosaminoglycans (GAGs)

Glycosaminoglycans (GAGs) are complex carbohydrate biopolymers also known as mucopolysaccharides because of their viscous lubricating properties as found in mucous secretions. GAGs interact with a wide range of proteins and exhibit important roles in various physiological and pathological processes such as inflammation, coagulation, angiogenesis, cell adhesion and viral invasion by interacting with several different proteins.[110-112]

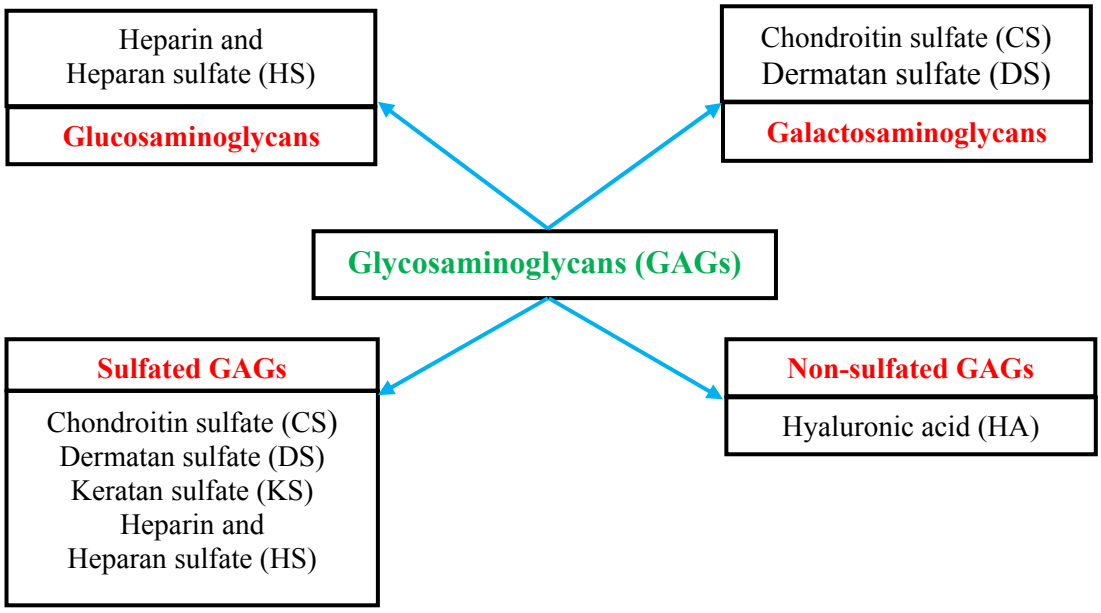
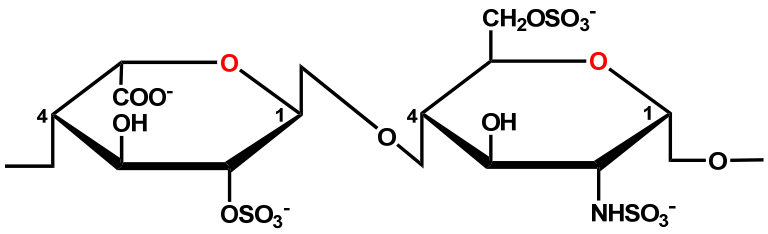


Figure 3. A simple classification of glycosaminoglycans.

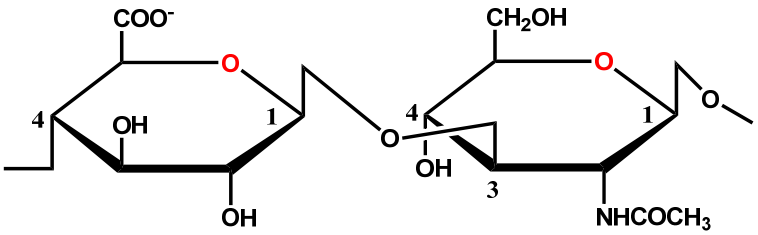
The linear, sulfated, negatively charged GAGs are built up of several different sulfated disaccharide units.[113] The repeating units are composed of alternating uronic acids (D-glucuronic acid or L-iduronic acid) and amino sugars (D-galactosamine or D-

glucosamine). GAGs are differentiated according to the type of hexosamine, hexose or hexuronic acid unit that they contain, as well as the geometry of the glycosidic linkage between these units (Figure 3).

The disaccharide unit may be sulfated or unsulfated at various positions and the ring system can exist in different conformation. GAGs also vary in the geometry of the glycosidic linkage (α or β). As a result, there exists exquisite structural complexity in the GAG sequences.[114] At physiological pH, all carboxylic acid and sulfate groups may be deprotonated, giving GAGs very high negative charge densities (heparin has the highest negative charge density of any known biomolecule).[115]



A. IdoA2S β (1→4) GlcNS6S



B. GlcA β (1→3) GlcNAc

Figure 4. Examples of basic structural units of GAGs. **A.** L-iduronic acid 2-*O*-sulfate and D-glucosamine 2-*N*-sulfate-6-*O*-sulfate linked together via α (1→4) glycosidic bond is a major disaccharide unit in heparin. **B.** D-glucuronic acid and D-*N*-acetyl glucosamine, linked together via alternating β (1→3) and β (1→4) glycosidic bonds is a repeating disaccharide unit in hyaluronic acid.

The structural complexity of GAGs and their roles in various physiological processes makes their targets excellent candidates for structure/ligand-based drug design, where recent developments in molecular modeling techniques could be applied to expedite the understanding of this complex system (structural complexity and molecular modeling details in Chapter 3).

Molecular modeling and docking as a tool for drug design has been successfully used in many instances where small organic molecules modulate the actions of macromolecules, but molecular docking of sulfated GAGs has not been explored much due to their high negative charge density and conformational flexibility. If all possible conformations of the GAG oligosaccharide and all rotamers of charged side chains in a protein are to be taken into account, an accurate prediction of GAG-protein binding becomes an extremely challenging task.

Molecular modeling techniques have been described for the successful prediction of sulfated GAG binding sites on the surface of proteins. These methods include energy mapping of ligand probes on the surface of proteins, molecular docking and scoring, and molecular dynamics simulations.[116,117]

The prediction of the location of GAG binding sites on the surface of the proteins have been attempted by mapping of sulfate interaction energies using GRID followed by ligand-protein docking to predict the most favorable anchoring position for a charged sulfate group on the surface of proteins.[118] Such studies have been performed with a number of proteins such as aFGF (acidic Fibroblast Growth Factor), bFGF (basic Fibroblast Growth Factor), antithrombin and IL-8 (Interleukin-8).[119]

Most docking studies on heparin-binding proteins have been focused on predicting the amino acids that make up the heparin binding sites.[120] Simulated annealing and genetic algorithms have been used to dock GAGs to their putative proteins or receptors.[121-126]

Since heparin and related GAGs are structurally complex, the prediction of binding modes and energies for these molecules to their biological receptors requires a sufficiently large conformational sampling from the large search space along with an adequate scoring function.

Lam et al. have reported molecular modeling studies to predict the binding of a heparin hexasaccharide to the multi-component complex between bFGF and FGFR1 (Fibroblast Growth Factor Receptor-1).[126] The proposed structural model of the biologically functional dimeric bFGF-heparin complex between heparin, bFGF and FGFR1 has a stoichiometry of 1 heparin: 2 bFGF: 2 FGFR1. This model is consistent with the binding mechanism of FGF to its receptor, the receptor dimerization, and the reported site-specific mutagenesis and biochemical cross-linking data.[126]

In another study, molecular docking predicted that long heparin fragments such as a dodecasaccharide or a tetradecasaccharide are required for binding to the dimer of chemokine stromal cell-derived factor 1α (SDF-1α).[127]

A study of the interaction between a heparin pentasaccharide and AT-III has been carried out by Grootenhuis and van Boeckel by homology modeling and manual docking. However, the geometry determined by them was different from the original crystal structure.[128]

The first approach of combinatorial library screening for Heparin/Heparan Sulfate (heparin/HS) GAGs has been reported and this study demonstrates that library screening is feasible for heparin/HS oligosaccharides, especially if a high-resolution crystal structure of the protein is available.[109] This 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 an average heparin/HS backbone hypothesis. The approach uses a genetic algorithm-based docking and scoring protocol (additional information in Chapters 2 and 3).

Current docking methods aimed at predicting high-affinity GAG sequences have certain limitations that need to be addressed. Since most docking methods based on coarse docking generally fail to take into account any conformational changes that may occur in the protein receptor, it is necessary to include sequences that have both $^{1}C_{4}$ and $^{2}S_{O}$ ring conformations for iduronic acid. Molecular dynamics simulations suggested that the chair form ($^{1}C_{4}$) predominates at monosaccharide level of IdoA2S and the skew-boat ($^{2}S_{O}$) may contribute from ~40% to 60% of the total IdoA2S conformational preference in the entire polysaccharide chain depending on the heparin sequence.[129-132]

In addition, GAG oligosaccharides have many rotatable bonds (large number of degrees of freedom), posing a significant challenge for the search of the correct binding mode. An additional problem arises due to the presence water-mediated interaction that is being neglected in many docking protocols.

Very few Molecular Dynamics (MD) simulations have been performed for sulfated GAGs such as heparin and HS. MD simulations have been performed for the complex of a heparin pentasaccharide with AT-III in order to characterize the energetic contribution of important amino acids required for the interaction with GAG fragments and the ability of GAG fragments to induce the observed conformational change in AT-III.[133]

1.6.2. Molecular Docking of Small Organic Non-Carbohydrate Sulfated Molecules

Most of the research on sulfated molecules has been focused on GAGs due to their extensive role in many physiological processes and disease conditions. This is based on the assumption that a GAG sequence is a critical pharmacophore needed for recognition and activation/inhibition of the target protein.

This concept was challenged by Gunnarsson et al. in Desai's lab when they designed a non-carbohydrate sulfated molecule to mimic the trisaccharide to activate antithrombin.[134] The rational design strategy is based on a study of complexes of natural and mutant antithrombins with heparin-based oligosaccharides using hydropathic interaction (HINT) technique, a quantitative computational tool for analysis of molecular interactions.

Using this approach, a small, nonsugar, aromatic molecule, (-)-epicatechin sulfate (ECS), was designed by Gunnarsson et al. to mimic the non-reducing end trisaccharide unit DEF of the sequence specific heparin pentasaccharide DEFGH. The designed (-)-epicatechin sulfate (ECS) was the first small non-saccharide molecule reported as an activator of antithrombin for the accelerated inhibition of factor Xa, a key proteinase of the

coagulation cascade.[134] (+)-Catechin sulfate (CS), a chiral stereoisomer of ECS, was found to be more active than the (-)-epicatechin sulfate (ECS).[135] The molecular docking study using HINT suggested plausible binding of CS in the extended heparin binding site, which is adjacent to the binding domain for the reference trisaccharide DEF.[135]

Following this dramatic change in designing non-carbohydrate sulfated molecules, many papers have been published by Desai's lab describing new methodologies for docking of sulfated molecules and advances in synthetic methods for efficient synthesis, characterization and biological evaluation (more details in Chapter 2).

1.7. Antithrombin-Based Anticoagulants

1.7.1. Antithrombin Structure and Mechanism of Inhibition

Clotting is a natural defense mechanism of the body to prevent excessive loss of blood and ingestion of microbes upon injury. Yet, unintended clot formation can be very dangerous and is a leading common cause of death. Anticoagulants are used therapeutically in the treatment and prevention of clot formation and clot growth in thromboembolic disorders.

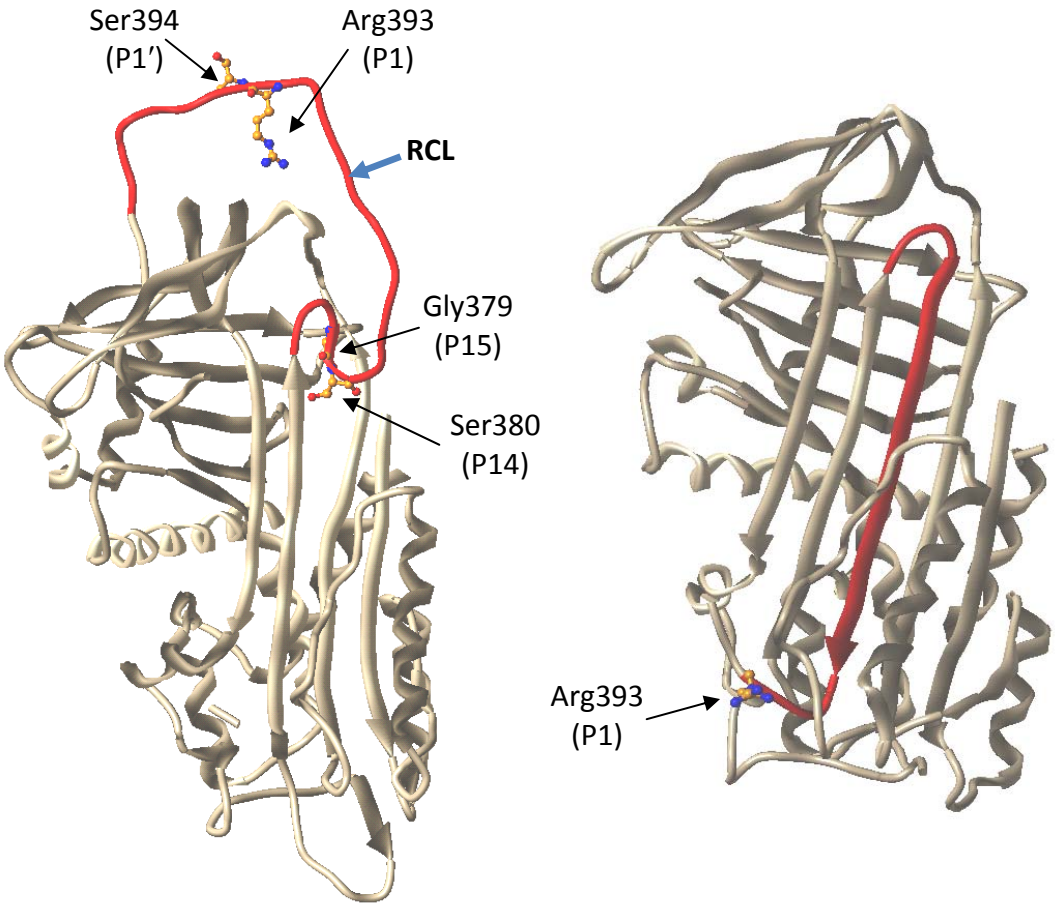
Endogenous antithrombin (AT) is a major regulator of blood clotting. It inactivates a number of proteinase enzymes of the coagulation cascade, especially thrombin and factor Xa.[136] Antithrombin alone is a poor inhibitor. However, heparin-based anticoagulants enhance its proteinase inhibitory activity. Because of this, heparin-based anticoagulants are also called antithrombin activators.

Human antithrombin is a glycoprotein with 432 residues with major (α-antithrombin) and minor form (β-antithrombin). Antithrombin is also a member of the serine proteinase inhibitors (serpin) superfamily of proteins. It shows structural and functional similarity with homologous serpin members including α_1-proteinase inhibitor, heparin co-factor II, and plasminogen activator inhibitor-I.[137,138]

The structure of intact, uncleaved, free antithrombin has nine α-helices surrounding three β-sheets.[139-141] There are two striking features: a dominant five-stranded β-sheet A, approximately in the center of the inhibitor and an exposed 15-residue sequence called reactive center loop (RCL) containing the reactive bond between Arg393 and Ser394 that serves as the substrate for serine protease (Figure 5A). These features are common to all serpins.[137,138,142,143]

Antithrombin has another unique feature known to be present in only one other serpin, heparin co-factor II.[144] Two residues, P15-P14 (Gly379-Ser380) at the N-terminal end of the reactive center loop, are partially inserted as a short β-strand between strands 3 and 4 of β-sheet A (Figure 5A).

The overall structure of antithrombin cleaved at the reactive center loop (RCL) is similar to the intact form, except that the RCL is inserted as strand 4a in β-sheet A (Figure 5B). This structural change results in the movement of the P1 residue to the opposite side of the protein, a distance of approximately 70 Å. This dramatic conformational change following cleavage by the target enzyme leads to significant thermodynamic stabilization of the molecule.[145,146]



A. Native antithrombin **B. Cleaved antithrombin**

Figure 5. Ribbon diagram of native (A) and cleaved (B) plasma antithrombin. The crystal structure is from PDB entry 1ath. The Reactive Center Loop (RCL) is shown in red. Reactive bond residues Arg393 (P1) and Ser394 (P1′) and the partial insertion loop Gly379 (P15) and Ser380 (P14) are shown in ball-and-stick representation. The P1 residue, in an exposed orientation in the native structure, moves to the opposite end of the molecule following cleavage with a proteinase, whereupon the RCL is inserted as strand 4 in 6-stranded β sheet A. The proteinase moves with the P1 residue to the bottom of the inhibitor and is covalently trapped. Adapted from Desai, U. R. New antithrombin-based anticoagulants. *Med Res Rev* **2004**, *24*, 151-181.

A recent crystal structure of the α_1-proteinase inhibitor–trypsin complex[147] and biochemical results[148,149] indicate that the dramatic structural change following cleavage of

the P1-P1′ bond is critical for the disruption of the catalytic triad of the proteinase, which results in the inactivation of the enzyme.

Thrombin and factor Xa inhibition by antithrombin is referred to as the serpin 'mousetrap' mechanism. Antithrombin (AT) acts as a bait to trap the target enzyme (E) in an equimolar, covalent, inactive complex (AT*-E*). In the first step, the RCL interacts with the active site of the proteinase to form a Michaelis complex (AT:E). This is rapidly followed by cleavage of the scissile bond P1-P1' in the RCL to form an acyl-enzyme intermediate (AT-E), which undergoes a major rearrangement to disrupt the enzyme's catalytic triad[147-150] resulting in inhibition (E*-AT*) (Inhibition Pathway, Figure 6). In the substrate pathway (Figure 6), structural perturbations in antithrombin (e.g., mutational changes) may facilitate rapid hydrolysis of the acyl-enzyme intermediate E-AT to yield an active enzyme (E) and a cleaved inhibitor (AT_C) that diminishes efficacy of inhibition.

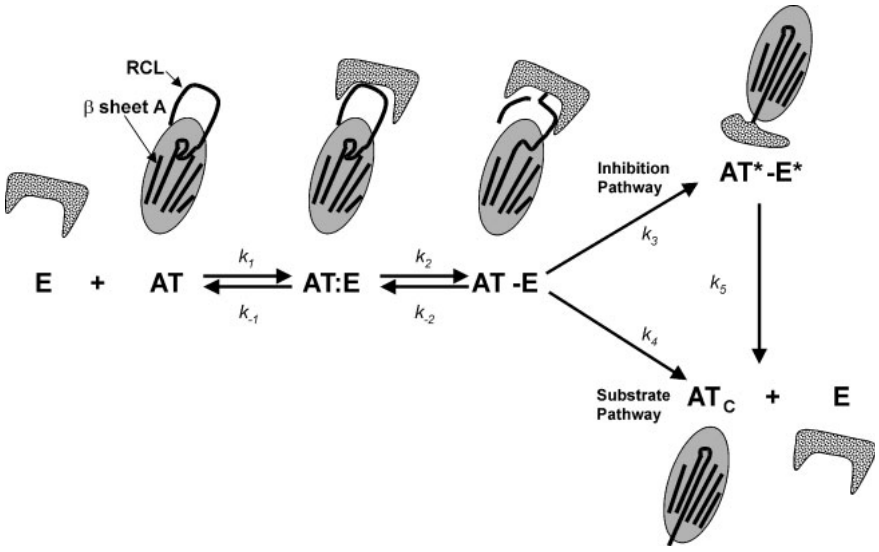


Figure 6. A model of the serpin 'mouse trap' mechanism of inhibition (See text for details). Reproduced from Desai, U. R. New antithrombin-based anticoagulants. *Med Res Rev* **2004**, *24*, 151-181.

1.7.2. Rates of Antithrombin Inhibition of Thrombin and Factor Xa

The antithrombin inhibition of factor Xa and thrombin is a slow process. The typical uncatalyzed *in vitro* thrombin inhibition rate at pH 7.4 and 25 °C is in the range of $7–11\times10^3$ M^{-1} sec^{-1},[151-153] while for factor Xa, it is $2–3\times10^3$ M^{-1} sec^{-1}.[154] The slow rates of factor Xa and thrombin inhibition are dramatically increased in the presence of heparin. The second-order rate constant for thrombin inhibition by antithrombin–heparin complex is in the range of $1– 4\times10^7$ M^{-1} sec^{-1}, which is an acceleration of more than 2,000-fold, and for factor Xa inhibition this constant reaches 1.5×10^6 M^{-1} sec^{-1} representing an increase of ~600-fold.[155-158] Under physiological conditions, the acceleration of factor Xa inhibition may be even higher due to the presence of calcium ions.[159]

1.7.3. Mechanism of Heparin Activation of Antithrombin

A two-step induced-fit mechanism is involved in antithrombin activation by the high-affinity ligands heparin or heparin pentasaccharide (H_5). In the first step, heparin recognizes the native antithrombin (AT) and forms a low-affinity complex (AT:H) in rapid equilibrium, then is followed by a major conformational change to give a high-affinity complex (AT*:H).[160]

$$AT + H \underset{}{\overset{K_1}{\leftrightarrows}} AT:H \underset{K_2}{\overset{K_{-2}}{\leftrightarrows}} AT^*:H \qquad \text{Eq. (10)}$$

In equation 1, K_1 is the rapid equilibrium constant of first step of the induced-fit pathway; K_2 and K_{-2} are forward and reverse rate constants for the conformational change

step. The conformational change results in the expulsion of the partially inserted RCL residues. This significantly changes the conformation of the P1-P1′ reactive center loop and exposes an exosite in antithrombin (Figure 7).[150,154,157,158,161] This process is called the conformational activation of antithrombin. The altered RCL in heparin-antithrombin co-complex is better recognized by factor Xa resulting in accelerated cleavage of the P1-P1′ bond and rapid formation of the covalent inhibited complex (E*-AT* in Figure 6). Thus, conformational activation of antithrombin is necessary and sufficient for accelerated factor Xa inhibition.

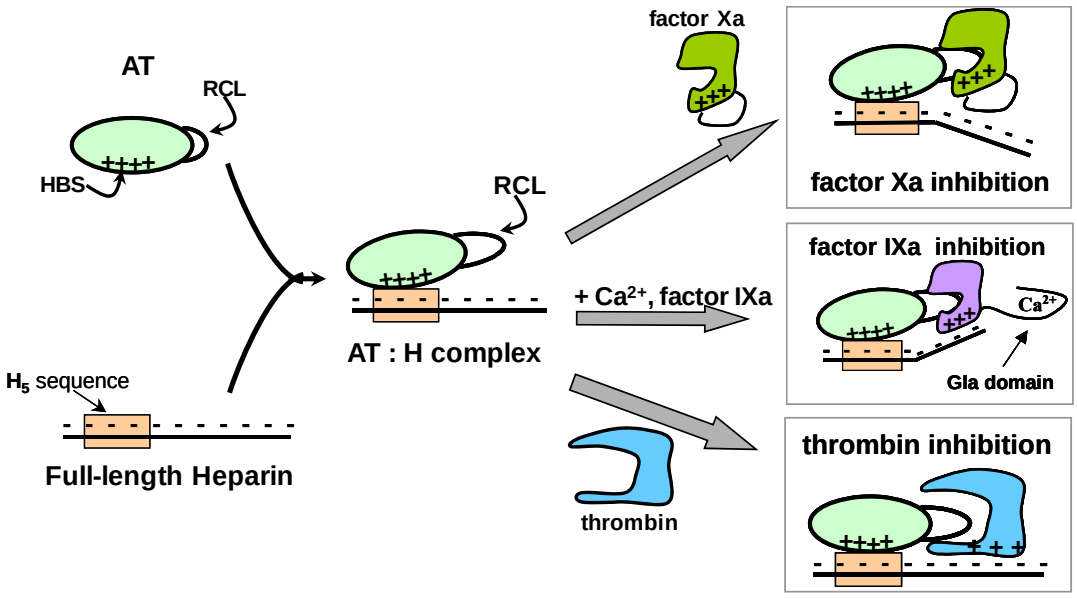


Figure 7. Two major mechanisms of heparin activation of AT and subsequent inhibition of factor Xa, factor IXa or thrombin: conformational activation and bridging mechanism. AT*:H = antithrombin–heparin complex; H_5 = high-affinity pentasaccharide sequence in heparin; RCL = reactive center loop; '+++' = exosite on enzyme; HBS = heparin-binding site. Reproduced from Desai, U. R. New antithrombin-based anticoagulants. *Med Res Rev* 2004, *24*, 151-181.

Thrombin inhibition is accelerated only two-fold through the conformational activation mechanism (Figure 7).[158] The acceleration of thrombin inhibition by heparin primarily arises from a bridging mechanism. The H_5 sequence in full length heparin tightly binds to antithrombin and is followed by the binding of thrombin to the same heparin chain at non-specific sites to form an antithrombin–heparin–thrombin ternary complex (Figure 7). Thrombin then diffuses along the polyanionic chain to encounter the inhibitor. This whole process results in a ~2,000-fold acceleration in inhibition under physiological conditions. A saccharide length of ~18 residues is needed to simultaneously hold thrombin and antithrombin for the accelerated inhibition.[162-163] A sequence-specific $\mathbf{H_5}$ is necessary for tight binding of heparin chains, but H_5 alone cannot potentiate antithrombin inhibition of thrombin.

1.7.4. Heparin Binding Site and Reactive Center Loop in Antithrombin

The heparin binding site in antithrombin is located about 20 Å away from the reactive center loop. This binding site specifically recognizes a heparin pentasaccharide sequence with high-affinity.[158] The heparin-binding domain in antithrombin is formed by positively charged residues of helices A and D, and the polypeptide N-terminus. The crystal structure of antithrombin pentasaccharide co-complex (PDB entry 1AZX, (Figure 8) shows that residues Arg47, Lys114, Lys125, and Arg129 in this region are commonly referred to as the pentasaccharide binding site (PBS), interact with H_5.[164] This observation is also consistent biochemical studies with antithrombin mutants.[165-169]

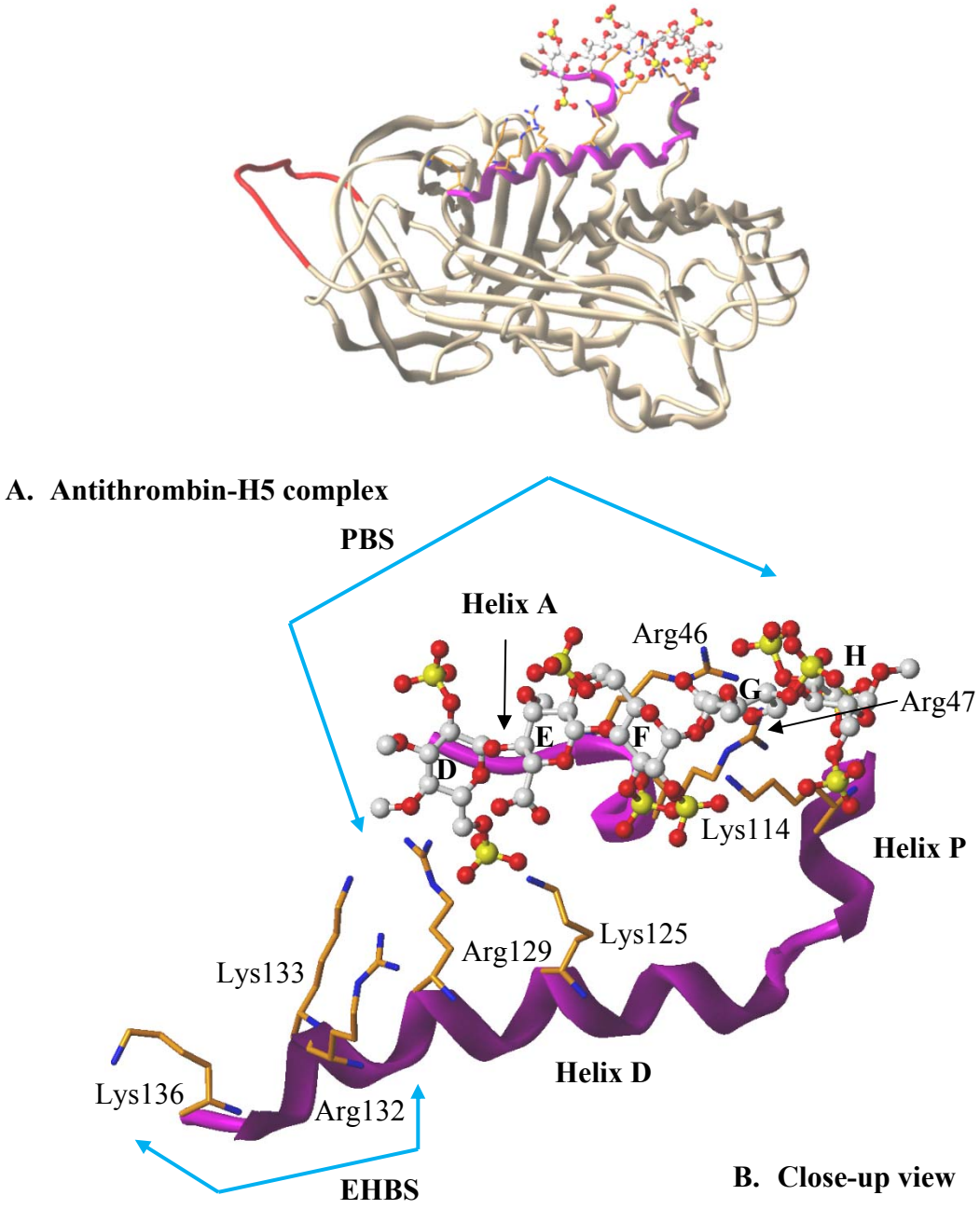


Figure 8. Ribbon diagram of plasma antithrombin complexed with natural pentasaccharide DEFGH (A) and a close-up view of the heparin-binding site (B). The crystal structure co-complex is obtained from PDB entry 1azx. Ribbon in red is the RCL, and magenta is the heparin-binding site. Pentasaccharide DEFGH is shown in ball-and-stick representation and individual residues are identified. Helices D, P and A form the heparin-binding site (HBS). Arg46, Arg47, Lys114, Lys125, and Arg129 form the pentasaccharide binding site (PBS) and Arg132, Lys133, and Lys136 form the extended heparin-binding site (EHBS).

A full-length heparin, in addition to interacting with PBS, also binds to an extended region formed by the basic residues Arg132, Lys133, and Lys136 at the C-terminal end of helix D (the EHBS).[170]

1.7.5. Heparin, LMWHs and Heparin Pentasaccharide DEFGH

Heparin is an anticoagulant that is a linear polysaccharide of varying chain length and is composed of uronic acid and glucosamine residues that are variably sulfated and acetylated.[158,171-172] Although heparin is a commonly used anticoagulant, its use as an anticoagulant is still limited by serious side effects such as excessive bleeding complications.[173]

In addition to bleeding complications, unfractionated heparin (UFH) and low molecular weight heparin (LMWH) suffer from significant intra- and inter-patient dose variability that requires laboratory monitoring. Heparin-induced thrombocytopenia (HIT) is a potentially lethal complication of heparin therapy that is associated with thrombosis. HIT refers to a significant drop in platelet count between 4 and 14 days after the initiation of therapy.[174]

Low Molecular Weight Heparins (LMWHs) is a better choice compared to heparin as anticoagulant.[175-177] LMWHs are much smaller in size (M_R 4,000–6,000) and are produced from heparin by chemical or enzymatic depolymerization. Ardeparin, dalteparin, enoxaparin, nadroparin, reviparin and tinzaparin are examples of available LMWHs.[178]

In comparison to heparin, LMWHs have greater bioavailability at low doses, better pharmacokinetics, and a more predictable dose response, which allows for fixed doses to

be administered without laboratory monitoring.[175-181] However, the risk of bleeding is not completely eliminated. In fact, there is not much difference between UFH and LMWHs when comparing preoperative hemorrhagic risk.[182-185] In addition, other concerns originating from the structural variations of LMWHs still need to be addressed. For example, different methods of preparation may introduce considerable variation in the *in vivo* efficacy among the LMWHs.[186] A report in 2007 revealed that heparin contaminated with oversulfated chondroitin sulfate (OSCS) was responsible for at least 81 reported deaths. Another concern arising from animal-derived heparin products due to the presence of potentially toxic substances that have a chemical structure similar to heparin.[187-192]

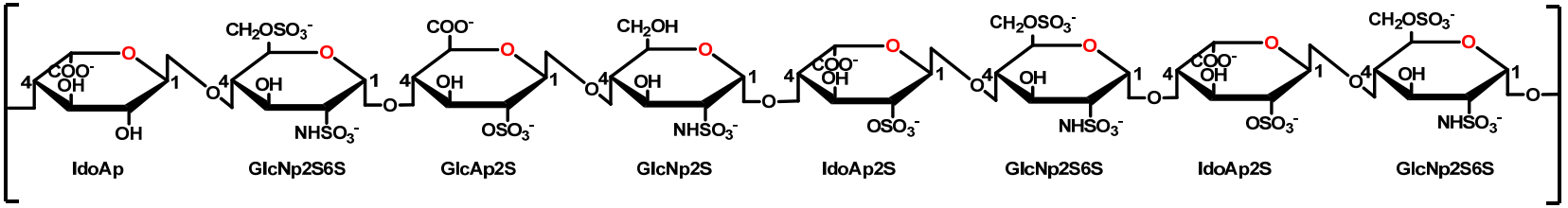
A sequence-specific heparin pentasaccharide-based compound, Fondaparinux, is perhaps the best anticoagulant in the heparin class due to the lack of HIT and a more predictable patient response. It also is not completely devoid of the risk of bleeding and lacks an effective antidote to reverse excessive iatrogenic bleeding.[193-198] A new pentasaccharide called idraparinux is being developed as a once-a-week injection, but because of its structural similarity to fondaparinux it is likely to carry a similar bleeding risk and antidote problem.[199-201]

Most of the side effect complications are attributed to the nature of structural complexity and diverse roles of heparin and heparan sulfate in a number of physiological and pathological processes. These side effects are reduced by using homogeneous heparin preparations such as fondaparinux, which is a synthetic five-residue sequence based on the naturally occurring DEFGH sequence (Figure 9).

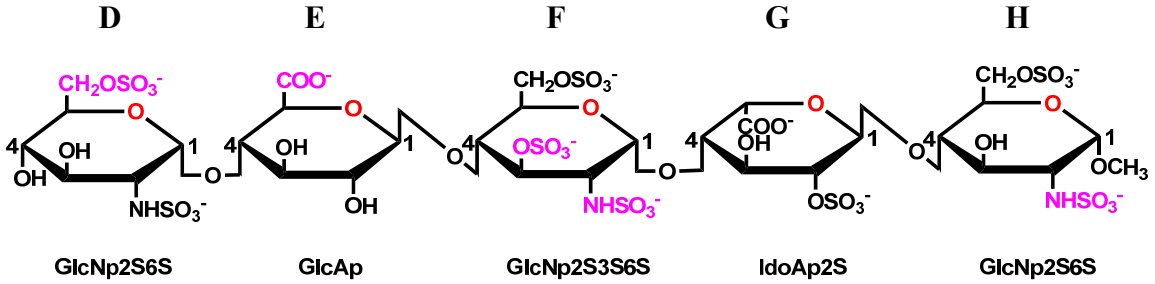
H_5 consists of three 2-*N*- and 6-*O*-sulfated glucosamines (β-D-GlcNp2S6S; residues D, F, H) interspersed with a 2-*O*- sulfated iduronic acid (α-L-IdoAp2S; residue G) and a glucuronic acid (β-D-GlcAp; residue E), in which the central glucosamine residue F has a unique 3-*O*-sulfate group (Figure 10). This sequence is abbreviated as GlcNp2S6S (1 → 4) IdoAp2S (1 → 4) GlcNp2S3S6S (1 → 4) GlcAp (1 → 4) GlcNp2S6S and labeled as DEFGH, in line with the history of its identification.[202-205]

Detailed structure-activity studies show that several anionic groups of DEFGH are critical for its high-affinity interaction with antithrombin (Figure 9).[206] More importantly, studies with truncated variants of DEFGH indicate that trisaccharide sequence DEF is the minimum structure that retains the functional role of heparin, i.e., antithrombin activation, *albeit* with significant loss of affinity under physiological conditions.[206]

The multidimensional limitations associated with heparin based anticoagulants warrants the need for the design of newer anticoagulants. For example, antithrombin (AT)-based anticoagulants including heparin, low molecular weight heparin (LMWH) and fondaparinux (FX)[207] are only effective when administered parenterally. Compared to heparin or LMWHs, results from clinical trials with fondaparinux are more promising with a low probability of causing HIT (Heparin Induced Thrombocytopenia) but its long term efficacy and safety is yet to be ascertained.[208] In addition, fondaparinux does not interact with protamine sulfate, the heparin antidote, making it difficult to manage drug-induced bleeding. However, the initial success of fondaparinux has validated factor Xa as a target for antithrombin-based anticoagulants.



Heparin or Low Molecular Weight Heparins



DEFGH Sequence

Figure 9. Structure of a heparin chain/LMWHs, and the sequence-specific heparin pentasaccharide DEFGH. Note the variations in the structure of glucosamine (GlcNp) and uronic acid residues (IdoAp or GlcAp). Numbers 1 and 4 refer to saccharide positions and the 1→4 linkages. Groups in magenta are critical for high affinity interaction with antithrombin.

1.8. Designing Synthetic Non-Sugar Antithrombin Activators

1.8.1. Hypothesis

Our research group has challenged the assumption that the saccharide-based skeleton is essential for the activity of specific glycosaminoglycan (GAG) sequences. We hypothesized that specific GAG sequences can be replaced by non-saccharide skeletons which may provide several advantages over the GAG skeleton. These include 1) ease of chemical synthesis, 2) likelihood of oral delivery due to enhanced hydrophobic character, 3) opportunity to gain additional non-ionic binding energy, 4) enhanced specificity for the target protein and 5) the ability to modulate responses in either an agonist or an antagonist manner. The following sections describe the methodologies used and the advances made in designing non-saccharide antithrombin activators.

1.8.2. First Generation Non-sugar Antithrombin Activators

Numerous attempts have been made to design or discover new molecules that activate antithrombin.[158,160,207] However, each of these searches has relied on utilizing a saccharide scaffold as a mimic of heparin. Implicit in these designs was the expectation that a saccharide scaffold was necessary to induce antithrombin activation.

Designing non-saccharide, antithrombin activators is not an easy task for some critical reasons. As mentioned earlier, antithrombin activation is a two-step induced-fit process that involves an initial recognition phase and followed by conformational transformation of AT.[160,206] Mimicking such a two-step process by design is challenging. Heparin mimics invariably must possess several negative charges, which can induce

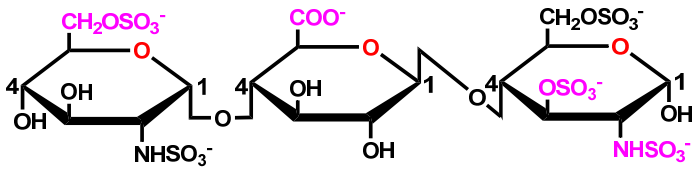
recognition of many electropositive domains on protein surfaces, leading to potential unwanted side effects. Mimicking heparin, or even pentasaccharide DEFGH (~20–25 Å long), by substituting organic scaffolds with multiple sulfate and carboxylate groups presents substantial design and synthetic challenges.[209]

Desai and co-workers have shown through structure-activity relationship studies that, while residues D, E, F, G, and H of the pentasaccharide sequence 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 10 A).[206] Thus, based on the trisaccharide DEF as a template, small, non-saccharide sulfated flavans were designed by Gunnarsson et al. using hydropathic interaction (HINT)[210] analysis (Figure 10B).

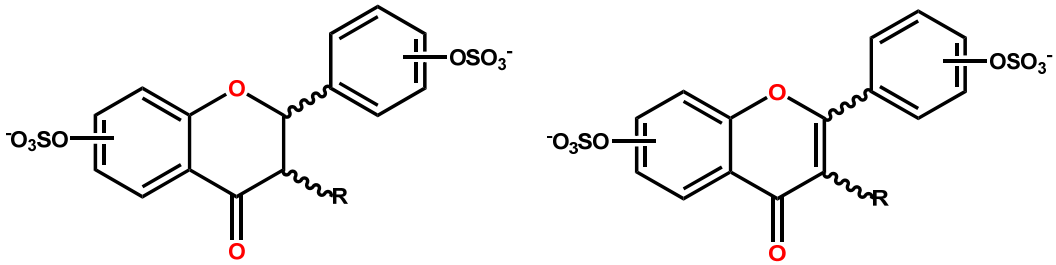
These designed molecules were found to be weak activators of the inhibitor AT (~10 fold).[134] The results indicated that sulfated flavans bind to AT with an affinity between 18 and 500 μM and bind in the adjoining extended heparin-binding site instead of the targeted pentasaccharide-binding site in AT.[135] 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 10B).

However, initial synthetic efforts towards sulfated flavones were unproductive because these non-saccharide designs possess significantly greater charge density than GAGs.[211] Hence, Gunnarsson et al. developed an alternate synthetic route to sulfated flavones that involved reductive cleavage of 2,2,2-trichloroethyloxysulfonyl-protected flavonoids (Figure 11).[212] Using this method, six sulfated flavones bearing different sulfate

group distributions were synthesized and evaluated. These sulfated flavans and flavones were also found to bind non-specifically to the smaller extended-heparin binding site in AT and resulted only in weak activation of antithrombin.



A. DEF sequence



B. Sulfated flavans and flavones

Figure 10. A) Structure of the DEF fragment. Critical groups for binding to antithrombin are shown in magenta. **B)** Non-saccharide DEF mimics.

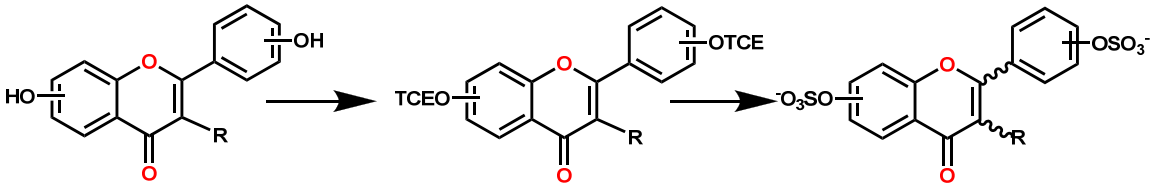


Figure 11. TCE-protection-deprotection strategy for the synthesis of sulfated flavones [TCE = 2,2,2–trichloroethyloxysulfonyl ($CCl_3CH_2OSO_2^-$)].

1.8.3. Tetrahydroisoquinoline-Based Lead Compounds

To improve on the antithrombin activation potential of these organic activators, a tetrahydroisoquinoline-based bicyclic-unicyclic sulfated activator IAS_5 (Figure 12) was designed using a pharmacophore-based approach (Figure 12). The pharmacophore was extracted from DEF and four critical groups were connected in three-dimensional space using a linear carbon linker to arrive at a first 'blueprint' of an activator.[209,213]

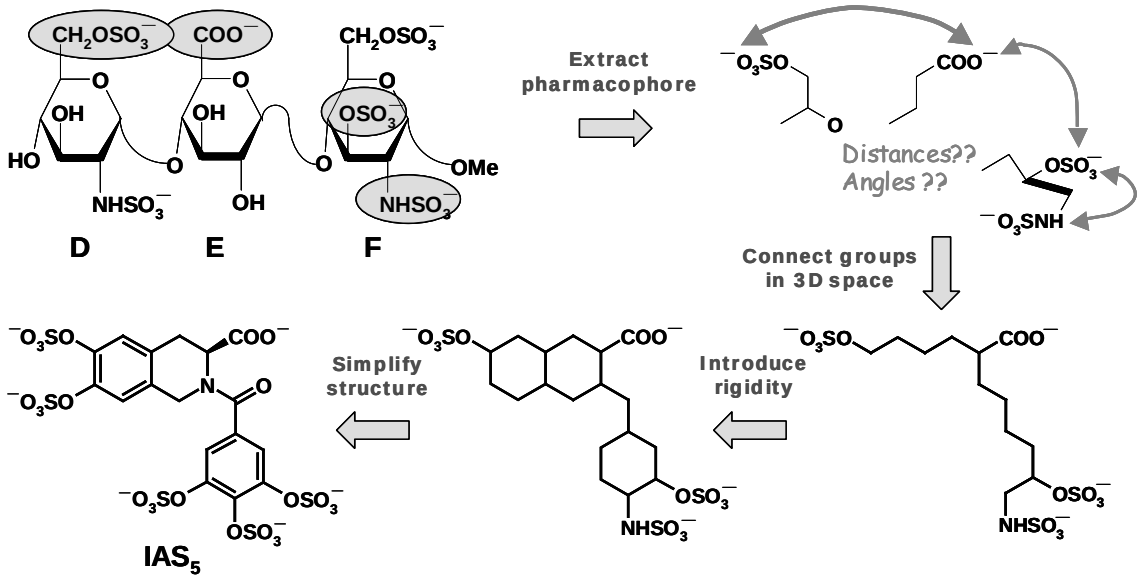


Figure 12. Rationale used in the design of tetrahydroisoquinoline-based organic activator IAS_5. Four critical anionic groups (highlighted as filled ovals) of trisaccharide DEF formed the pharmacophore. Connecting the groups using a carbon framework followed by engineering of rigidity that matches their orientation gave rise to a 'blueprint', which was transformed into a synthetically plausible target, IAS_5. Reproduced from Raghuraman, A.; Liang, A.; Krishnasamy, C.; Lauck, T.; Gunnarsson, G. T.; Desai, U. R. On designing non-saccharide, allosteric activators of antithrombin. *Eur J Med Chem* 2009, *44*, 2626-2631.

IAS_5, the tetrahydroisoquinoline-based activator containing an acid functionality and five sulfate groups was synthesized from commercially available precursors by Raghuraman et al.[209] These studies showed that IAS_5 activates antithrombin nearly 30-fold, an increase of nearly 2 to 3-fold higher than the first-generation rationally designed agents.[134,135]

The previously reported non-saccharide antithrombin (AT) activators in our laboratory, including sulfated flavonoids and sulfated tetrahydroisoquinoline derivatives, were found to bind the extended heparin-binding site (EHBS), although they were designed to target the pentasaccharide-binding site (PBS). The affinity capillary electrophoretic based competitive binding assays developed by our lab (Dr. Aiye Liang) showed that the tetrahydroisoquinoline-based molecules do not compete with a high affinity heparin pentasaccharide. In contrast, the affinity decreased dramatically in the presence of an extended heparin binding site ligand.[214]

The X-ray crystal structures of antithrombin alone[215,216] and in complex with heparins[217,218] 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.[219] 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[167-169] that 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.

1.8.4. Questions Remaining Unanswered

The previously reported non-saccharide sulfated flavans and tetrahydroisoquinoline-based molecules are binding at the EHBS of activated antithrombin and are weak activators of antithrombin. Thus, the questions that arise include: Can we design molecules that bind in the PBS of activated antithrombin with high affinity? Further, can we design molecules that interact with the PBS of activated antithrombin with high specificity? Chapter 2 describes the computational and experimental methodologies used to address these questions.

1.9. Desinging Specific Thrombin Exosite-II Modulators

As mentioned earlier, heparin is a complex carbohydrate biopolymer made of linear polysaccharides of varying chain length and is composed of uronic acid and glucosamine residues.[113] Although heparin is a commonly used anticoagulant it suffers from serious side effects such as excessive bleeding complications, heparin induced thrombocytopenia, and significant intra- and inter-patient dose response variability.[173,174]

Most of the side effect complications arise primarily from its structural complexity and its interaction with numerous proteins other than coagulation proteins factor Xa (fXa), thrombin, and antithrombin. Implicitly these diverse roles must arise from an optimal

combination of specificity and affinity. However with the exception of a few cases, the specificity of heparin interactions has been poorly understood and not explored in detail.

In antithrombin, three important amino acids are involved in heparin pentasaccharide binding, while in thrombin, at least seven basic amino acids are predicted to be involved. For biological systems, one would expect greater specificity with more interacting points. However, the heparin–thrombin system interestingly displays a lack of specificity. The molecular basis for this lack of specificity is not clear.

1.9.1. Questions Remaining Unanswered

The heparin binding sites (HBS) of antithrombin and thrombin are lined with Arg and Lys residues. Whereas the antithrombin-heparin interaction is specific, the thrombin-heparin interaction is considered to be non-specific. Thus, the questions that arise include: What is the structural and molecular basis for the specificity of the antithrombin-heparin interaction and the non-specificity of the thrombin-heparin interaction? Is it possible to design ligands that are specific for thrombin exosite-II?

Chapter 3 describes the structural biology and molecular modeling approaches used to study the specific and nonspecific interactions of heparin with antithrombin and thrombin. Also the mehodolgies used in designing specific thrombin exosite-II ligands using molecular docking techniques have been described.

CHAPTER 2

VIRTUAL SCREENING OF TETRAHYDROISOQUINOLINE SCAFFOLD AS ANTITHROMBIN ACTIVATORS

2.1. Virtual Screening of a Library of Tetrahydroisoquinolines

As mentioned in Chapter 1, the previously designed small, sulfated non-saccharide molecules are binding to the extended heparin binding site (EHBS) of antithrombin. Thus, they are poor activators of antithrombin. The lack of ability of the small, sulfated non-saccharide molecules to fully activate the antithrombin could be remediated by devising a credible non-saccharide design strategy that would facilitate the design of better non-saccharide anionic molecules with appropriate charge distribution with which to engage the the pentasaccharide binding site (PBS) amino acids residues Lys114, Lys125 and Arg129, and thus, turn the 'key' to open the 'lock'.

As mentioned in Chaper 1, the design of sulfated flavonoids and sulfated tetrahydroisoquinoline derivatives was based on the structure of trisaccharide DEF. It was time-consuming and user-biased because it involved manual pre-positioning of the ligand in the pentasaccharide-binding site in AT. This reduces the number of structural variations that could be effectively tested.

In addition, to increase our chances of targeting the PBS, we decided to use an approach that would categorize the ligands as either PBS-binding or EHBS-binding structures in AT (native and activated AT). Also, to explore a large number of

possibilities, such a method would have to be rapid and automated. Virtual screening (VS) was selected as the most appropriate method to achieve the aforementioned goals. VS is a computational technique that involves the rapid *in silico* assessment of large libraries of chemical structures in order to identify those structures most likely to bind to a drug target, typically the protein receptor or enzymes.

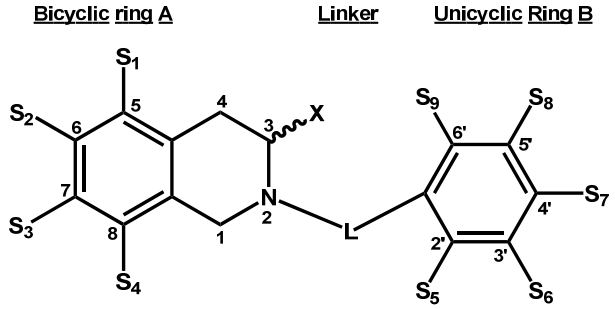
We have used the GOLD docking program for virtual screening of molecules bearing the tetrahydroisoquinoline scaffold with customized fitness functions to identify the hit molecules in an efficient manner, and also with high degree of reproducibility.

VS of the tetrahydroisoquinolines included the following components: i) constructing a library of tetrahydroisoquinoline based compounds using LEGION, ii) structural optimization and energy-minimization of the tetrahydroisoquinoline library for docking, iii) molecular docking of the library of compounds using the GOLD docking program onto activated and native antithrombin at the pentasaccharide biding site (PBS) and extended heparin binding site (EHBS), iv) identifying and analyzing the hit molecules and the core structural features proposed to be necessary for AT binding, v) analysis of protein-ligand interactions and their putative binding modes.

2.1.1. Designing Virtual Tetrahydroisoquinoline (ISOQ) Library

The success of combinatorial virtual library screening for glycosaminoglycans (GAGs) using GOLD in our lab suggested that it should be possible to use the previously reported protocol with modifications to study the tetrahydroisoquinoline based bicyclic-

unicyclic chemical structures for antithrombin activation. The core structure is derived from the tetrahydroisoquinoline-based compound IAS_5 (Figure 12). In order to find the appropriate position of sulfate group, substitutions were allowed at different positions. This library also consisted of 16 types of linkers (L) to study the effect of type and size of different linkers that connects the tetrahydroisoquinoline and phenyl ring system on antithrombin affinity. This combination of linkers and variation in the substitution pattern in the rings (S_1 to S_9 and X, Figure 13) resulted in a virtual library of 24576 compounds.



L: Linker, S_1 to S_9: H/-OSO_3^-,
X: H/COO^- (Both *R*- and *S*-isomers included)

–CO– (keto),	$–CH_2CH_2CH_2CH_2–$ (n-butylene)
–CONH–(amide),	$–CH_2CH_2CH_2CH_2CH_2–$ (n-pentylene)
$–CH_2–$ (methylene),	$–CH_2CH_2CH_2CH_2CH_2CH_2–$ (n-hexylene)
$–CH_2CH_2–$ (ethylene),	$–CH=CHCH_2CH_2–$ (1-butenyl)
$–CH_2CH_2CH_2–$ (n-propylene),	$–CH_2CH=CHCH_2–$ (2-butenyl)
–CH=CH– (ethenyl),	$–CH_2CH_2CH=CH–$ (3-butenyl)
$–CH=CHCH_2–$ (1-propenyl, allyl)	$–CH=CHCH_2CH_2CH_2–$ (1-pentenyl)
$–CH_2CH=CH–$ (2-propenyl, isoallyl),	$–CH_2CH_2CH_2CH=CH–$ (4-pentenyl)

Figure 13. Combinatorial virtual library based on IAS_5 with linkers containing 1 to 6 atoms. The library was created by variation of substituents S_1 to S_9, X, and the linker L.

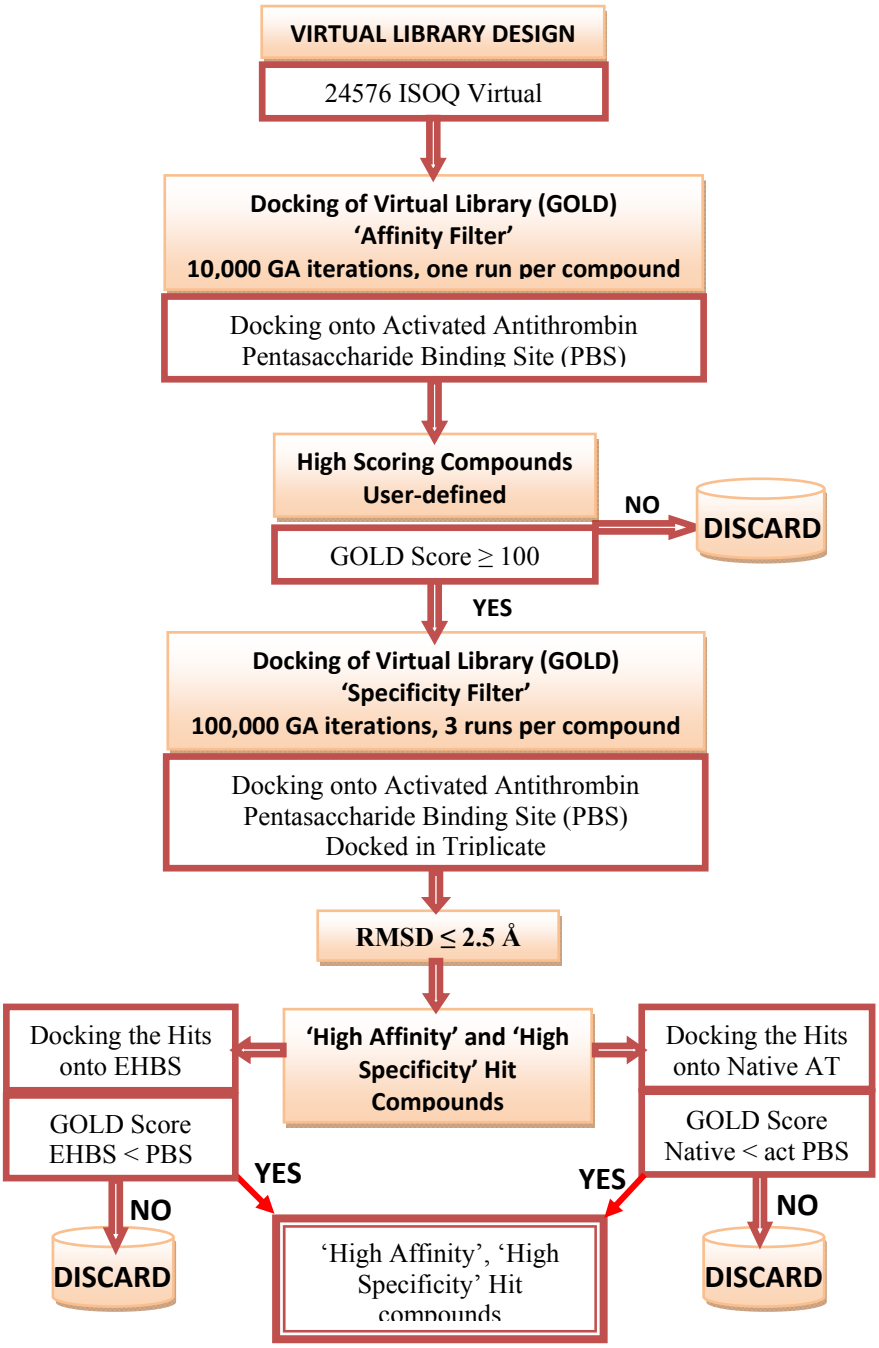


Figure 14. Docking protocol used to screen a combinatorial library of 24576 tetrahydroisoquinoline-based compounds. Color filled box: general dual-filter algorithm. (See text for details).

Molecular modeling of the combinatorial virtual library was performed using SYBYL (Tripos, St. Louis, MO). The small molecule non-saccharide library was built using the LEGION combinatorial library design module. LEGION rapidly generates 2D structures in SYBYL Line Notation (SLN) in a combinatorial manner. A plausible 3D geometry of the 2D structures generated by LEGION was obtained using CONCORD and these structures were subjected to energy minimizations (Section 2.9.2) in an automated manner using in-house SYBYL Programming Language (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. The protocol used for virtual screening of non-saccharide molecules was identical to the previously reported dual-filter algorithm used for saccharides, but with additional functions (Figure 14).[109]

2.2. Docking of ISOQ Library onto Activated Antithrombin (PBS and EHBS)

Previous studies in our lab on the docking of the H_5 sequence and its variants to the activated crystal structure of AT predicted that the modified GOLDScore (see experimental methods) of the H_5 sequence is 140 and its binding geometry was repeatedly predicted to within 2.5 Å in multiple docking experiments.[109] Thus, our search for small organic activators used these benchmarks as targets.

We docked these virtual 24576 molecules onto the activated form of antithrombin (PDB entry 1TB6) as mentioned in the docking protocol.[109,248] Following the affinity filter step, a normal Gaussian distribution of molecules was noted, as expected, with the majority showing poor binding to antithrombin (Figure 15).

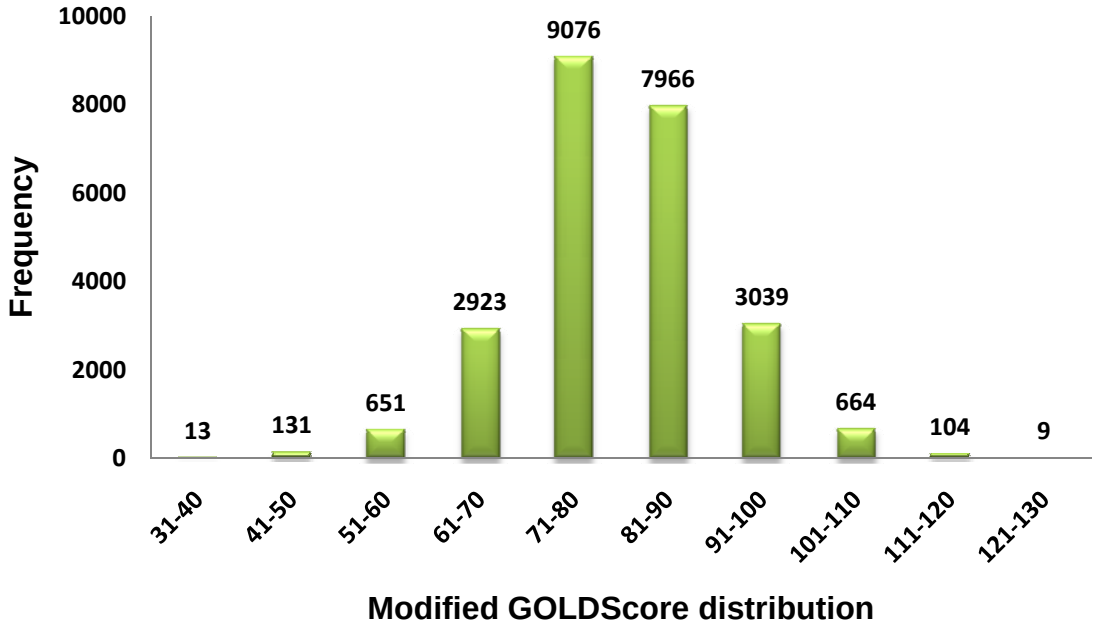


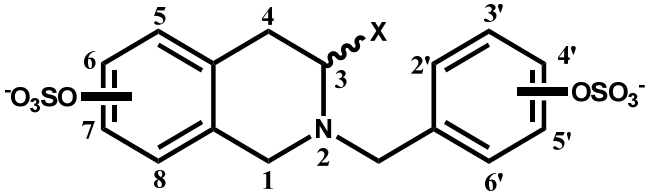
Figure 15. Histogram depicting the distribution of modified GOLDScore in 10 unit intervals for the 24576 tetrahydroisoquinoline compounds docked onto activated antithrombin PBS following the first phase (affinity filter) of the dual-filter screening algorithm.

The virtual 24576 molecules were docked to the pentasaccharide binding site (PBS) of antithrombin in affinity filter step (10,000 iterations). The molecules that had GOLD score greater than 100 (GOLD score for DEF = 115, while that for DEFGH = 140) were then docked in triplicate (100,000 iterations) onto the activated form of AT with 'specificity' as the second filter, where only 211 molecules passed the 2.5 Å cut-off filter (i.e., the top ranked 2 solutions were within 2.5 Å in each run). Of these 211 molecules, only 92 molecules passed the stringent test of specificity filter (The RMSD of the top ranked 2 solutions from each run and across runs is within 2.5 Å (Table 2: RMSD data for 92 hit compound, a measure of 'specificity' (At the end of this chapter with experimental

section). Thus, our combinatorial virtual screening results indicate that 92 molecules are likely to bind in the PBS of activated antithrombin with good affinity (Table 1).

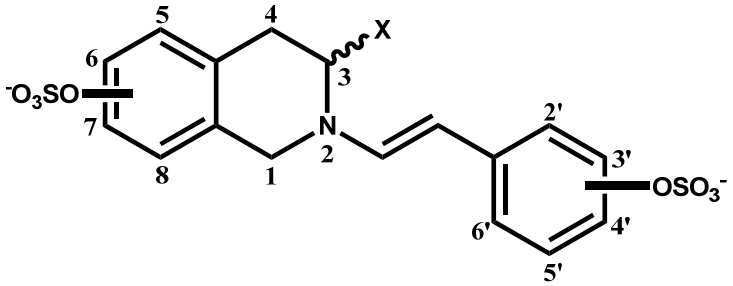
Table 1. Modified GOLDScore for 92 hit compounds and their structural information.

Compounds with 1-Carbon Linker



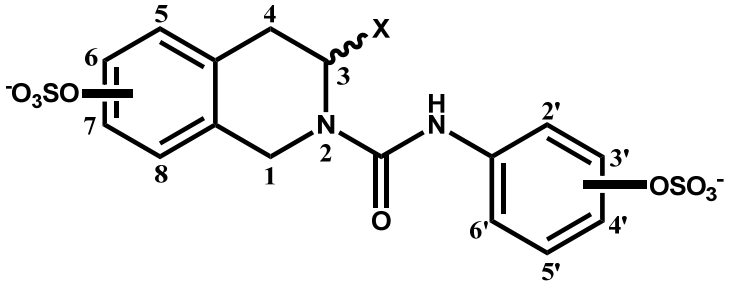
Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_6162	5, 6	−	2′,3′,6′	−5	108
ISOQ_R_6082	5,6	(*R*) COO^-	2′,4′	−5	104
ISOQ_S_6153	5,6	(*S*) COO^-	2′,3′	−5	107
ISOQ_S_6330	5,6,8	(*S*) COO^-	3′	−5	107

Compounds with 2-Carbon Linker



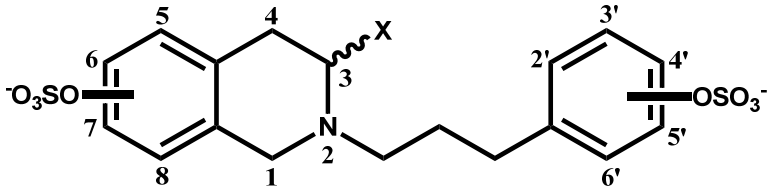
Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_S_1873	6,8	(*S*) COO^-	2′,3′,4′	−6	102
ISOQ_S_4255	5,8	(*S*) COO^-	2′,4′,5′	−6	117

Compounds with−CONH− Linker

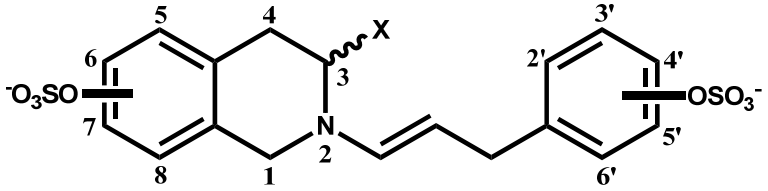


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_R_1059	7, 8	(*R*) COO^-	2′,3′,6′	−6	116
ISOQ_R_277	7	(*R*) COO^-	2′,4′,5′	−5	109
ISOQ_R_5183	5,7,8	(*R*) COO^-	2′	−5	109
ISOQ_R_5218	5,7,8	(*R*) COO^-	2′,3′	−5	119
ISOQ_R_5325	5,7,8	(*R*) COO^-	2′,5′	−6	121
ISOQ_S_2072	6,8	(*S*) COO^-	2′,5′	−5	108
ISOQ_S_6463	5,6,8	(*S*) COO^-	2′,5′	−6	127
ISOQ_S_6623	5,6,8	(*S*) COO^-	2′,4′	−6	117
ISOQ_S_7547	5,6,7,8	(*S*) COO^-	3′,4′	−7	121
ISOQ_S_7832	5,6,7,8	(*S*) COO^-	2′,4′,6′	−8	117

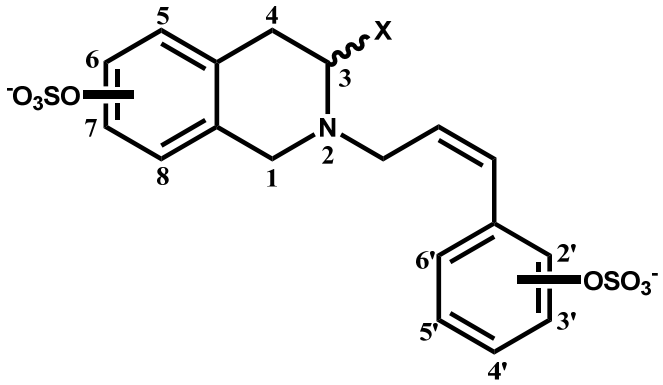
Compounds with 3-Carbon Linker



Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_5809	5,6	–	2′,4′	−4	113
ISOQ_R_1960	6,8	(*R*) COO^-	3′,4′	−5	105
ISOQ_S_6652	5,6,8	(*S*) COO^-	2′,5′	−6	120

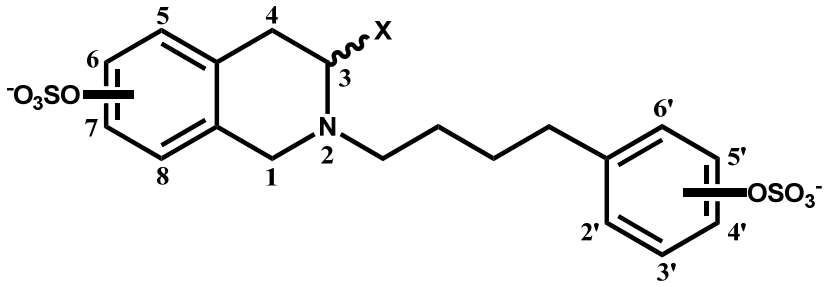


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_1903	6,8	–	2′,5′	−4	118
ISOQ_2561	6,7	–	3′,4′,5′	−5	109
ISOQ_6435	5,6,8	–	3′	−4	112
ISOQ_R_1875	6,8	(*R*) COO^{-}	2′,3′,4′	−6	111
ISOQ_R_2106	6,8	(*R*) COO^{-}	2′,4′	−5	106
ISOQ_R_5769	5,6	(*R*) COO^{-}	3′	−4	109
ISOQ_R_6018	5,6	(*R*) COO^{-}	2′	−4	107
ISOQ_R_6053	5,6	(*R*) COO^{-}	2′,5′	−5	117
ISOQ_R_6408	5,6,8	(*R*) COO^{-}	3′,4′	−6	120
ISOQ_S_6960	5,6,7	(*S*) COO^{-}	2′,5′	−6	114

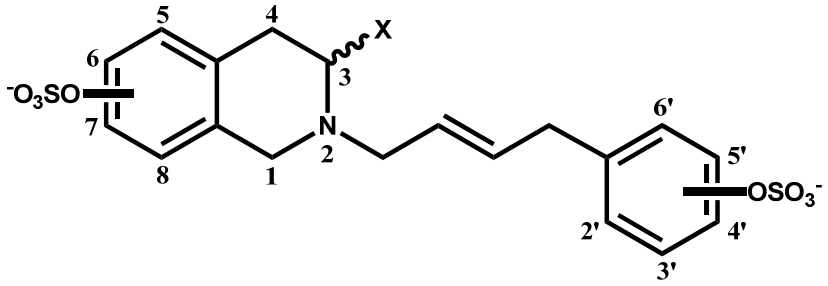


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_2096	6,8	–	2′,4′	−4	106
ISOQ_S_5963	5,6	(*S*) COO^{-}	2′,4′,5′	−6	117
ISOQ_S_6159	5,6	(*S*) COO^{-}	2′,3′	−5	108

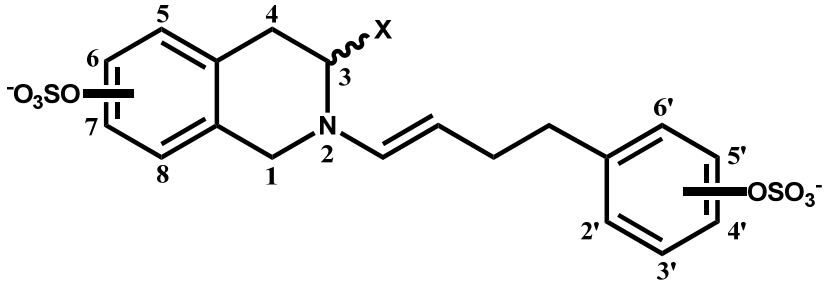
Compounds with 4-Carbon Linker



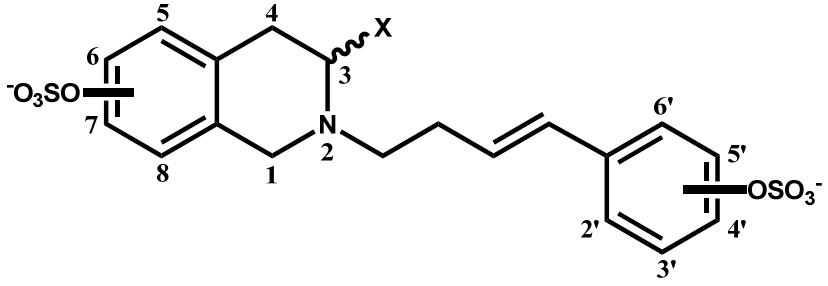
Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6462	5,6	(*S*) COO^-	2′,3′,6′	−6	124
ISOQ_EXT_5142	5,7	−	2′	−3	101
ISOQ_EXT_6294	5,6	−	2′,5′	−4	123
ISOQ_EXT_6550	5,6	−	2′,3′,6′	−5	123
ISOQ_EXT_R_3182	6,7	(*R*) COO^-	3′,4′	−5	117
ISOQ_EXT_R_6318	5,6	(*R*) COO^-	3′,5′	−5	117
ISOQ_EXT_R_6542	5,6	(*R*) COO^-	2′,3′	−5	121



Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_1229	7	(*S*) COO^-	3′,4′	−4	113
ISOQ_EXT_R_2861	6,8	(*R*) COO^-	2′,5′	−5	115
ISOQ_EXT_R_6253	5,6	(*R*) COO^-	3′,4′	−5	117

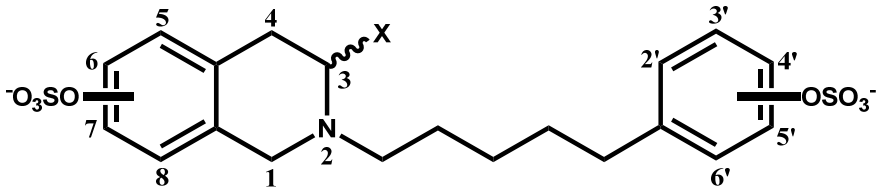


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6185	5,6	(*S*) COO^-	3′	−4	129
ISOQ_EXT_3201	6,7	–	3′	−3	108
ISOQ_EXT_6225	5,6	–	2′,4′	−4	110
ISOQ_EXT_R_6217	5,6	(*R*) COO^-	4′	−4	112
ISOQ_EXT_R_6281	5,6	(*R*) COO^-	3′	−4	117
ISOQ_EXT_R_6297	5,6	(*R*) COO^-	2′,5′	−5	120
ISOQ_EXT_R_6345	5,6	(*R*) COO^-	3′,4′	−5	126
ISOQ_EXT_R_6409	5,6	(*R*) COO^-	2′	−4	118
ISOQ_EXT_R_6473	5,6	(*R*) COO^-	2′,5′	−5	124
ISOQ_EXT_R_6857	5,6,8	(*R*) COO^-	3′,4′	−6	122
ISOQ_EXT_R_6921	5,6,8	(*R*) COO^-	2′	−5	123

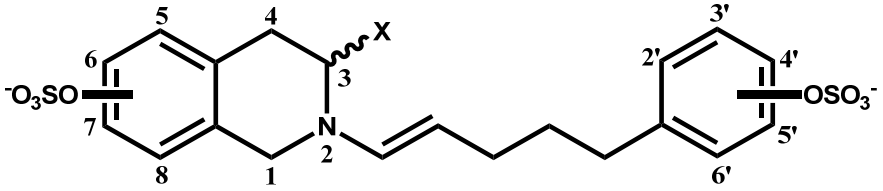


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_1243	7	(*S*) COO^-	2′,4′,5′	−5	118
ISOQ_EXT_S_6411	5,6	(*S*) COO^-	2′	−4	113
ISOQ_EXT_S_6971	5,6,8	(*S*) COO^-	2′,3′,6′	−7	138
ISOQ_EXT_S_6987	5,6,8	(*S*) COO^-	2′,4′	−6	132
ISOQ_EXT_6467	5,6	–	2′,4′	−4	110
ISOQ_EXT_979	8	–	2′,3′,4′,6′	−5	123
ISOQ_EXT_R_6411	5,6	(*R*) COO^-	2′	−4	121
ISOQ_EXT_R_6443	5,6	(*R*) COO^-	2′,5′	−5	127
ISOQ_EXT_R_6555	5,6	(*R*) COO^-	2′,3′,6′	−6	125
ISOQ_EXT_R_6955	5,6,8	(*R*) COO^-	2′,5′	−6	132
ISOQ_EXT_R_731	8	(*R*) COO^-	2′,3′,6′	−5	120

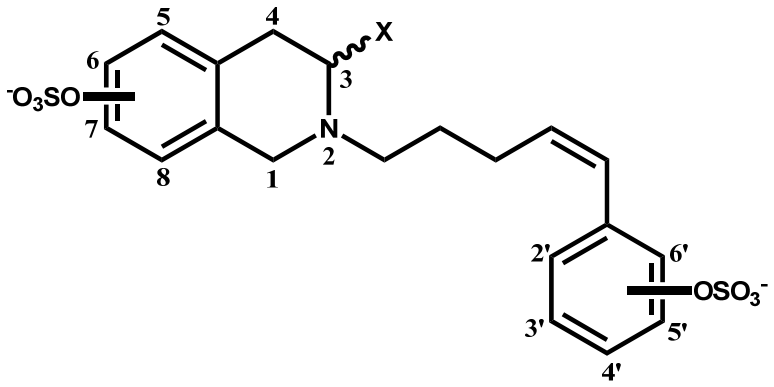
Compounds with 5-Carbon Linker



Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6815	5,6,8	(*S*) COO^{-}	2′,5′	−6	129
ISOQ_EXT_855	8	–	2′,4′,6′	−4	108
ISOQ_EXT_R_2591	6,8	(*R*) COO^{-}	2′	−4	110
ISOQ_EXT_R_3279	6,7	(*R*) COO^{-}	3′,4′	−5	121
ISOQ_EXT_R_6687	5,6,8	(*R*) COO^{-}	2′	−5	125

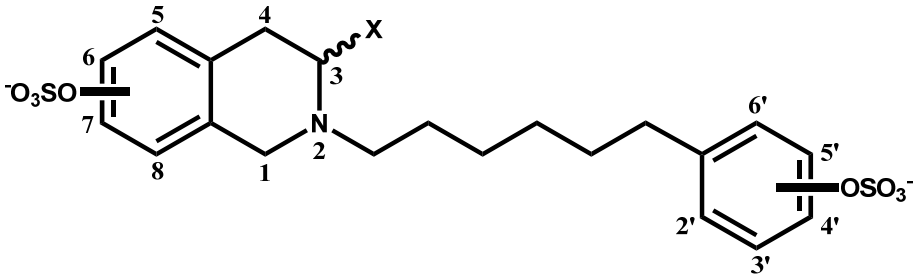


Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6234	5,6	(*S*) COO^{-}	2′,4′	−5	121
ISOQ_EXT_S_6410	5,6	(*S*) COO^{-}	2′	−4	105
ISOQ_EXT_2850	6,8	–	2′,5′	−4	121
ISOQ_EXT_2866	6,8	–	2′,3′,6′	−5	119
ISOQ_EXT_3906	6,7,8	–	2′,4′	−4	125
ISOQ_EXT_6402	5,6	–	2′	−3	107
ISOQ_EXT_6434	5,6	–	2′,5′	−4	118
ISOQ_EXT_706	8	–	3′,4′	−3	111
ISOQ_EXT_7938	5,6,7,8	–	2′	−5	112
ISOQ_EXT_R_6170	5,6	(*R*) COO^{-}	2′	−4	120
ISOQ_EXT_R_6474	5,6	(*R*) COO^{-}	2′,4′	−5	127
ISOQ_EXT_R_6538	5,6	(*R*) COO^{-}	2′,3′	−5	128



Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6236	6	(*S*) COO^-	2′,4′	−4	113
ISOQ_EXT_S_6988	5,6,8	(*S*) COO^-	2′,4′	−6	120
ISOQ_EXT_6532	5,6	–	2′,3′	−4	118
ISOQ_EXT_R_2956	6,8	(*R*) COO^-	2′,3′	−5	111
ISOQ_EXT_R_6444	5,6	(*R*) COO^-	2′,5′	−5	119
ISOQ_EXT_R_6732	5,6,8	(*R*) COO^-	4′	−4	106
ISOQ_EXT_R_6988	5,6,8	(*R*) COO^-	2′,4′	−6	131

Compounds with 6-Carbon Linker



Compound ID	*ISOQ Ring*	*Carboxylate Group*	*Phenyl Ring*	*Total Charge*	*Mod. GOLD Score*
ISOQ_EXT_S_6448	5,6	(*S*) COO^-	2′,5′	−5	123
ISOQ_EXT_7064	5,6,8	–	2′,3′,6′	−6	132
ISOQ_EXT_R_3408	6,7	(*R*) COO^-	2′,4′	−5	118
ISOQ_EXT_R_6224	5,6	(*R*) COO^-	4′	−4	109

2.2.1. Can these hit molecules preferentially bind to PBS and not to EHBS?

Experimental evidence clearly showed that (-)-epicatechin sulfate, one of the reported earlier lead compounds, has reasonably good binding affinity with antithrombin but weaker activation. It has been found that epicatechin and related molecules that were designed and synthesized in our lab are able to bind antithrombin tightly, but unable to activate is due to fact that they were binding in the EHBS of antithrombin and not PBS.

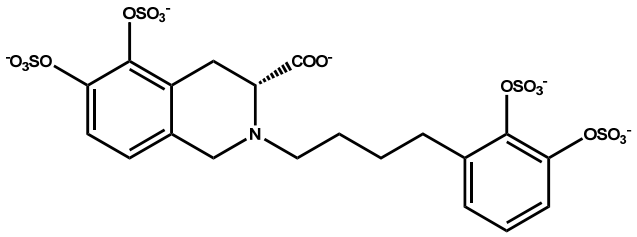
Docking results of these newly designed molecules (ISOQ-hits) in contrast are showing GOLD scores equivalent to DEF and DEFGH on PBS of activated antithrombin. In order to confirm that these compounds are indeed preferentially binding in the PBS but not in the EHBS, we have docked the hit molecules onto EHBS of activated antithrombin. The binding site has been predefined to include only the EHBS, so that GOLD docking of hit molecules is guided into the EHBS. Out of 92 hits that recognize the PBS consistently, 90 of them do not recognize the EHBS consistently in triplicate, indicating that our hit molecules are predicted to bind only the PBS and not the EHBS.

Only 2 molecules, ISOQ_5809 and ISOQ_EXT_R_6542 were predicted to have significant affinity for both binding sites. However, the comparison of the GOLD scores and binding geometries at PBS and EHBS of these 2 molecules (Figure 16 and Table 3) shows that these molecules also tend to be docked preferably at the PBS, and not to the EHBS.

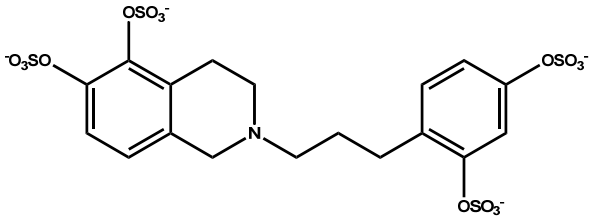
ISOQ_EXT_R_6542 is taken as a model for explaining the preferential binding to PBS and not to EHBS. Figure 19 show that the predicted binding geometry in the EHBS

makes minimal interactions, whereas the predicted geometry in PBS has the maximal number of interactions.

Table 3. Predicted interaction profile for hit ISOQ_EXT_R_6542 in PBS and EHBS



ISOQ-EXT_R_6542
GOLD Score for activated Antithrombin PBS = 121
GOLD Score for activated Antithrombin EHBS = 94



ISOQ_5809
GOLD Score for activaed Antithrombin PBS = 113
GOLD Score for activated Antithrombin EHBS = 87

Binding site	***Ligand group***	***Amino acid***	***Distance in Å***
PBS	2′-*O*-sulfate	Arg129	2.8
	3-Carboxylate	Lys125	3.5
	5-*O*-sulfate	Lys125	3.1
		Lys114	3.0
	6-*O*-sulfate	Lys114	3.0
		Arg47	2.7
		Arg46	3.1
EHBS	2′-*O*-sulfate	Arg132	2.5
	5-*O*-sulfate	Arg129	3.0
		Lys133	3.4
	6-*O*-sulfate	Lys125	2.7

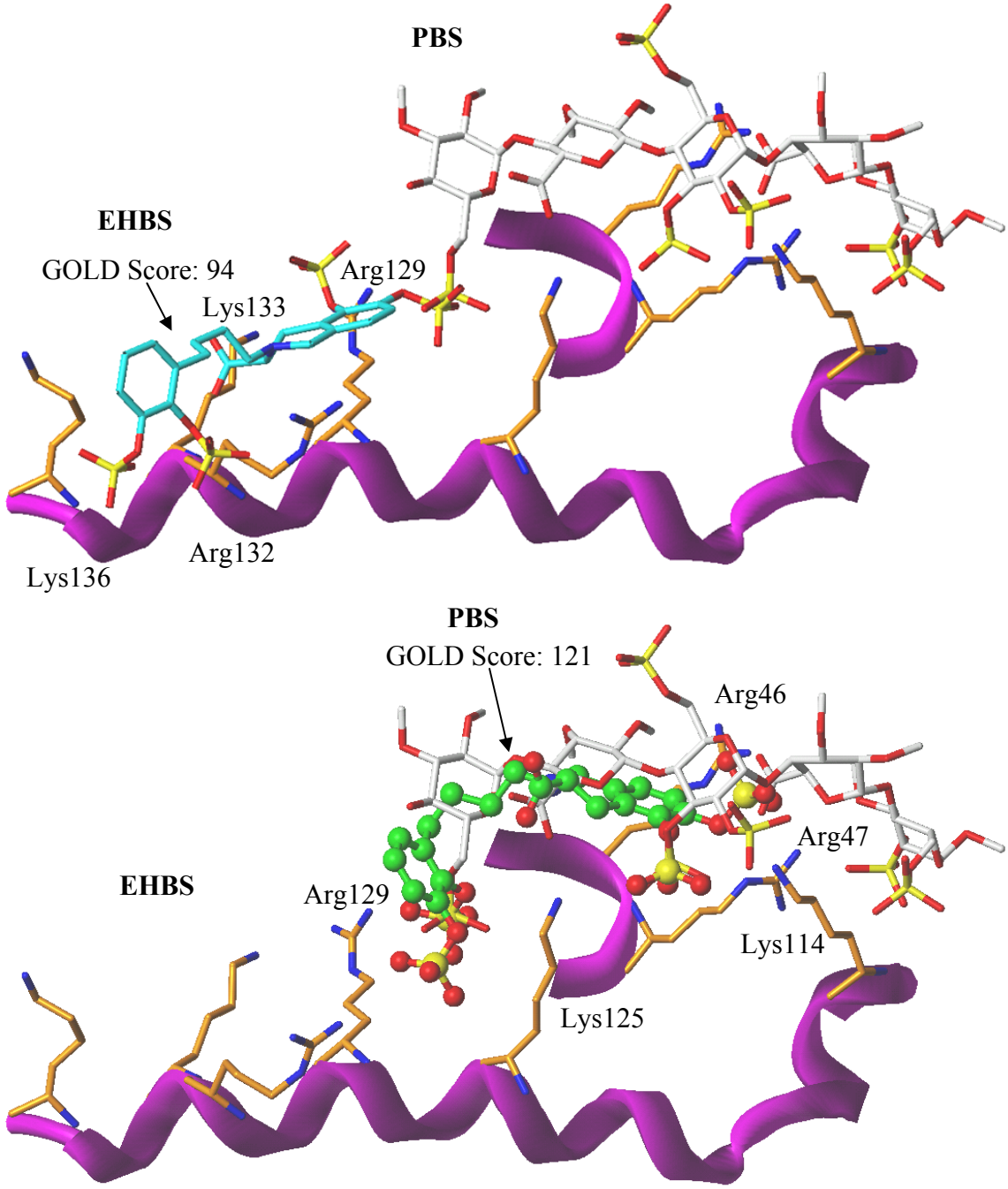


Figure 16. Predicted binding geometry for hit ISOQ_EXT_R_6542 in PBS and EHBS. A. Hit 6542 is shown in cyan stick form. B. Hit 6542 is shown green ball-and-stick form. In both figures, heparin binding amino acids are shown in capped stick form in orange and the helices A and D in magenta. The pentasaccharide sequence is shown in capped stick form in SYBYL atom type color, for reference.

This study predicts that indeed these compounds would preferentially bind at PBS and not at EHBS. This finding that our newly designed molecules predicted to be preferably binding at PBS, and not on EHBS, is a key in activation of antithrombin-mediated indirect fXa inhibition. As mentioned earlier, previous lead compounds from our lab (epicatechin and related compounds) are able to bind antithrombin tightly, but are weakly activating antithrombin; this is due to the binding of these molecules at EHBS, and not at PBS. Docking experiments of these newly designed molecules suggest that these hit molecules bind at PBS and are approximately equivalent to DEFGH in GOLD score.

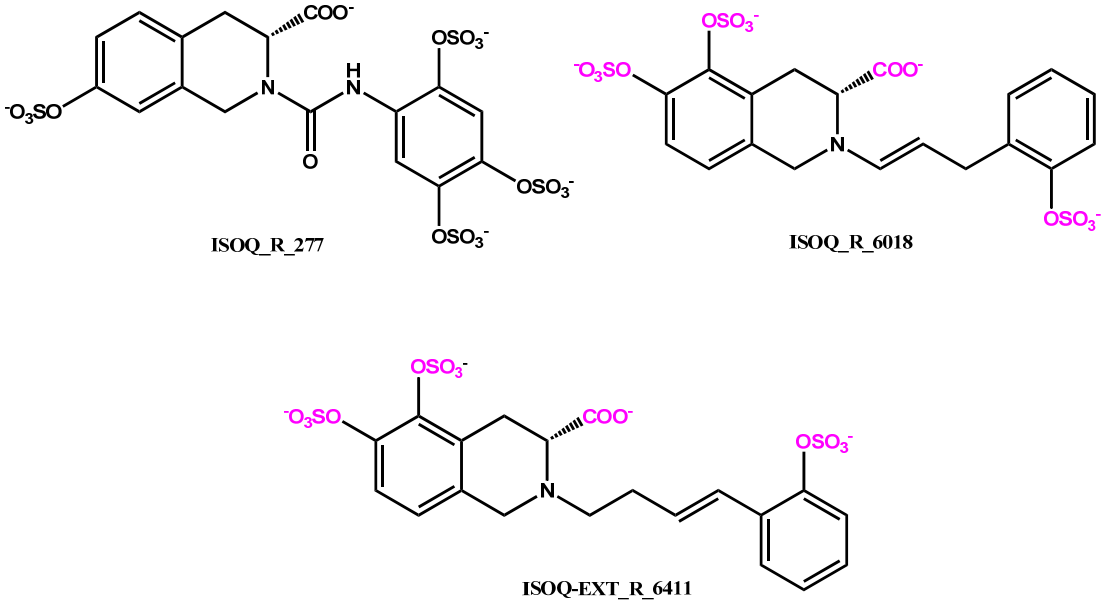
2.3. Docking onto Native Antithrombin

As discussed earlier, a two-step induced-fit mechanism is involved in antithrombin activation by the high-affinity heparin or a heparin pentasaccharide (H_5). In the first step, heparin recognizes the native antithrombin (AT) and forms a low-affinity complex (AT: H) in rapid equilibrium, then is followed by a major conformational change to give a high-affinity complex (AT*:H). The conformational change results in the expulsion of the partially inserted RCL residues. This is called conformational activation of antithrombin. The conformationally activated antithrombin recognizes the factor Xa better, thus increasing the effectiveness of the inhibition.

2.3.1. Can the ISOQ hits produce the necessary conformational changes in AT as H_5 does?

To predict whether or not this can happen, we docked our 92 hit compounds in triplicate onto the native antithrombin PBS (PDB entry 2ANT)[141] to see if they have any

preferential binding to the native form of antithrombin. These molecules were in general had lower GOLD scores when docked to native antithrombin. Also they were not docked consistently as in the PBS of activated form. This can be interpreted to mean that they are mimicking the initial recognition (low affinity complex formation) with the native antithrombin. Since the conversion of native antithrombin to activated antithrombin happens in rapid equilibrium, it is hard to say how much GOLD score would be required to bring about this change.



Compound	*GOLD score for native AT*	*GOLD score for activated AT*
ISOQ_R_277	93	109
ISOQ_R_6018	107	107
ISOQ_EXT_R_6411	92	121

Figure 17. Structures of ISOQ_R_277, ISOQ_R_6018 and ISOQ_EXT_R_6411 that are docked consistently to both native and activated forms of antithrombin in triplicate docking experiments. Groups in magenta shows the structural similarity between ISOQ_R_6018 and ISOQ_EXT_R6411.

Interestingly, there are 3 molecules of the 92 hit docked onto native antithrombin in triplicate experiments within 2.5 Å RMSD: ISOQ_R_277, ISOQ_R_6018 and ISOQ_EXT_R_6411 (Figure 17). Comparison of GOLD scores for ISOQ_R_6018 and ISOQ_EXT_R_6411 in native and activated antithrombin shows an interesting phenomenon. Both structures are structurally very similar, except that the hit 6411 has a 4-carbon linker and a double bond close to the phenyl ring, whereas the hit 6018 has a 3-carbon linker and a double bond close to the bicyclic ring.

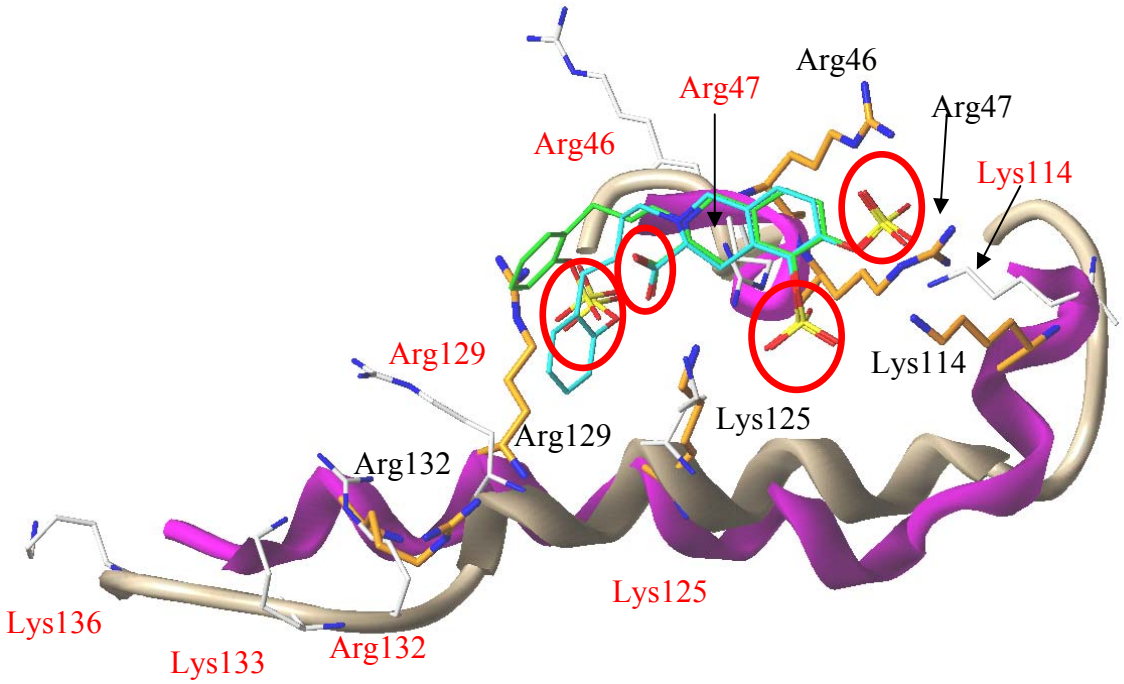


Figure 18. Predicted binding geometry for hit ISOQ_R_6018 and ISOQ_EXT_R_6411 in the PBS of activated antithrombin. Hit 6411 shown in cyan capped stick and hit 6018 in green. Activated antithrombin heparin-binding amino acids (labeled in black) are shown in capped stick form in orange and the helices A, D and P in magenta; whereas native form amino acids are shown in capped stick form in white (labeled in red), and the helices A and D in tan. The protein-ligand interactions are circled. Note the differences in orientation of heparin binding amino acids in native and activated antithrombin.

The predicted binding mode of both the compounds are similar as shown in Figure 18, except that the phenyl ring orientation is better for hit 6411, perhaps due to the π-bond effect in the phenyl ring end (3-ene) placing the ring system more appropriately to gain hydrophobic interaction with the long side chain of the spatially conserved Arg129 (see detail in Chapter 3, Section 3.4.1). This explain why the GOLD scores in the PBS of activated thrombin for the hit 6018 with a short linker (3-carbon) and the hit 6411 with an extended linker (4-carbon, thus the notation EXT for 4 or more carbon length) are 107 and 121, respectively.

Interestingly, the hit 6018 with a 3-carbon linker has the better GOLD score of 107 for native antithrombin than the hit with a 4-carbon linker that has the GOLD score of 92 for native, but a better profile for activated antithrombin. A closer look at the predicted binding geometries of these hit molecules show that the short linker hit 6018 interacts with Arg129, Lys125 and Arg47 of native antithrombin well within H-bond distances. However, the extended linker hit 6411 makes only a few interaction in the native antithrombin PBS (Figure 19A and 19B).

The results show that our hit molecules predicted to preferentially bind to activated antithrombin rather than native antithrombin. However, they are also predicted to have significant recognition capability for native antithrombin, which is very important for the two-step induced-fit mechanism of antithrombin activation. The compounds with a 4-carbon linker may be promising candidates for initial recognition of native antithrombin, and also to activate fully antithrombin to inhibit factor Xa.

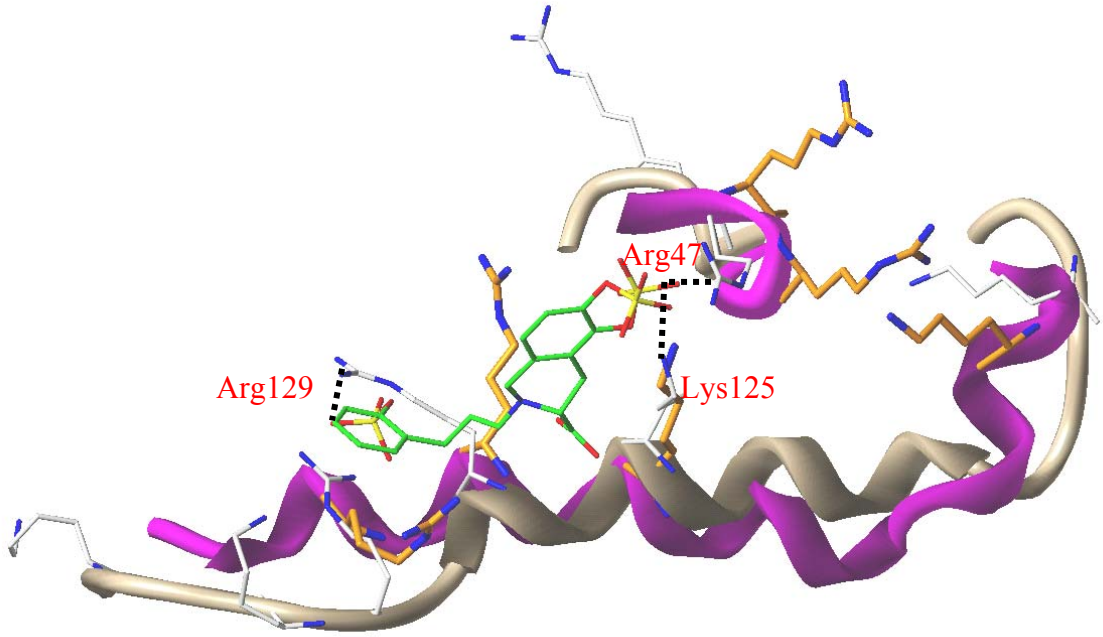


Figure 19. A. Predicted interaction of the short linker hit 6018 with native AT.

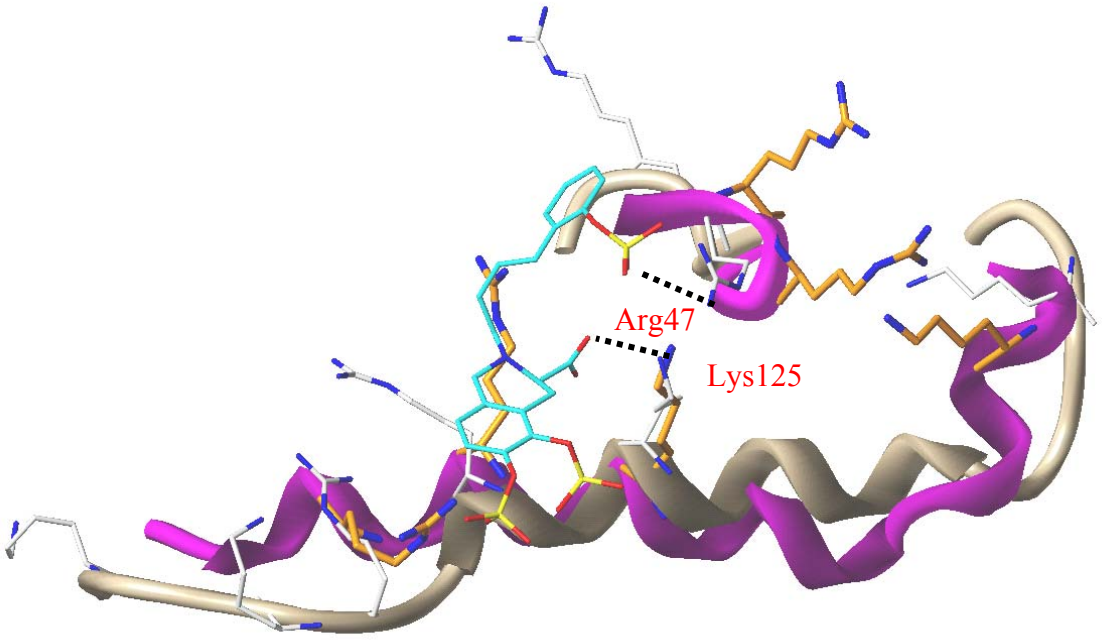


Figure 19. B. Predicted interaction of the extended linker hit 6411 with native AT.

2.4. Analysis of Predicted Binding Modes in the PBS of Activated Antithrombin

a. Linker types

In total, 16 types of linkers were explored that connect the bicyclic-unicyclic ring systems in the core structure (Figure 13). Structural analysis of the 92 hit molecules in the linker region showed the order of significance of the linkers as follows based on how many molecules had the particular linker type.

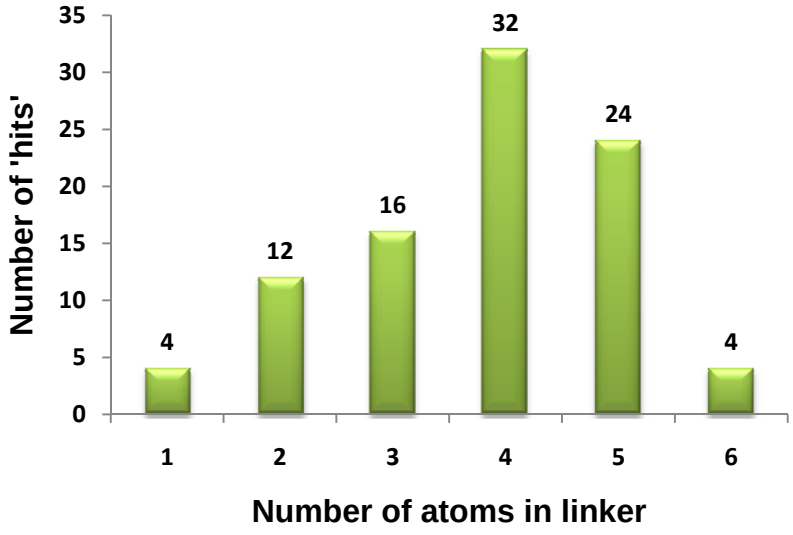


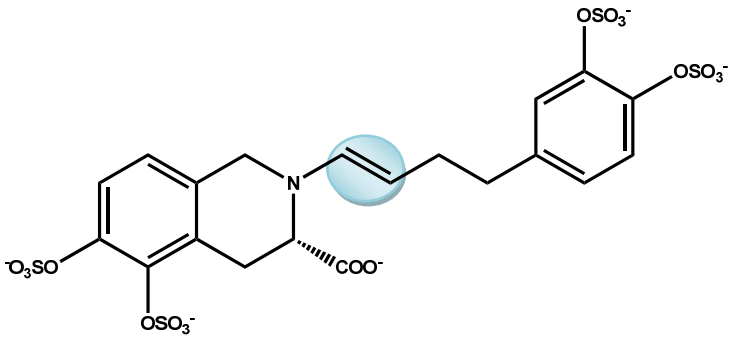
Figure 20. Histogram plot of the number of atoms in the linker for the 92 hits obtained from virtual screening of a non-saccharide sulfated library

It is predicted that a 4- or 5- carbon linker with a double bond at either end of the linker provides optimal geometries for these structures, allowing them to have strong interactions with antithrombin. Of the 92 hits, 32 had an all-carbon 4-atom linker, while 24 had an all-carbon 5-atom linker (Figure 20). Of the 56 hits that had a 4- or 5-carbon linker, 34 had as (*E)* double bond configuration with a strong preference to be alpha to either the

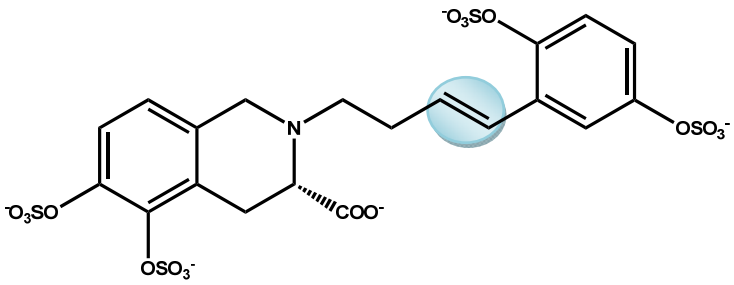
nitrogen atom or the phenyl ring (Figure 21) while 10 contained a (*Z)* double bond configuration alpha to the phenyl ring system. The remaining 12 contained a fully saturated linker.

The order of preference for the linkers is as follows:

-CH=CHCH$_2$CH$_2$CH$_2$- (12) > -CH=CH$_2$CH$_2$CH$_2$- (11) = -CH$_2$CH$_2$CH=CH- (11) > -CH=CHCH$_2$- (10) > -CONH- (10) > CH$_2$CH$_2$CH$_2$CH$_2$- (7) > -CH$_2$CH$_2$CH$_2$CH=CH- (7) > -CH$_2$CH$_2$CH$_2$CH$_2$CH$_2$- (5) > -CH$_2$CH$_2$CH$_2$CH$_2$CH$_2$CH$_2$- (4) = -CH$_2$- (4) > -CH$_2$CH=CHCH$_2$- (3)> -CH$_2$CH$_2$CH$_2$- (3) CH$_2$CH=CH$_2$ (3) > CH=CH- (2).



ISOQ_EXT_R_6345



ISOQ_EXT_R_6443

Figure 21. The hit structures with 4-carbon linkers showing double bond positions either close to tetrahydroisoquinoline ring (6345) or phenyl ring system (6443).

Of 25 molecules that have a double bond in the 4 carbon linker, 22 of them have the double bond either close to isoquinoline ring or phenyl ring system (1-ene or 3-ene). This indicates that the restricted rotation around the bulky ring system may help in orientation of negatively charged functional groups in the binding site (Figure 21).

b. Importance of 3-carboxylate group

The 3-carboxylate in the tetrahydroisoquinoline ring may form a hydrogen bond interaction with Lys125, provided that other sulfate groups in the tetrahydroisoquinoline ring are appropriately positioned. Of 92 hit molecules, 69 (75%) of them has the COO^- group at position 3 of tetrahydroisoquinoline ring. The large percentage of hits containing the 3-carboxylate indicates that substitution at this position with a negatively charged group is of great significance.

The protein-ligand interaction analysis shows that out of 69 molecules with COO^- group, 43 of them are (*R*)-isomer and 26 are (*S*)-isomer. Though statistically the (*R*)-isomer is preferred, (*S*)-isomers also have equally high GOLD scores (Table 1), indicating that either isomer should be fine in that position, depending on other substitutions.

For example, the unique feature that facilitates the COO^- interaction with Lys125 is the presence of 5-*O*-sulfate and 6-*O*-sulfate in the tetrahydroisoquinoline ring. In particular, if the structure has 6, 7 or 6, 8 position sulfates, it is the 7-*O*-sulfate or 8-*O*-sulfate that interacts with Lys125 and not the 3-carboxylate (Figure 22).

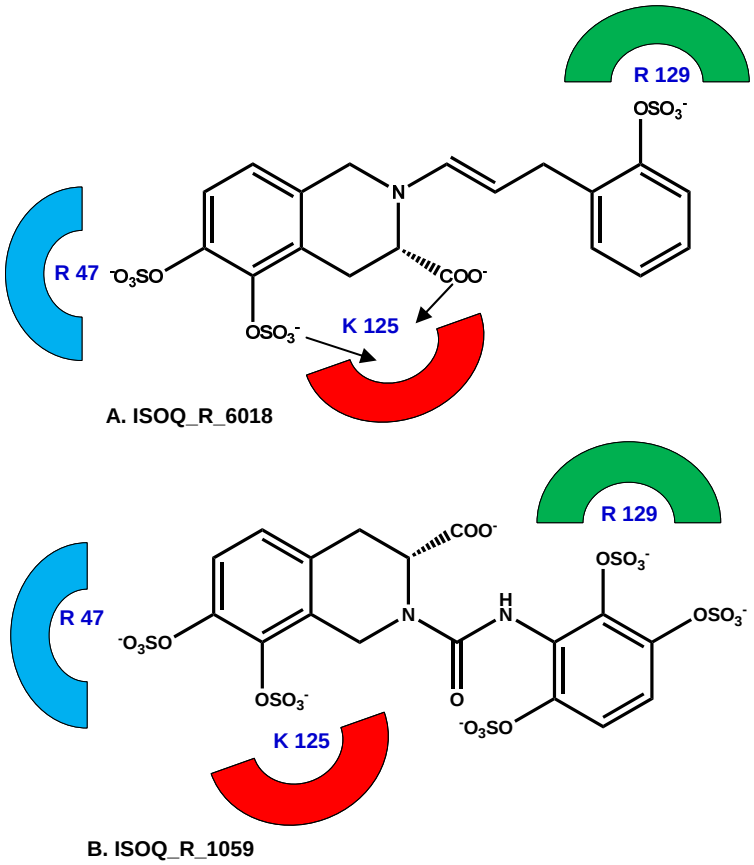


Figure 22. ISOQ_R_6018 shows double hydrogen bond with Lys125 through the 5-*O*-sulfate and 3- carboxylate, but ISOQ_R_1059 hydrogen bonds with Lys125 through the 8-*O*-sulfate.

It is predicted that the presence of 5-*O*-sulfate and 6-*O*-sulfates in the ligand structure helps Lys125 of antithrombin to form a double hydrogen bond, which makes the ligand a better candidate for affinity (Figure 23). This distribution favors the molecule to bind in such a way that facilitates the interaction of 3-carboxylate to Lys125. This is also the reason why other compounds do not interact with Lys125. The presence of sulfate groups only at the 6- and 8-, or 7-and 8- positions in the tetrahydroisoquinoline ring system changes the face of the molecule in the opposite way such that the 8-*O*-sulfate interacts

with Lys125 rather than the COO^{-}. There is a mutual competition between the 8-*O*-sulfate group and the 3-COO^{-} for Lys125, which is determined by the distribution of the rest of the sulfate groups in the system.

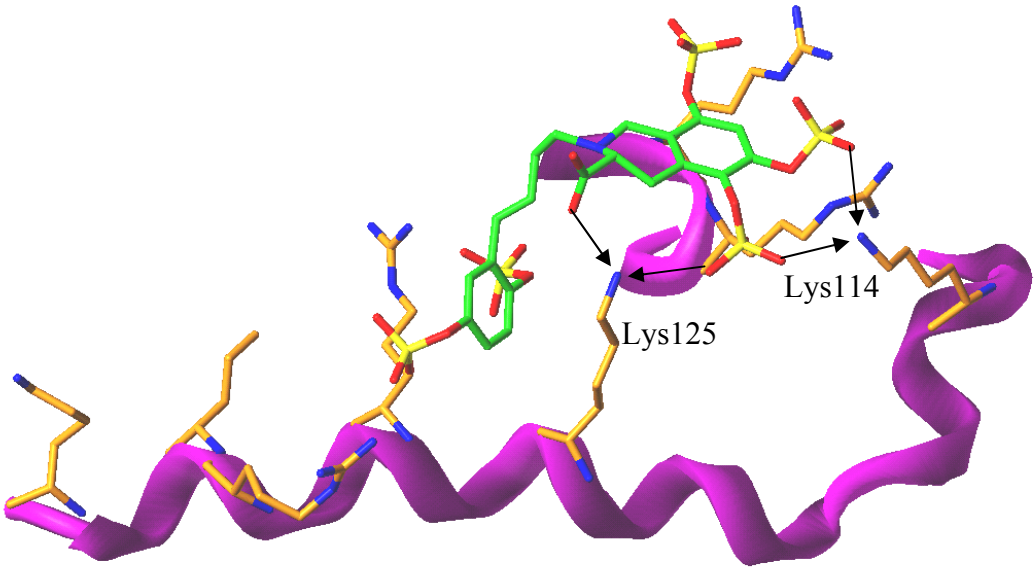


Figure 23. Predicted binding geometry for hit ISOQ_R_6955 in the PBS of activated antithrombin showing the influence of the 5- and 6-*O*-sulfates on the 3-carboxylate interaction. The capped stick form of 'hit' 6955 is shown in green and sulfate groups in SYBYL atom type. Activated antithrombin heparin binding amino acids are shown in capped stick form in orange and the helices A, D and P in magenta. Note that the 3-carboxylate and 5- and 6-*O*-sulfates are in the same direction, whereas the 8-*O*-sulfate faces in the opposite direction.

c. Distribution of sulfate groups determines the binding mode

Overall, it is the distribution of sulfate groups in the core structure that make hydrogen bond and ionic interactions with basic amino acids that determines whether a molecule is going to bind and activate antithrombin or not. Although it looks simple, there

is a significant preference for the presence of sulfate group in certain positions over others. Structural information analysis of GOLD-docked dual filtered hit molecules reveals that certain patterns of sulfate group substitution occur present more often than others. The average number of sulfate groups in the tetrahydroisoquinoline and phenyl ring system that is necessary is 2.

The frequency of charges at various positions of the core structure is shown in the histogram (Figure 24), indicating the importance of charged functionalities at different positions.

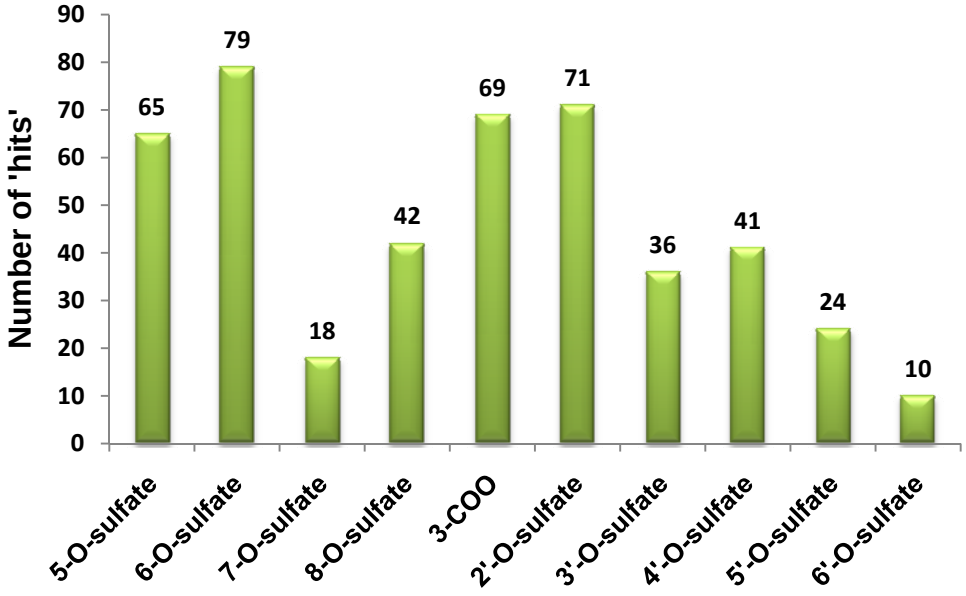


Figure 24. Histogram plot of the frequency of charged groups (carboxylate and sulfate) in various positions of the 92 hit compounds obtained from virtual screening of a non-saccharide sulfated library.

In Figure 25, the important positions are shown in color-filled ovals with their respective percent sulfation among the 92 hit molecules. With respect to number and distribution of negative charges, the majority of hits contained 4 to 6 negative charges.

Most interestingly, 48 of the 92 hits contained a 3,5,6-trisubstitutedtetrahydro isoquinoline fragment (Figure 26) suggesting that sulfate groups at the 5- and 6-position of the tetrahydro isoquinoline ring favorably contribute towards affinity .

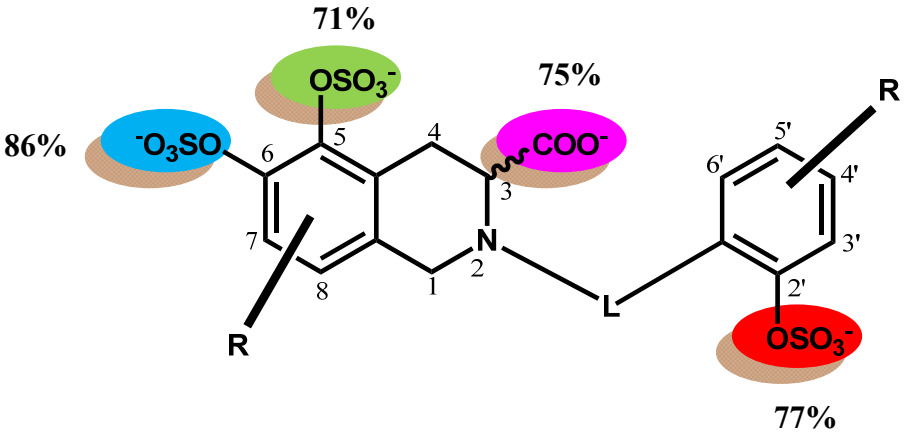
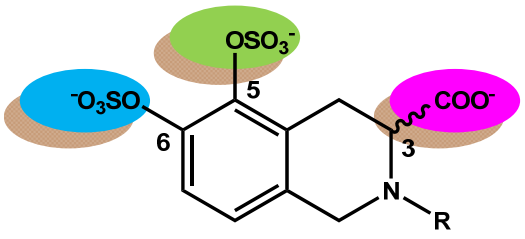


Figure 25. Core structure showing the putative pharmacophore (colored ovals) for a series of ISOQ analogs binding in the pentasaccharide binding site of activated antithrombin (L = Linker size of 4- to 5-carbon length, R = $–OSO_3^-$).



3,5,6-trisubstituted tetrahydroisoquinoline fragment

Figure 26. Key pharmacophore of the bicyclic ring system (referred as fragment I for discussion purpose in the text) identified by docking to the PBS.

The number of sulfate groups in the ring systems predicts the binding mode to be one of two poses. The following example highlights the uniqueness of the distribution of the sulfate groups and their impact in determining binding mode. Molecule ISOQ_R_277 is binding to activated antithrombin at PBS just exactly opposite direction compared to molecule ISOQ_R_5183 (Figure 27).

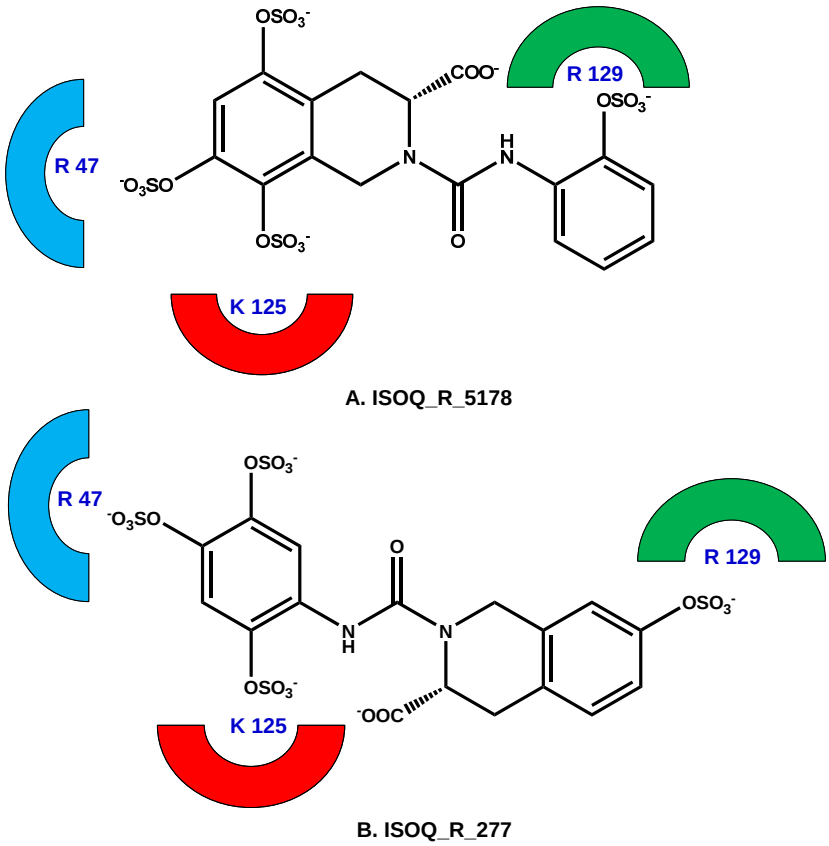


Figure 27. ISOQ _R_5178 and ISOQ_R_277 interactions with key basic amino acids showing a complete reversal in its binding mode based on sulfate group distribution.

The two molecules hit 5178 and hit 277 are essentially the same in its core structure except in the distribution of sulfate groups. Hit 5183 has 5-, 7-, and 8-*O*-sulfate groups in the tetrahydroisoquinoline ring system and the 2′-*O*-sulfate in phenyl ring system whereas hit 277 has 2′-, 4′-, and 5′-*O*-sulfates in phenyl ring and 7-*O*-sulfate in tetrahydroisoquinoline ring. This results in a complete predicted reversal of binding mode.

Furthermore, this kind of 'flip' is only observed when the tetrahydroisoquinoline ring has one sulfate group and phenyl ring has more than 1 or 2, but if the number of sulfate groups in the tetrahydroisoquinoline ring becomes 2 then the tetrahydroisoquinoline stays in its preferred original binding mode even if the phenyl ring has 3 sulfate groups. This prediction indicates that the number of sulfate group in the tetrahydroisoquinoline ring predominantly determines the binding mode.

d. Predicted Binding Geometry of the Representative Hit Molecule

Hit molecule ISOQ_EXT_R_6955 represents one of the high-scoring favorable molecules that have all of the above mentioned structural features. Hit 6955 contain a 4-carbon linker with a double bond alpha to the phenyl ring. 3-carboxylate, 5-, 6-, and 8-*O*-sulfates in the bicyclic ring, and 2′-, and 5′-*O*-sulfates in the phenyl ring system. Since this molecule represents the common binding mode among most of the hit molecules, we will discuss in detail its interactions with activated antithrombin at PBS (Figure 28 and Table 4).

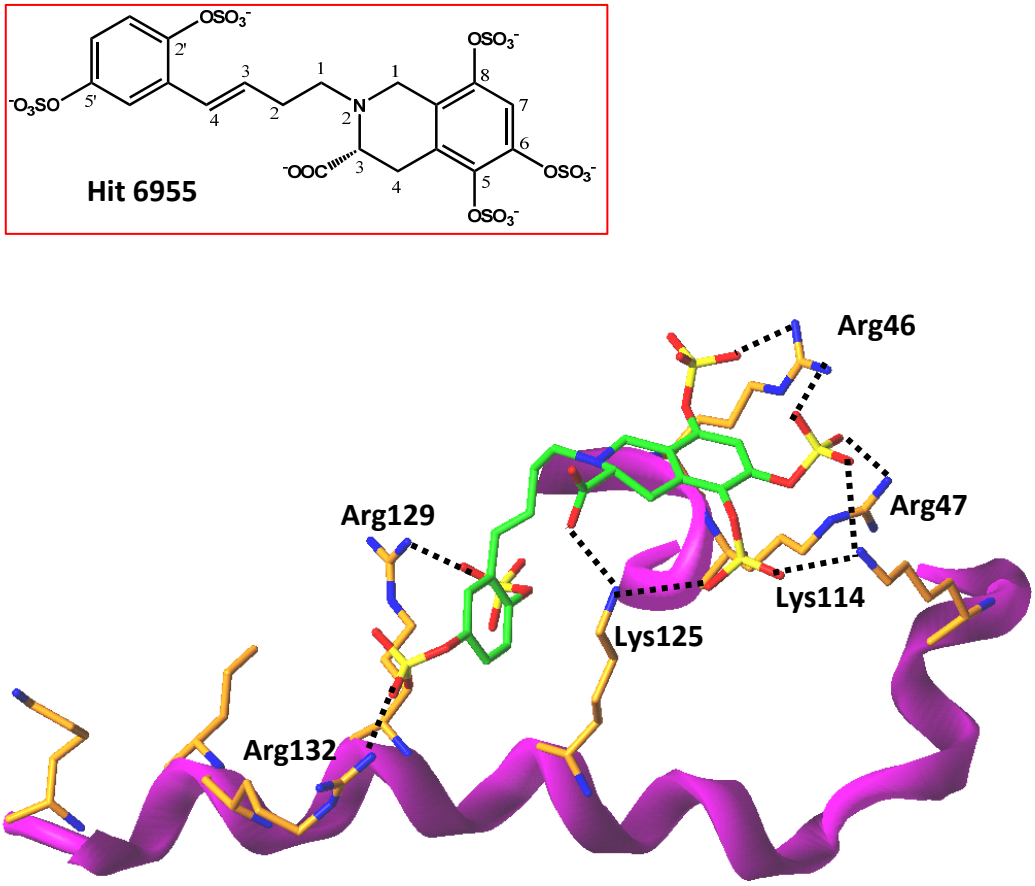


Figure 28. Predicted binding geometry for hit ISOQ_R_6955 in the PBS of activated antithrombin. The capped stick form of hit 6955 is shown in green and sulfate groups in SYBYL atom type. Activated antithrombin heparin-binding amino acids are shown in capped stick form in orange and the helices A, D and P in magenta. The interactions are shown as dotted lines between ligand and amino acids.

Analysis of protein-ligand interaction predicted that the 3-COO^- to interact with Lys125, the 5-*O*-sulfate with Lys125 and Lys114, and the 8-*O*-sulfate with Arg46. The 6-*O*-sulfate is predicted to interact with Arg46, Arg47 and/or Lys114. Its position is optimal

in that it can anchor itself in the middle of Lys114, Arg46 and Arg47. It is interesting to see that 79 of 92 hit molecules (86%) have a 6-*O*-sulfate, which in all cases interact with Arg46, and Arg47/or Lys 114. This implies that this group may preferentially recognize a specific feature in the protein. The 6-*O*-sulfate and other sulfate groups H-bond specifically and exhibit a preferential binding mode. Analysis of sulfate group distribution also reveals that with some preferable combinations it is possible to attain positive interactions with critical basic amino acids (3-COO^-, 5-*O*-sulfate and 6-*O*-sulfate).

In the phenyl ring system, the 2′-*O*-sulfate is interacting with Arg129 and the 5′-O-sulfate with Arg132, with additional interactions beyond those observed for the pentasaccharide. A combination of 2′- and 5′-*O*-sulfates appears to be optimal on this end.

Table 4. Predicted interaction profile for hit molecule ISOQ_EXT_R_6955 at activated antithrombin PBS.

Fragment	***Ligand group***	***Amino acid***	***Distance in Å***
Bicyclic ring (Tetrahydro isoquinoline)	3-Carboxylate	Lys125	3.5
	5-*O*-sulfate	Lys125	3.3
		Lys114	2.8
	6-*O*-sulfate	Lys114	3.0
		Arg47	2.9
		Arg46	2.6
	8-*O*-sulfate	Arg46	3.1
Unicyclic ring (Phenyl)	2′-*O*-sulfate	Arg129	3.0
	5′-*O*-sulfate	Arg132	2.9

Structural and predicted binding geometry analysis of the following hit molecules show that although they share many common features, differences in one or two important positions can have a significant or insignificant effect on binding affinity in terms of GOLD score (Figure 29 and 30).

Comparison of structures of ISOQ-EXT_R_6955, ISOQ-EXT_R_6411, ISOQ-EXT_R_6443 and ISOQ_EXT_6467 allows us to identify critical features needed for a high GOLD score. For example, ISOQ-EXT_R_6443 is essentially the same as ISOQ-E6955 except for the absence of an 8-*O*-sulfate.

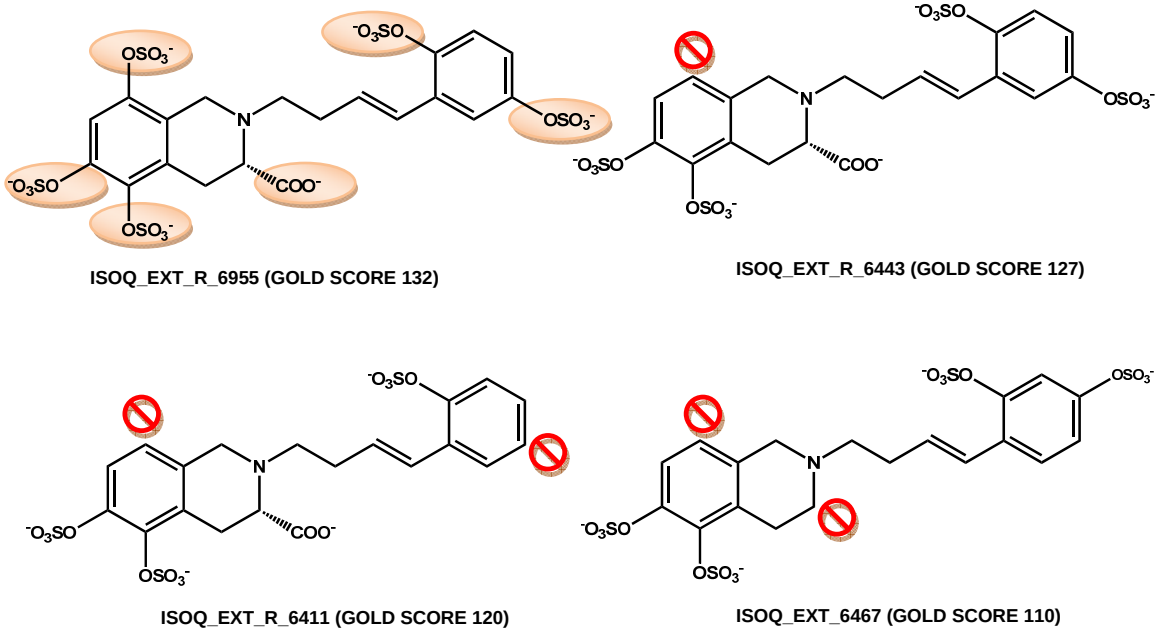


Figure 29. Representative structures showing the importance of sulfate and carboxylate groups and their contribution to the modified GOLDScore. Important groups in the reference hit 6955 are shown in the color filled ovals. Red icons indicate the absence of that particular substitution in comparison to the reference structure.

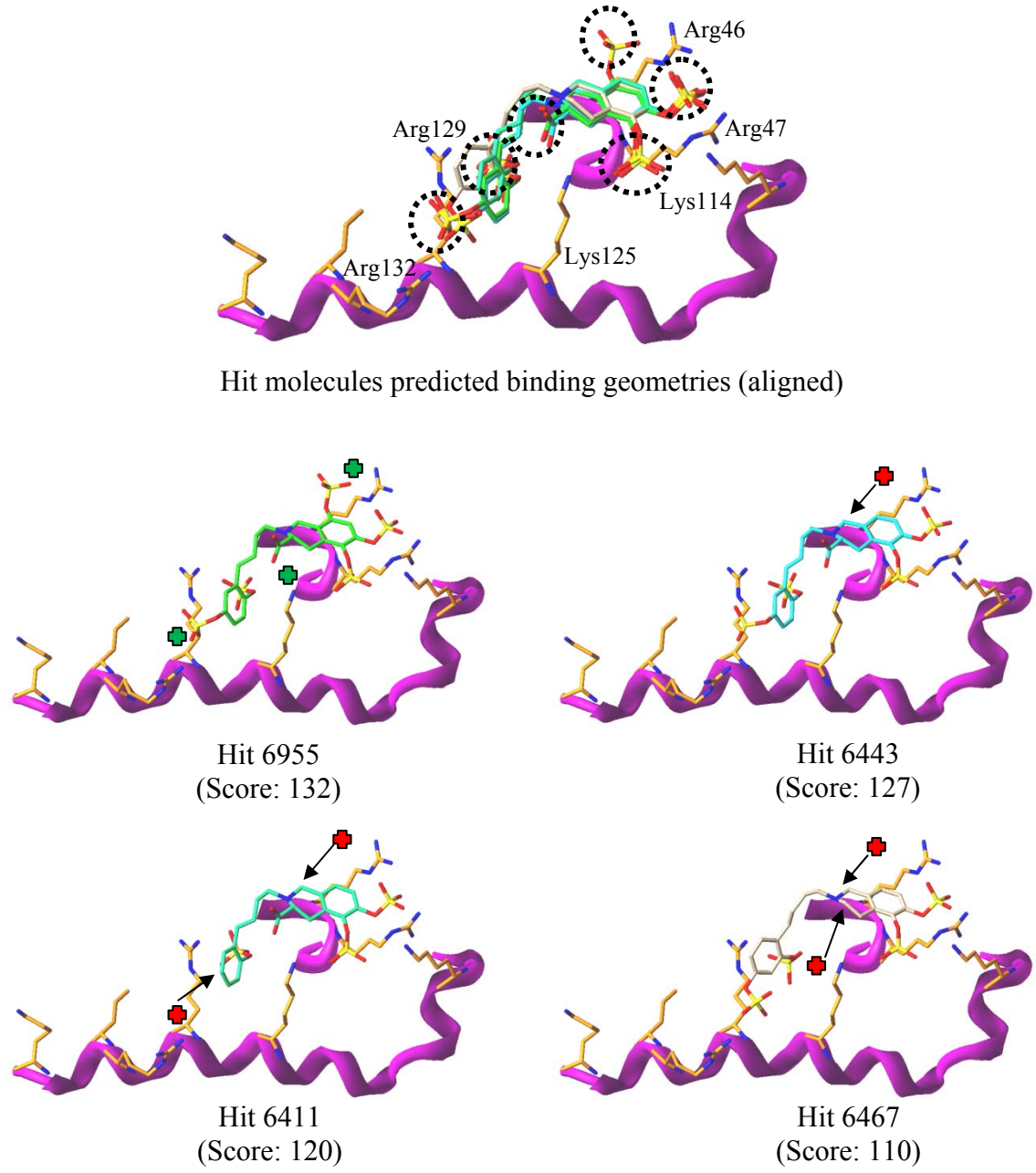


Figure 30. Predicted binding geometry of hit molecules in the activated antithrombin PBS. Green crosses show the interactions of reference hit 6955. Red crosses show the missing interactions in comparison to the reference compound responsible for reduction in the GOLD score.

The absence of the 3-carboxylate group makes a large difference in GOLD score of ISOQ-EXT_6467 (127 to 110) and is well below the average GOLD score for a 'hit' candidate. (The average GOLD Score for hit compounds is 119). This indicates the importance of 3-carboxylate group.

Hit 6443 is predicted to lose one interaction with Arg46 due to the absence of the 8-*O*-sulfate. Hit 6411, in addition to the 8-*O*-sulfate, does not have a 5′-*O*-sulfate group which is predicted to result in the loss of another hydrogen bond interaction with Arg132. This supports the idea of having 2 negatively charged groups in the phenyl ring system preferably in the 2′- and 5′- positions. The 2′-*O*-sulfate is predicted to interact with Arg129 and the 5′-*O*-sulfate with Arg132.

The 3-ene double bond in the linker provides an additional structural feature that restricts flexibility along the linker and places both 2′- and 5′-*O*-sulfate groups appropriately for interactions with Arg129 and Arg132 respectively.

Interestingly, even though some of the interactions are missing in these molecules in comparison to the reference structure (hit 6955), the presence of 5- and 6-*O*-sulfates and the 4-carbon linker appears to play an important role in maintaining the same binding geometry (Figure 30).

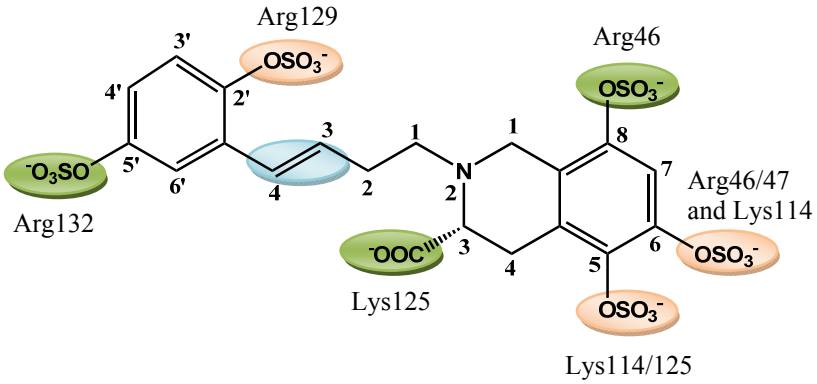


Figure 31. Hit molecule ISOQ_EXT_R_6955 showing the critical pharmacophore. Ovals in tan color represent the minimal requirements to recognize the PBS, whereas the ovals in light green make additional interactions.

Based on the molecular docking predictions, we propose that the bicyclic ring (tetrahydroisoquinoline) and the unicyclic ring (phenyl) connected by a 4-carbon linker with a '*trans'* double bond alpha to the phenyl ring, and 5-, 6- and 2′-*O*-sulfates form the basis for recognition of activated antithrombin PBS in a consistent manner. Additional key interactions are made by 8-*O*-sulfate (Arg46), 3-carboxylate (Lys125) and 5′-*O*-sulfate (Arg132) groups for higher GOLD score, and presumably higher affinity (Figure 31).

e. Summary of structural features of most favorable hits

The representative hit ISOQ_EXT_R_6955 from the virtual screening study shows the best fit in the PBS of activated antithrombin (Figure 28 and Table 4), but how well does this molecule mimic the interactions of DEFGH, a heparin pentasaccharide? Figure 32 shows an overlay between pentasaccharide $H5_{CRYS}$ (Figure 9) and the hit 6955 from the virtual screening study that contains the 3, 5, 6-trisubstituted tetrahydroisoquinoline fragment shown above and a 2′-*O*-sulfate group in the unicyclic ring. The indicated pose of

ISOQ_EXT_R_6955 was repeatedly predicted in triplicate docking experiments to be within 2.5 Å RMSD (Table 2).

The figure shows the predicted striking similarities in the location of 4 negative charges: 1) the 2′-*O*-sulfate group of ISOQ_EXT_R_6955 overlaps with the D-ring 6-*O*-sulfate group of $H5_{CRYS}$, 2) the 3-carboxylate overlaps with the E-ring carboxylate of the $H5_{CRYS}$, 3) the 5-*O*-sulfate overlaps with the critical unique F-ring 3-*O*-sulfate of the $H5_{CRYS}$ and 4) the 6-*O*-sulfate overlaps with the G-ring carboxylate group of the $H5_{CRYS}$.

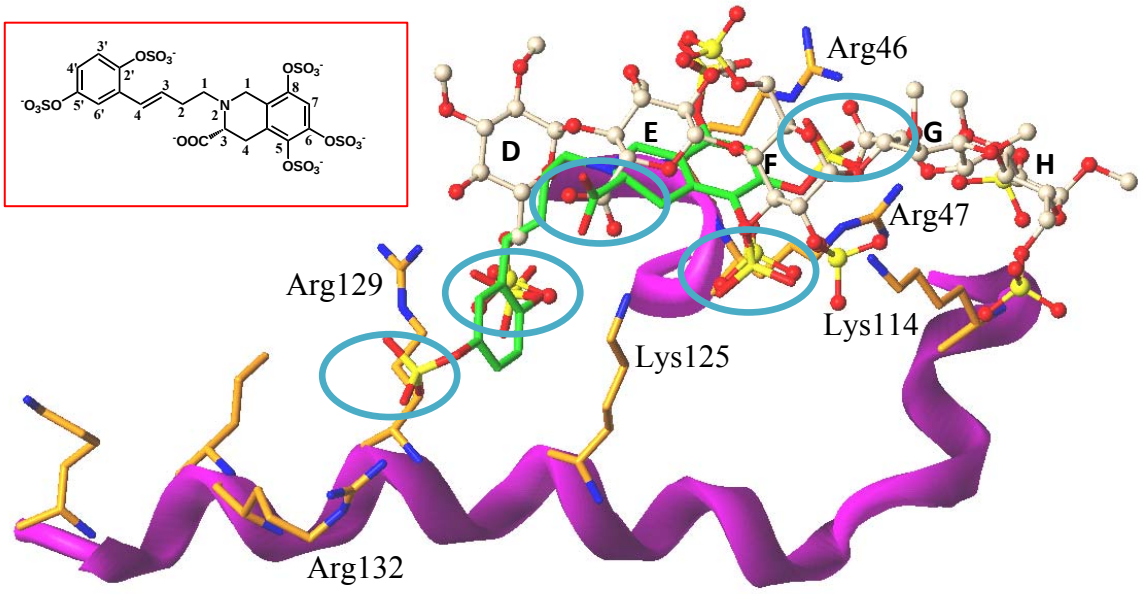


Figure 32. Overlay of heparin pentasaccharide H5 (gray), obtained from the crystal structure (PDB: 1TB6) and the consistently-predicted docking pose of hit ISOQ_EXT_R_6955 (green, structure shown in in-lay) from the virtual screening study of sulfated non-saccharide molecules. Overlying negative charges (sulfate and carboxylate groups) are circled. The critical PBS residues Lys114, Lys125 and Arg129 interact with the 6-*O*-sulfate, 5-*O*-sulfate and 2′-*O*-sulfate groups of ISOQ_EXT_R_6955, respectively, which in turn overlay with the G-ring carboxylate, F-ring *N*-sulfate and D-ring 6-*O*-sulfate groups of the H5 sequence. In addition, the 3-carboxylate of ISOQ_EXT_R_6955 and the E-ring carboxylate of H5 overlap and interact with Lys125. The 5′-*O*-sulfate makes an additional interaction with Arg132.

These charges are predicted to interact with the critically important positively charged triad (Lys114, Lys125 and Arg129) which is located in the pentasaccharide-binding site (PBS) in antithrombin (AT). Additionally, the 5′-*O*-sulfate group of ISOQ_EXT_R_6955 is predicted to interact with the extended heparin-binding site residue Arg132. Thus, the bicyclic ring is predicted to bind in the center of the PBS, and the unicyclic ring at the interface of the PBS and EHBS.

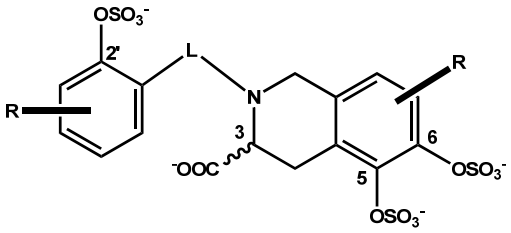
It is important to note that the 6-*O*-sulfate group in ISOQ_EXT_R_6955 overlaps with a carboxylate group of $H5_{CRYS}$. Since the 6-*O*-sulfate group is well-placed in the middle of Lys114, Arg46 and Arg47, and the sulfate group contains three potential interaction points, it is predicted to simultaneously interact with Lys114 (Helix P) and the N-terminal residues Arg46 and Arg47 (Helix A) (Figure 27 and table 6). At the same time, the G ring carboxylate of $H5_{CRYS}$ appears to interact only with Arg46 (hydrogen bond and ionic interaction).

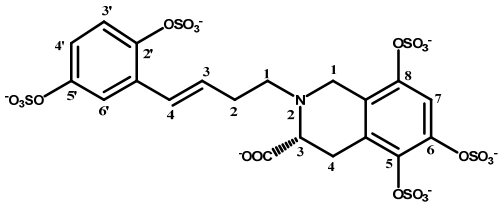
Another interesting result of our virtual screening study is the dependence of the overlap of structures shown in Figure 33 with linker length for the series of hits that contain fragment I and the 2′-*O*-sulfate group of the unicyclic ring.

Figure 34 reveals that, irrespective of the exact nature of the linker, the docking pose orients the interacting negative charges in an identical manner. For example, ISOQ_EXT_R_6748 with a 5-carbon linker containing a (*Z*)-double bond has a predicted binding mode similar to ISOQ_EXT_R_6687 that contains a fully saturated linker. This binding mode does not depend on the configuration of the carboxylate group (*R*- and *S*-isomers dock similarly) or even the absence of this group (not shown). This prediction

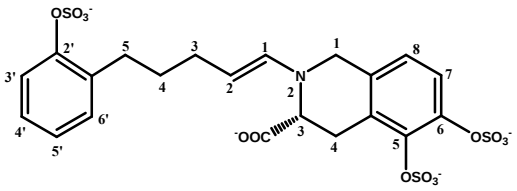
supports our conclusion that while 5- and 6-*O*-sulfates are necessary for favorable binding mode in PBS, a 4- to 5- carbon linker is predicted to be optimal for placing the unicyclic ring to reach for interaction with Arg129.

Table 5. Summary of structural features of hit molecules

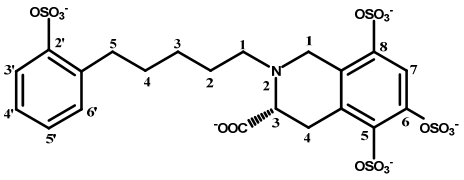
Fragment I	Number of 'hits'	GOLD Scores	Percentage with 4 or 5 carbon linkers
	37	104-132	76%



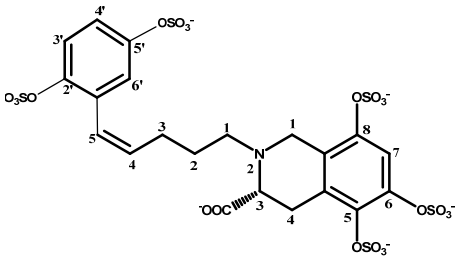
ISOQ_EXT_R_6955 (Green)



ISOQ_EXT_R_6170 (Cyan)



ISOQ_EXT_R_6687 (Gray)



ISOQ_EXT_R_6988 (Black)

Figure 33. The hit structures with fragment I, but with variation in the linker

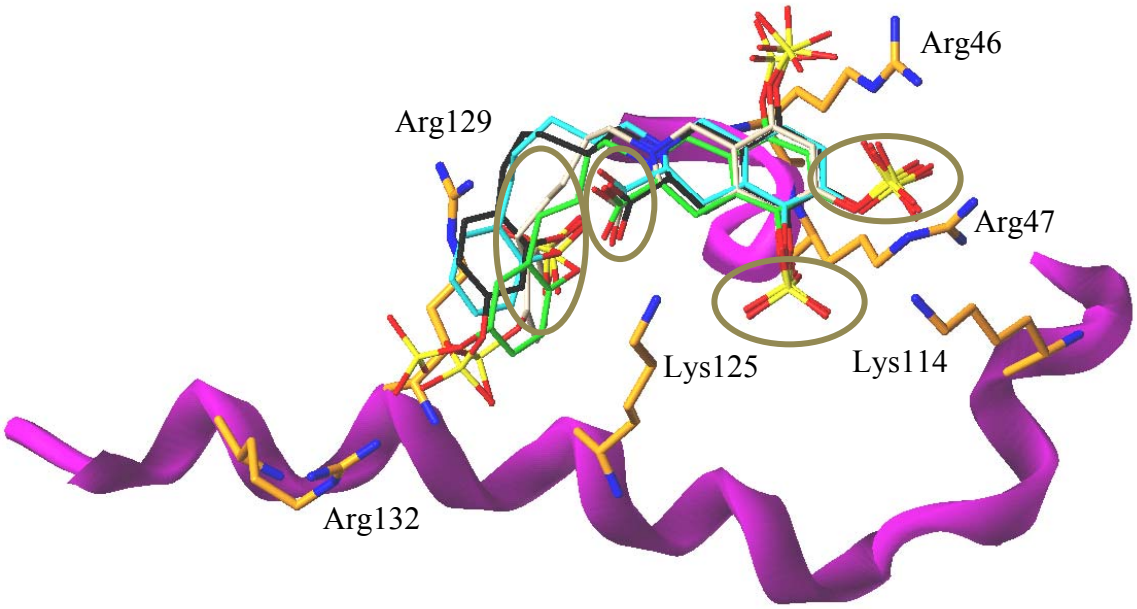


Figure 34. Overlay of different docked poses of hits cotaining core fragment I and 2'-substituted unicyclic ring. Note the striking similarity of ortientation of the bicyclic ring and circled negative charges. Green capped stick: ISOQ_EXT_R_6955; Cyan capped stick: ISOQ_EXT_R_6170; Gray capped stick: ISOQ_EXT_R_6687; Black capped stick: ISOQ_EXT_R_6988.

2.5. Initial Biochemical Validation of Modeling Results

The molecular modeling results presented in the above sections lead to several hypotheses, as may be apparent from the discussion following results. Of special importance are three hypotheses. The modeling results suggest 1) an optimal length of the 4 to 5 atom linker; 2) 2′,5′-disulfated unicyclic ring; and 3) 3,5,6-trisubstituted bicyclic ring are to be important. The hypotheses form the basis of synthetic work currently in progress in the Desai laboratory. The synthesis of highly sulfated molecules is challenging as demonstrated by previous work on such molecules in the laboratory.[209] To date, no

molecule with all three features has been synthesized. Yet, the laboratory has just synthesized and tested one molecule that may serve as initial test of molecular modeling experiments.

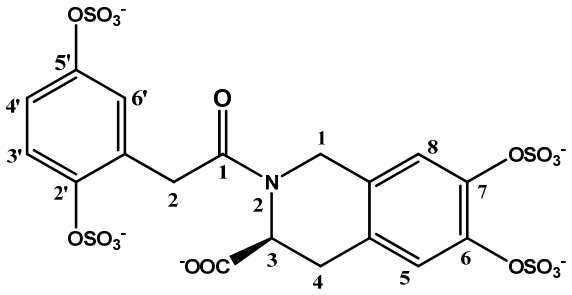


Figure 35. Structure of synthesized molecule 67A2L25.

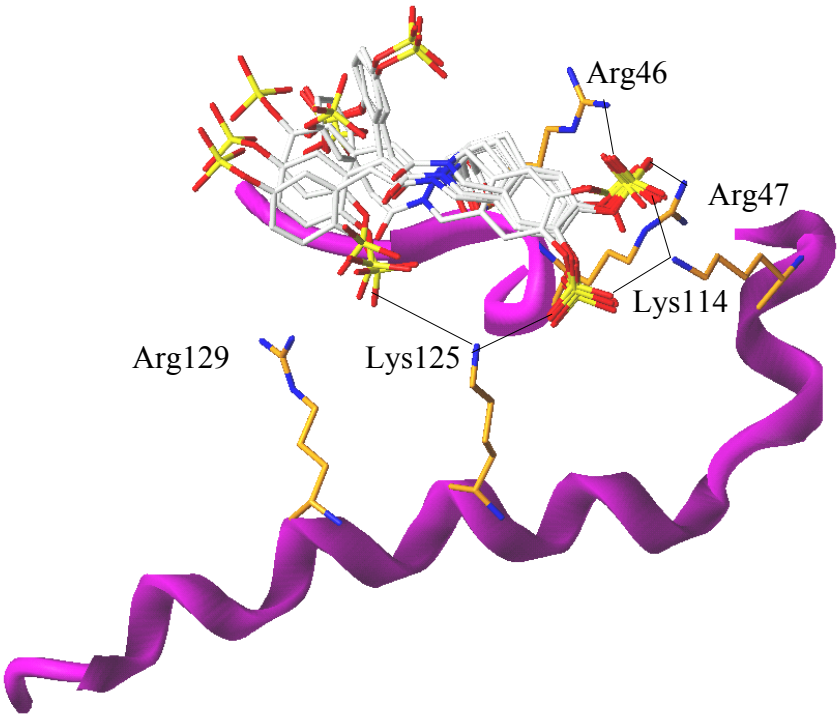


Figure 36. The predicted binding modes for the synthesized molecule 67A2L25 in triplicate docking experiment (2 top ranked solutions from each experiment). The solid black lines show the possible interactions between the ligand and the protein. Maximum GOLDScore is 97.2.

Potential activator 67A2L25 (Figure 35) was synthesized (by Mr. Al-Horani) and evaluated (by Dr. Liang) using the standard antithrombin activation protocol developed by the laboratory. Activator 67A2L25 has a 6,7-disulfated bicyclic ring (instead of 5,6-disubstitution), a two carbon linker (instead of 4 or 5 carbon linker); and a 2,5-disubstituted unicyclic ring (as desired).

Biochemical analysis of 67A2L25 shows an antithrombin activation potential in the region of 50- to 100-fold (the high variance is because of the very high salt content as impurity in the sample). The activation potential of 67A2L25 is much less than full antithrombin activation (300-fold), but significantly more than 8 – 30-fold activation achieved with previous designs.[134,135,209] The antithrombin affinity was found to be between 1 and 5 μM at pH 7.4. This is significantly higher than affinities of all previous designs (50 to 500 μM),[171,212,214] though still less than most potent saccharide-based activators (50 to 100 nM).

Since the synthesized molecule was not originally present in the virtual library, this molecule was docked using the same docking parameters. Modeling results for the synthesized molecule shows that the GOLD score for 67A2L25 is 97.2, which is lower than the scores for the hit molecules identified in the virtual library screening, and the binding modes are also divergent in triplicate docking experiments.

Several conclusions may be derived based on this single data point. 1) The activation is not full (50-100 fold in comparison to 300-fold) for compound 67A2L25 probably because of the absence of optimal structural features, as predicted by GOLD modeling; 2) the activation is much better than that realized with previous designs

probably because of the slightly longer linker and/or 2′,5′-disulfation of the unicyclic ring; 3) the antithrombin affinity of 67A2L25 is found to be higher perhaps because of more optimal features (slightly longer linker and/or 2,5-disulfation of unicyclic ring) than those present in previous designs; and 4) the antithrombin affinity of 67A2L25 is not as great as the most optimal saccharide activators because all features suggested by the docking experiments have not been introduced. Overall, the antithrombin binding and activation results obtained with 67A2L25 are consistent with the GOLD-based molecular modeling results obtained and provide stronger impetus to synthesize and test the 'high-affinity, high-specificity' molecules identified in this work.

2.6. Experimental Section

2.6.1. Software/Hardware

SYBYL 7.2 (Tripos Associates, St. Louis, MO) was used for molecular visualization, for minimization, and for adding hydrogens to protein structures from the Protein Data Bank. All modeling was performed on MIPS R16K or R14K IRIX 6.5-based SGI Tezro and Fuel graphical workstations. GOLD version 3.0 was used for docking experiments. Combinatorial tetrahydroisoquinoline bicyclic-unicyclic small molecule structures were built using LEGION, a structure building module that creates all possible structures in compliance with the given input limits for a core structure. Then the structures were converted into 3D form by the program CONCORD and checked for atom types in an automated manner using in-house SPL (SYBYL Programming Language) scripts.

2.6.2. Energy Minimizations

Energy minimization of modeled structures was performed to optimize the geometric conformation of small molecules in the library and the protein antithrombin. Except where stated, energy minimization was performed using the Tripos Force Field with Gasteiger-Hückel charges, a fixed dielectric constant of 80, and a non-bonded cutoff radius of 8Å. Minimization was carried out for a maximum of 5000 iterations subject to a termination gradient of 0.05 kcal/(mol-Å) for protein whereas for small molecules 10000 iterations subject to a termination gradient of 0.01 kcal/(mol-Å).

2.6.3. Protein Coordinates

The coordinates for the activated form of AT were extracted from the crystal structure of the ternary AT–pentasaccharide–thrombin complex (PDB entry 1TB6). Hydrogen atoms were added in SYBYL 7.2 and the structure was minimized with fixed heavy-atom coordinates using the Tripos force field for 1000 iterations subject to a termination gradient of 0.05 kcal / (mol-Å). Single-point correction was made on AT (1TB6) at Lys133 side chain which is incomplete in its original crystal structure (AT_K133). The side-chain of the corrected residue was optimized through energy–minimization in which the all other side chains were held rigid. Native antithrombin coordinates were extracted from PDB structure 2ANT.

2.6.4. Coordinates for ISOQ Virtual Library

The coordinates for the virtual library of 24576 compounds based on the tetrahydroisoquinoline scaffold were generated with a series of SPL scripts from a core

structure with variable substitutions using LEGION and CONCORD. Atom type checks and appropriate changes were made by an in–house SPL script.[109] The structures were then minimized using the Tripos force field for 10000 iterations subject to a termination gradient of 0.01 kcal / (mol-Å) in an automated manner.

2.6.5. Docking Protocol

Docking of the virtual library onto the activated and native form of antithrombin was performed with GOLD v.3.0. The Pentasaccharide Binding Site (PBS) in activated antithrombin is predefined as all atoms within 16 Å from the C^{ζ} atom of Phe121 in the D helix. This definition of the binding site covers all important known heparin binding residues including Lys11, Arg13, Arg46, Arg47, Trp49, Lys114, Phe121, Lys125, Arg129, and Arg132.

The Extended Heparin Binding Site (EHBS) in activated antithrombin is predefined as all atoms within 14 Å from C^{α} of Lys133 in the D helix which includes Lys125, Arg129, Arg132, Lys133, Lys136, and Lys139.

The binding site in native antithrombin is predefined as all atoms within 19 Å from the terminal *N* atom of Lys125 in the D helix which includes Lys11, Arg13, Arg46, Arg47, Trp49, Lys114, Phe121, Lys125, Arg129, and Arg132.

GOLD 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 Lys residues as part of the docking process. Unless specified otherwise, default parameters were employed during the GOLD docking runs.

When docking the virtual combinatorial library, a two-step docking protocol was utilized. The first step consisted of docking all structures in the library using 10000 GA iterations (7×8 speed up) and GOLD score evaluation of only the top-ranked solution. In this step, most promising candidates were identified (structures with GOLD Score of 100 and above, GOLD Score for DEF is 115 and for DEFGH is 140). The second step consisted of docking of hits obtained from the first step in **triplicate** using **no speed-up** and a **genetic algorithmic search** with 100000 iterations for each of 10 runs to ensure reproducibility and to reduce false positives. In this search, GOLD 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 toward 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 preterminate if at any point during the docking run the top two ranked solutions were within 2.5 Å RMSD. The top-ranked solution of each docking experiment was considered for further analysis.

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 for this purpose. This modified GOLDScore,

which utilizes hydrogen-bonding and van der Waals interactions (Eq.1) was used to rank the final docked solutions.

$$GOLDScore_{mod} = HB_{ext} + 1.375 \times VDW_{ext} \qquad Eq.(2)$$

HB_{ext} and VDW_{ext} are the "external" (non-bonded interactions taking place between the ligand and receptor) hydrogen bonding and van der Waals terms, respectively. Unless otherwise noted, the terms 'GOLD Score' and 'modified GOLDScore' in this work both refer to Equation 2.

Table 2. RMSD data for hit molecules which passed the affinity and specificity filter convergence cut off value 2.5 Å in triplicate.

SL. NO.	*COMPOUND*	DOCK 1 RMSD	*DOCK 2* RMSD	*DOCK 3* RMSD	*RMSD* 1 to 2	*RMSD* 2 to 3	*RMSD* 1 to 3
1	ISOQ_1903	1.4	1.5	1.2	1.5	1.8	1.6
2	ISOQ_2561	2.4	2.2	1.8	2.2	2.4	0.9
3	ISOQ_6435	0.9	1.4	0.8	1.5	1.1	1.1
4	ISOQ_2096	2.0	1.3	0.5	1.5	1.5	0.7
5	ISOQ_5809	1.6	2.3	1.7	1.3	1.6	1.8
6	ISOQ_6162	1.1	0.8	1.4	1.3	1.3	1.3
7	ISOQ_R_1875	0.9	2.4	1.0	1.7	1.2	1.7
8	ISOQ_R_2106	1.8	1.4	1.1	1.8	1.3	1.7
9	ISOQ_R_5769	1.3	1.4	1.8	1.5	1.1	1.5
10	ISOQ_R_6018	0.5	0.5	0.5	0.8	1.1	1.1
11	ISOQ_R_6053	0.2	0.7	2.3	1.0	1.2	1.2
12	ISOQ_R_6408	1.5	1.1	2.2	1.8	1.2	1.8
13	ISOQ_R_1960	0.5	1.4	1.5	1.1	1.8	1.6
14	ISOQ_R_6082	1.8	1.6	0.8	1.4	1.4	1.2
15	ISOQ_R_1059	1.1	0.5	0.6	0.8	1.0	0.9
16	ISOQ_R_277	0.4	1.1	0.3	1.1	1.1	0.9
17	ISOQ_R_5183	2.3	0.2	0.9	1.1	1.0	0.9
18	ISOQ_R_5218	1.3	0.7	2.3	1.0	0.9	1.1
19	ISOQ_R_5325	1.4	1.0	1.8	1.7	2.2	1.8
20	ISOQ_S_1873	1.4	1.2	1.2	1.7	1.2	1.5

21	ISOQ_S_2072	0.2	1.9	0.7	1.9	1.6	1.5
22	ISOQ_S_4255	1.4	0.5	1.3	1.5	1.3	1.2
23	ISOQ_S_5963	0.8	1.0	0.6	1.0	0.9	1.0
24	ISOQ_S_6153	2.4	1.3	2.3	2.4	1.1	2.5
25	ISOQ_S_6159	1.6	0.6	1.0	1.0	1.2	1.0
26	ISOQ_S_6330	0.8	1.7	1.0	1.3	0.9	1.1
27	ISOQ_S_6463	1.7	1.9	0.9	1.6	1.3	1.4
28	ISOQ_S_6623	0.9	1.0	1.0	1.4	0.6	1.4
29	ISOQ_S_6652	1.6	1.3	0.6	1.8	1.5	1.7
30	ISOQ_S_6960'	2.4	2.2	2.4	2.4	2.4	1.2
31	ISOQ_S_7547	0.9	1.6	0.7	1.1	1.5	1.7
32	ISOQ_S_7832	0.8	0.7	1.4	0.9	1.2	1.1
33	ISOQ_EXT_S_1229	1.1	0.3	1.1	1.0	1.0	1.0
34	ISOQ_EXT_S_1243	0.7	1.3	0.2	1.3	0.7	1.3
35	ISOQ_EXT_S_6185	0.6	1.1	0.7	1.2	1.1	1.4
36	ISOQ_EXT_S_6234	1.6	2.0	2.5	1.4	1.5	1.5
37	ISOQ_EXT_S_6236	2.1	1.0	1.1	1.7	1.7	1.1
38	ISOQ_EXT_S_6410	2.3	1.5	2.0	2.1	1.8	1.6
39	ISOQ_EXT_S_6411	1.5	0.5	1.4	1.1	1.2	1.2
40	ISOQ_EXT_S_6448	0.7	0.6	1.2	1.7	1.4	2.1
41	ISOQ_EXT_S_6462	1.5	1.4	1.1	0.9	2.0	1.9
42	ISOQ_EXT_S_6815	2.0	1.6	1.3	1.8	1.4	1.8
43	ISOQ_EXT_S_6971	1.4	0.9	1.7	1.7	1.1	1.6
44	ISOQ_EXT_S_6987	2.0	2.3	0.6	1.7	1.1	1.6
45	ISOQ_EXT_S_6988	0.8	1.7	2.4	1.5	1.6	1.9
46	ISOQ-EXT_5142	1.6	1.3	1.8	1.1	1.5	1.0
47	ISOQ-EXT_6294	1.2	0.4	0.8	0.9	1.3	0.9
48	ISOQ-EXT_6550	0.5	1.8	1.3	1.2	1.5	1.8
49	ISOQ-EXT_3201	2.2	0.5	1.0	2.2	1.8	1.5
50	ISOQ-EXT_6225	1.2	0.9	0.4	1.3	1.6	1.0
51	ISOQ-EXT_6467	1.1	1.7	2.3	1.5	1.4	1.4
52	ISOQ-EXT_979	1.4	1.0	1.2	1.9	1.8	1.6
53	ISOQ-EXT_2850	1.2	1.6	0.9	1.8	1.6	1.7
54	ISOQ-EXT_2866	2.2	2.4	1.1	2.2	1.8	1.4
55	ISOQ-EXT_3906	1.4	1.9	1.5	1.9	2.0	1.1
56	ISOQ-EXT_6402	2.0	1.1	2.1	1.2	1.1	0.9
57	ISOQ-EXT_6434	0.8	2.4	1.4	1.4	1.7	1.9
58	ISOQ-EXT_706	1.0	0.6	1.2	1.4	1.4	1.0
59	ISOQ-EXT_7938	2.3	1.2	1.3	1.4	1.8	1.7
60	ISOQ-EXT_6532	0.6	2.3	2.5	1.2	1.6	1.4
61	ISOQ-EXT_855	1.7	2.5	0.5	1.5	1.5	1.0
62	ISOQ-EXT_7064	2.1	2.2	1.6	2.4	2.4	1.7
63	ISOQ-EXT_R_3182	1.3	0.9	2.1	2.5	1.3	2.4

64	ISOQ-EXT_R_6318	1.4	1.6	1.5	2.0	2.4	1.7
65	ISOQ-EXT_R_6542	0.6	0.9	1.3	1.4	1.0	1.6
66	ISOQ-EXT_R_6217	1.3	1.8	2.3	1.7	1.0	1.5
67	ISOQ-EXT_R_6281	0.9	0.9	2.3	0.8	1.4	1.4
68	ISOQ-EXT_R_6297	0.6	0.4	1.1	1.4	1.1	1.2
69	ISOQ-EXT_R_6345	0.9	1.5	1.0	1.5	1.3	1.2
70	ISOQ-EXT_R_6409	2.4	2.3	2.0	0.7	1.9	2.2
71	ISOQ-EXT_R_6473	2.1	0.6	0.7	1.6	1.1	1.6
72	ISOQ-EXT_R_6857	2.4	2.5	2.2	1.5	2.1	2.0
73	ISOQ-EXT_R_6921	1.1	2.3	1.8	1.5	1.4	1.8
74	ISOQ-EXT_R_2861	1.8	0.9	2.2	1.8	1.4	1.3
75	ISOQ-EXT_R_6253	2.5	0.9	2.2	2.4	1.2	2.5
76	ISOQ-EXT_R_6411	0.6	1.2	0.4	1.0	1.3	1.4
77	ISOQ-EXT_R_6443	1.1	1.1	0.4	0.9	1.0	0.9
78	ISOQ-EXT_R_6555	1.6	1.1	1.6	0.8	1.4	1.4
79	ISOQ-EXT_R_6955	1.2	1.5	1.2	1.1	1.0	0.8
80	ISOQ-EXT_R_731	0.7	1.5	1.2	1.3	2.5	2.2
81	ISOQ-EXT_R_6170	1.7	1.5	1.2	2.0	2.0	1.1
82	ISOQ-EXT_R_6474	1.5	1.5	1.3	1.2	1.8	1.8
83	ISOQ-EXT_R_6538	1.4	1.4	0.9	1.0	1.6	1.5
84	ISOQ-EXT_R_2956	1.8	2.3	2.2	1.9	2.0	1.5
85	ISOQ-EXT_R_6444	1.8	2.2	2.2	1.9	2.0	1.4
86	ISOQ-EXT_R_6732	2.2	1.6	2.2	2.5	2.4	1.8
87	ISOQ-EXT_R_6988	1.3	2.4	1.6	1.5	1.6	1.4
88	ISOQ-EXT_R_2591	1.3	2.0	0.8	1.6	1.4	1.2
89	ISOQ-EXT_R_3279	1.6	2.3	1.7	2.4	1.4	2.5
90	ISOQ-EXT_R_6687	1.9	0.6	2.1	2.4	2.5	0.9
91	ISOQ-EXT_R_3408	1.8	1.7	2.5	1.3	1.5	1.7
92	ISOQ-EXT_R_6224	1.9	1.3	1.7	1.8	2.4	2.2

CHAPTER 3

DESIGNING SPECIFIC THROMBIN EXOSITE-II MODULATORS

3.1. Introduction

As described in chapter 1, heparin is a complex carbohydrate biopolymer made of linear polysaccharides of varying chain length and is composed of uronic acid and glucosamine residues.[113] Although heparin is a commonly used anticoagulant it suffers from serious side effects such as excessive bleeding complications, heparin induced thrombocytopenia, and significant intra- and inter-patient dose response variability.[173,174]

Most of the side effect complications arise primarily from its structural complexity and its interaction with numerous proteins other than coagulation proteins factor Xa (fXa), thrombin, and antithrombin (Table 6). Implicitly these diverse roles must arise from an optimal combination of specificity and affinity. However with the exception of a few cases, the specificity of heparin interactions has been poorly understood and not explored in detail.

A major reason for the limited understanding of heparin–protein interactions is due to the phenomenal structural diversity of heparin. Structural diversity of heparin is a result of the complex, highly anionic polysaccharides composed of alternating 1→4-linked glucosamine and uronic acid residues, which are variously modified through sulfation, acetylation and epimerization.[172] These modifications can produce 48 different

disaccharide building blocks, of which 23 have been found to date (Figures 37 and 38).[220] In addition, the iduronic acid residue (IdoAp) can exist in multiple conformations, especially $^{1}C_{4}$ and $^{2}S_{O}$ for internal locations that can inter-convert relatively easily (Figure 39).[132,221] As a result, the combination of structural and conformational variability generates millions of sequences, of which few are expected to specifically recognize a target protein.

Specificity in heparin-protein interaction is a function of both the target protein and the heparin sequence that binds the protein. Capila, I and Linhardt R. J. have reviewed the interactions of heparin with many different proteins (Table 6).[115]

Table 6. Characteristics of selected list of GAG binding proteins. Adapted from Capila, I.; Linhardt, R. J. Heparin–Protein Interactions. *Angew. Chem. Int. Ed.* **2002,** *41,* 390-412.

Heparin-binding protein	**Physiological/ pathological role**	K_D	**Oligosaccharide size**	**Sequence features**[#]
Antithrombin	Coagulation cascade	~20nM	5-mer	GlcNS3S6S
Thrombin*	Coagulation cascade	7μM	8-mer	HS
FGF-1	Cell proliferation, differentiation, morphogenesis and angiogenesis	nM	4-mer to 6-mer	IdoA2S-GlcNS6S
FGF-2	As FGF-1	nM	4-mer to 6-mer	IdoA2S-GlcNS
PF-4	Inflammation and wound healing	nM	12-mer	HS/LS/IS
IL-8	Pro-inflammatory cytokine	~6μM	18-mer to 20-mer	HS/LS/IS
SDF-1α	Pro-inflammatory mediator	~20nM	12-mer to 14-mer	HS
HIV-1 gp120	Viral entry inhibition	0.3μM	10-mer	HS
HSV gB and gC	Viral entry inhibition	–	–	–

*Data is obtained from Huntington et al. *J Biol. Chem.* **2005,** *280,* 2745-2749. [#]HS: high sulfation, IS: intermediate sulfation, LS: Low sulfation

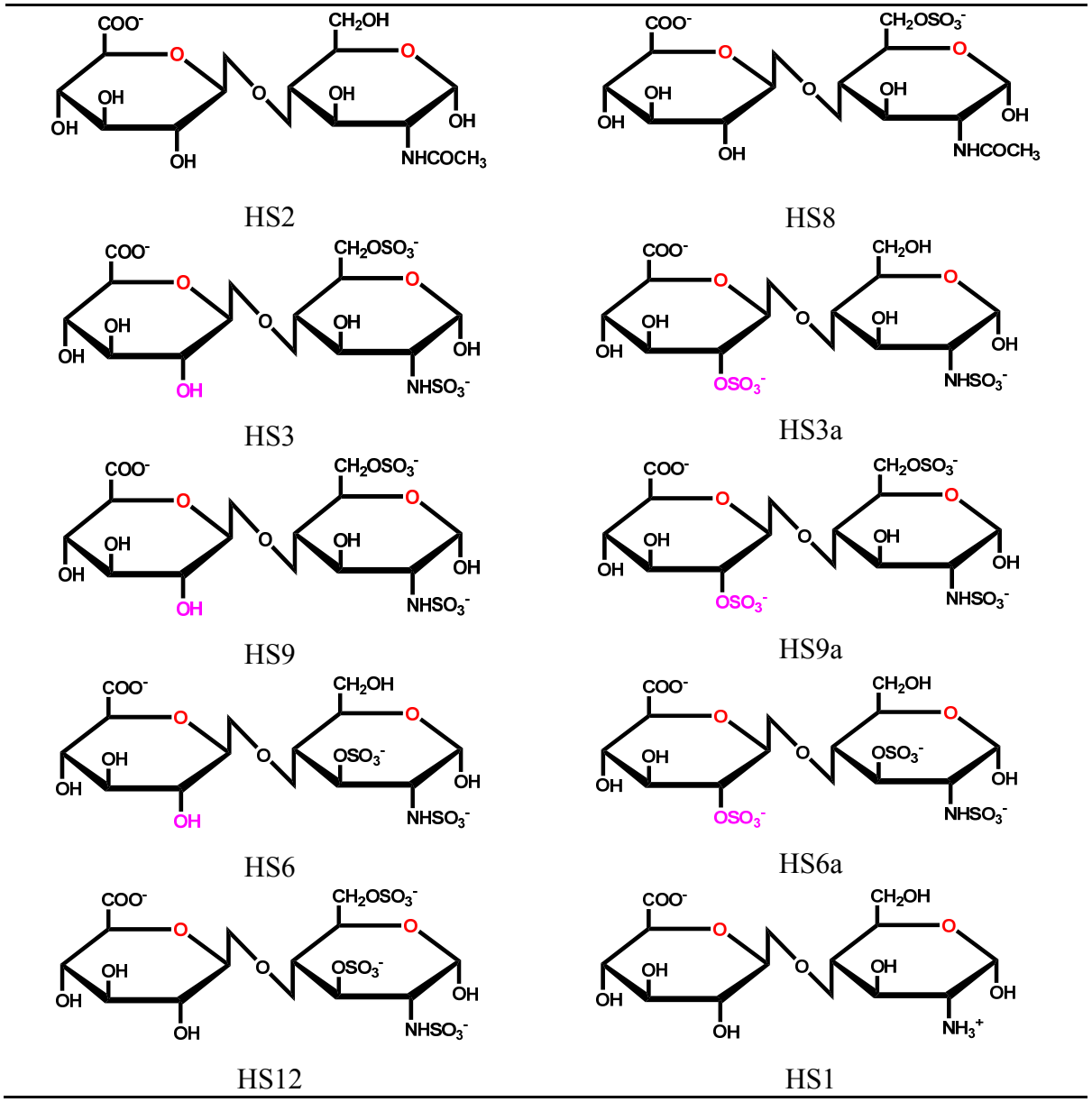
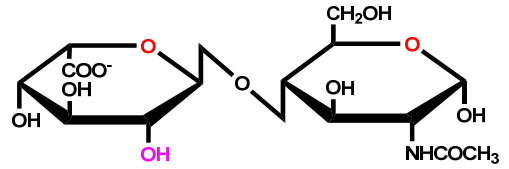
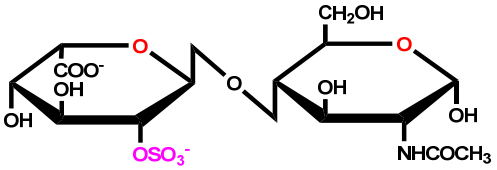


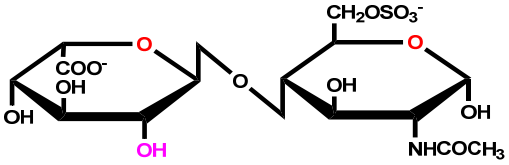
Figure 37. The glucuronic acid-containing disaccharide subunits known to be present in heparin/heparan sulfate. Groups in magenta show the difference in sulfation pattern between the pairs (the names of the disaccharides containing a 2-*O*-sulfate glucuronic acid residue end with the letter 'a'). Adapted from 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.



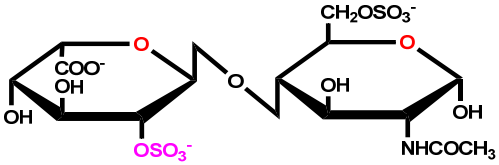
HS14 ($^1C_4$ and ^2S_O)



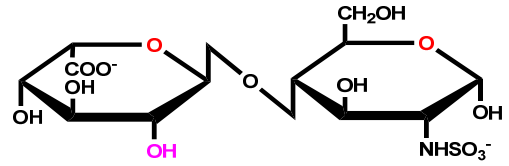
HS14a ($^1C_4$ and ^2S_O)



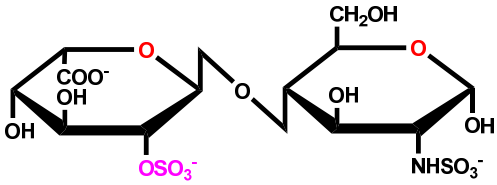
HS20 (1C4 and 2SO)



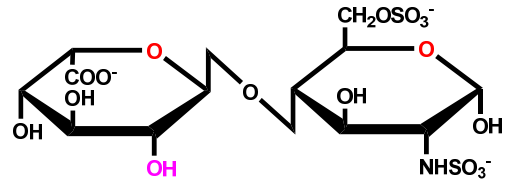
HS20a ($^1C_4$ and ^2S_O)



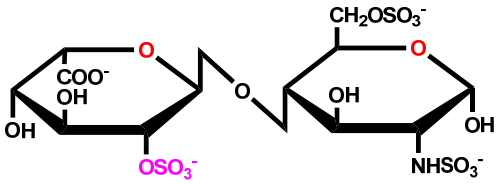
HS15 ($^1C_4$ and ^2S_O)



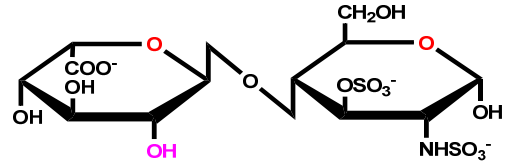
HS15a ($^1C_4$ and ^2S_O)



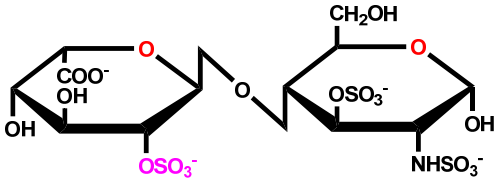
HS21 ($^1C_4$ and ^2S_O)



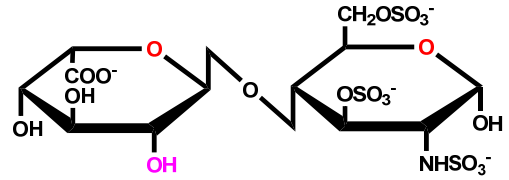
HS21a ($^1C_4$ and ^2S_O)



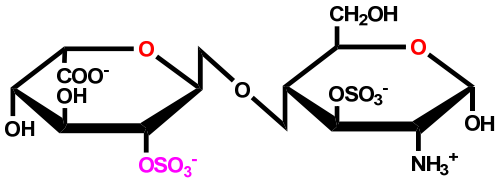
HS18 ($^1C_4$ and ^2S_O)



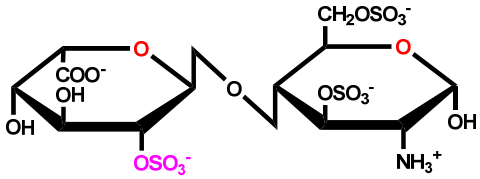
HS18a ($^1C_4$ and ^2S_O)



HS24 ($^1C_4$ and ^2S_O)



HS16a ($^1C_4$ and ^2S_O)



HS22a ($^1C_4$ and ^2S_O)

Figure 38. The iduronic acid-containing disaccharide subunits known to be present in heparin/heparan sulfate. Groups in magenta show the difference in sulfation pattern between the pairs (the names of the disaccharides containing a 2-*O*-sulfate iduronic acid residue end with the letter 'a').

Adapted from 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.

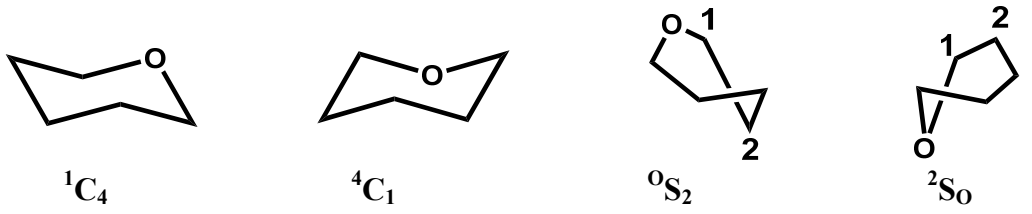


Figure 39. Iduronic acid conformations ($^1C_4$ and ^2S_O predominantly exist in GAGs)

On the protein front, Cardin and Weintraub were the first to study the general structural requirements for GAG-protein interactions.[222] They have demonstrated that the heparin binding domain in the protein has defined binding motifs of linear consensus sequences with specific repeat pattern. Others have suggested a spatial distance relationship to be important for heparin binding.[223,224]

Arginine and lysine residues in the protein surfaces dominate heparin-binding sites and are known to be critical in heparin binding. It is also known that not all arginine and lysine interactions with sulfate and carboxylate groups of heparin are identical. Arginine has been shown to bind 2.5 times more tightly than lysine due to the strong bidendate interaction of the guanidino group with sulfate and carboxylate groups.[225] Even though heparin–protein interactions are known to be primarily charge-based interactions, the degree of ionic contribution varies significantly among heparin-binding proteins. In the case of antithrombin and basic fibroblast growth factor, the ionic binding energy contribution is ~40% and 30% respectively, but with thrombin, it is ~80%.[206,226,227]

Although the importance of individual amino acids in recognizing heparin fragments has been studied in many proteins, the optimal 3D orientation that generates high specificity and affinity remains unclear. Considering the complexity of heparin

structure and its diverse physiological functions, molecular modeling and computational docking approaches represent a powerful means to address the structural complexity of heparin and exploring the issue of specificity and binding affinity of heparin–protein interactions.

In the present work, we have used molecular modeling approaches to explore the nature of the heparin binding site in antithrombin and thrombin with a special focus on understanding specificity/non-specificity. Understanding the features that are important for specificity in this complex system both in protein and GAGs, extracting the hidden information and decoding them will enable us to design new therapeutic agents possibly without the above-mentioned side effects and limitations associated with GAG-based anticoagulants.

3.2. Rationale

The interaction of heparin with antithrombin and thrombin is the basis for anticoagulation therapy. On one hand, the interaction of heparin with antithrombin is a known example of specific interaction. A specific five-residue present in heparin, called pentasaccharide DEFGH sequence (Figure 40), mediates the interaction with antithrombin. In fact, the absence of this sequence makes the heparin chain incapable of binding to antithrombin under physiological conditions. The importance of individual amino acids interacting with this sequence-specific pentasaccharide and the role of each residue in the pentasaccharide responsible for recognition and affinity to antithrombin has been extensively studied.[166,167,206,207,209]

On the other hand, heparin interaction with thrombin is considered to be a non-specific interaction. Equilibrium binding studies of heparin binding to thrombin by Olson et al. clearly showed that heparin interaction is a non-specific electrostatic interaction.[227] This is also supported by the crystal structure study of thrombin bound to heparin octasaccharide.[228] In the crystal structure, heparin octasaccharide has two different orientations (binding mode) in the heparin binding site also known as thrombin exosite-II (Figure 44).

In antithrombin, there are three important amino acids involved in heparin pentasaccharide binding: Lys114, Lys125 and Arg129. In thrombin, there are at least seven basic amino acids predicted to be involved in heparin octasaccharide binding: Arg93, Arg101, Arg126, Arg165, Arg233, Lys236 and Lys240. In general, one would expect less specificity with fewer interacting points in the target protein, or alternatively greater specificity with more interacting points. However, what has been observed is the reverse: whereas thrombin with more Lys and Arg shows non-specific interaction, antithrombin with fewer Lys and Arg makes specific interaction. The molecular basis responsible for the origin of this specificity/non-specificity is not clear. We reasoned that molecular modeling techniques may be used to analyze this issue.

3.3. Antithrombin-Heparin: Specific Interaction

As mentioned earlier, a specific five-residue GAG sequence present in heparin, called pentasaccharide DEFGH (Figure 40), mediates the interaction with antithrombin.

Absence of this sequence renders the heparin chain incapable of binding to antithrombin under physiological conditions.

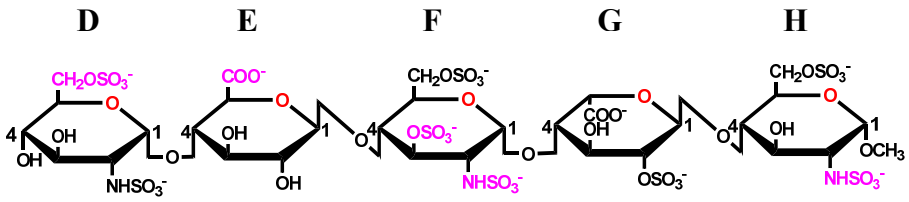


Figure 40. Structure of a specific five-residue heparin pentasaccharide DEFGH. Groups in magenta are known to be important for specificity.

Heparin chains containing the DEFGH sequence are referred to as high-affinity heparin (HAH) chains. Pentasaccharide DEFGH binds to antithrombin with ~50 nM affinity under physiological conditions. This binding affinity accounts for ~95% of the free energy of binding of the full-length polymer. This binding energy is contributed due to the interactions of 6-*O*-sulfate group of D ring, 6-COO^{-} group of ring E, 2-*O*-sulfate and 3-*O*-sulfate of F ring, and 2-*N*-sulfate of ring H in the pentasaccharide binding site (PBS).[217] Of these, the rare 3-*O*-sulfate group of the central glucosamine F plays the key role without which the affinity of the sequence is lost.[217,229] This sequence-specific interaction of DEFGH with antithrombin serves as a model for specificity. (Figure 41)

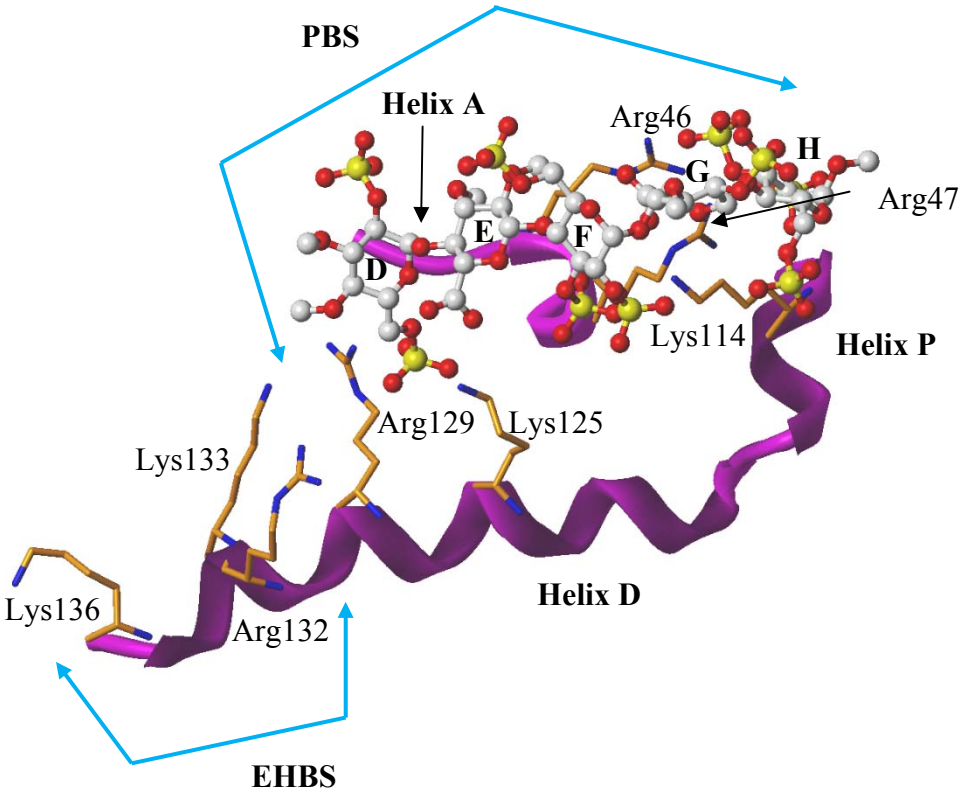


Figure 41. A close-up view of the structure of the heparin-binding sites in antithrombin (PBS and EHBS). The structure of co-complex was obtained from PDB entry 1AZX. Pentasaccharide DEFGH is shown in ball-and-stick representation. Helices D, P and A (C-terminal end) form the heparin-binding site. Arg46, Arg47, Lys114, Lys125 and Arg129 form the pentasaccharide binding site (PBS), while Arg132, Lys133 and Lys136 form the extended heparin-binding site (EHBS).

It appears that DEFGH specificity in antithrombin is a result of its multiple interactions with the Lys and Arg residues of the HBS. These include residues of the N-terminus, helix A, and helix D. Of these, Lys114, Lys125 and Arg129 of helix D contribute ~50%, ~25-33% and ~28-35% of the total binding energy, respectively, and form a coordinated network of interactions.[166-170,230] The domain formed by these three residues is called the pentasaccharide-binding site (PBS, Figure 41). The PBS also contains

other residues that play an important role in heparin binding, including Lys11, Arg13, Arg24, Arg47, Trp47, and Phe122.[231-233] A full-length heparin binds to PBS in antithrombin and to an extended region formed by residues Arg132, Lys133 and Lys136 at the C-terminal end of helix D. This extended region is termed the extended heparin binding site (EHBS, Figure 41). Although the detailed architecture responsible for specificity in antithrombin remains unclear, it can be concluded that DEFGH interaction with antithrombin is specific and well-defined.

3.3.1. Molecular Modeling of Antithrombin-Heparin Specific Interaction

Previous computational attempts to model heparin pentasaccharide binding site on antithrombin using molecular dynamics and docking by both Grootenhuis and van Boeckel[234] and Bitomsky and Wade[119] resulted in geometries that are significantly different from the co-crystal structure. In contrast, with a robust docking protocol based on dual-filter strategy the Desai lab predicted the high-specificity GAG sequences that bind antithrombin with remarkable accuracy.[109] The docking protocol using GOLD program predicted the binding geometry of natural pentasaccharide sequence DEFGH with an "average GAG backbone" conformation to within 2.5 Å (Figure 42).

The docking results suggested that the sulfate groups organized in a specific three-dimensional orientation to afford the specificity of interaction, a conclusion proposed earlier on the basis of a large number of structure-activity studies.[109]

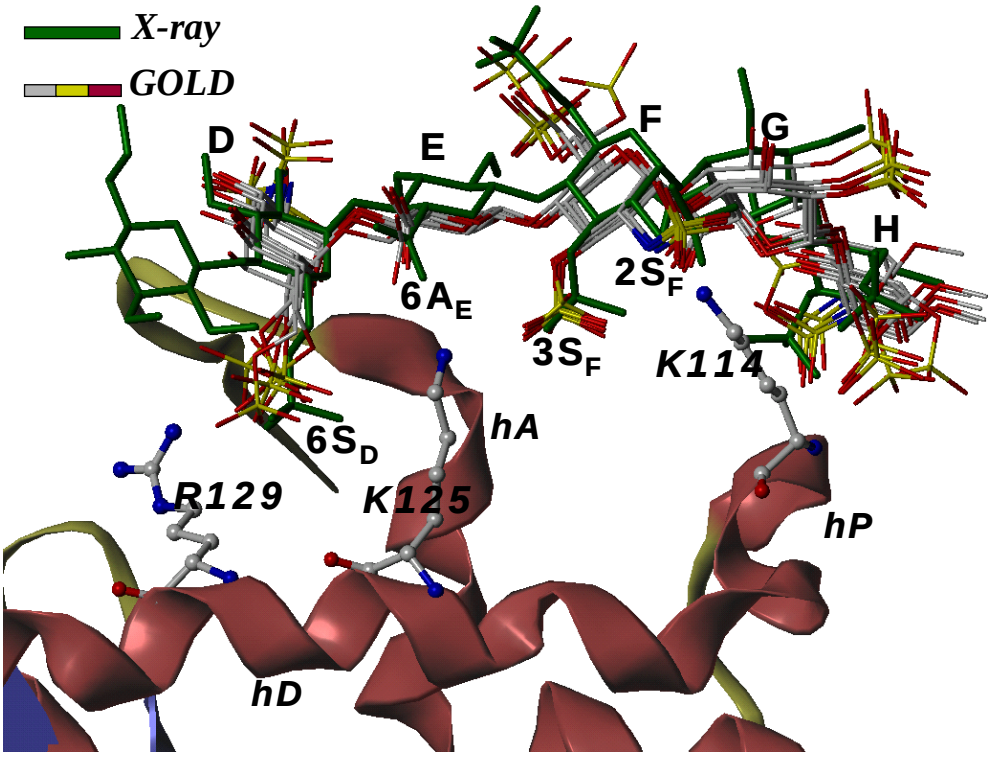


Figure 42. Comparison of GOLD predicted binding geometry of natural pentasaccharide H5 having 'average backbone' with that of $H5_{CRYS}$ determined in the crystal structure. An overlay of 6 solutions from three independent docking runs shows high consistency in the predicted binding geometry, which matches the crystal structure geometry with an RMSD less than 2.5 Å. The structure in green is the crystal structure geometry (DEFGH with an additional residue in the non-reducing end), 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 by *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. Reproduced from Raghuraman, A.; Mosier, P.D.; Desai, U. R. Finding a Needle in a Haystack: Development of a Virtual Screening Method for Identifying High Specificity Heparin/Heparan Sulfate Sequence(s). *J Med Chem.* 2006, 49, 3553-3562.

Raghuraman et al. also docked pentasaccharide variants to antithrombin and found a significant correlation between GOLD scores and binding affinity (slope of 4.2 GOLD score units per kcal/mol and an intercept of 78.9 GOLD score units). The result indicated

that self-consistency of docking geometries is sensitive to sulfate group distribution and pentasaccharide topology. The dual-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'.

While this work by Raghuraman et al. has added a significant contribution in the understanding of the specificity in terms of the GAG-ligand, the counterpart, the protein architecture responsible for such a defined specificity in antithrombin, has not been addressed. Though there are many papers published in detailing the importance of amino acids in heparin binding site, the detail architectural details responsible for specificity remains unclear. However, in one case, an observation was made in reference to the appropriate position of the Arg129 side chain in activated antithrombin due to the H-bond effect of primarily Glu414 and Thr44 to keep the pentasaccharide anchored to the activated state of the inhibitor.[169]

3.4. Thrombin-Heparin: Specific or Non-specific Interaction?

Thrombin is a serine proteinase enzyme and has four important binding sites: the active site, Na^+ binding site, exosite-I and exosite-II (Figure 43a-b). Of these, exosite-I and II are cationic domains located at approximately opposite ends of the thrombin molecule and approximately 10–20 Å away from the active site.[235]

* In this paper, 'specificity' was defined as 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.

Exosite-I, also called anion-binding site-I is located on the right of the active site ('east' in standard view shown in Figure 43a) and is involved in binding fibrinogen. In addition to fibrinogen, several other physiologic and non-physiologic anions bind in exosite-I. For example, fibrin, coagulation factors V, VIII, and XIII, hirudin, bivalirudin, dipetalogastin II, bothrojaracin, and many others interact with exosite-I.

Exosite-II, also called anion-binding exosite-II, is biophysically and spatially discrete from exosite I and binds to highly-charged heparin. Thrombin exosite-II domain include His91, Arg93, Arg101, Arg126, Arg165, His230, Arg233, Lys236, Trp237, and Lys240 (Figure 43b). Thrombin exosite-II also interacts with heparan sulfate, chondroitin sulfate and dermatan sulfate, and glycoprotein Ibα. Thrombomodulin interacts with both exosites simultaneously by a hydrophobic domain in its EGF-like repeats to interact with exosite-I and an anionic chondroitin sulfate sequence to interact with exosite-II.[236]

The binding of ligands in exosites I and II induces significant conformational changes in the active site of thrombin. This allosteric effect has major influence on the specificity and reactivity of thrombin towards its macromolecular substrates.[236-239] For example, the binding of thrombomodulin to exosite I alters the preference of thrombin from fibrinogen to protein C. Similarly, the binding of heparin to exosite II induces a much greater reactivity with antithrombin. Exosite II, being a more positively charged domain than exosite-I, has a preference for sequences that contain a greater proportion of acidic residues. This simple analysis is supported by binding energy contributions measured in solution. Salt dependence studies of the dissociation constant of heparin–thrombin interaction show ~80% contribution from ionic interactions.[227]

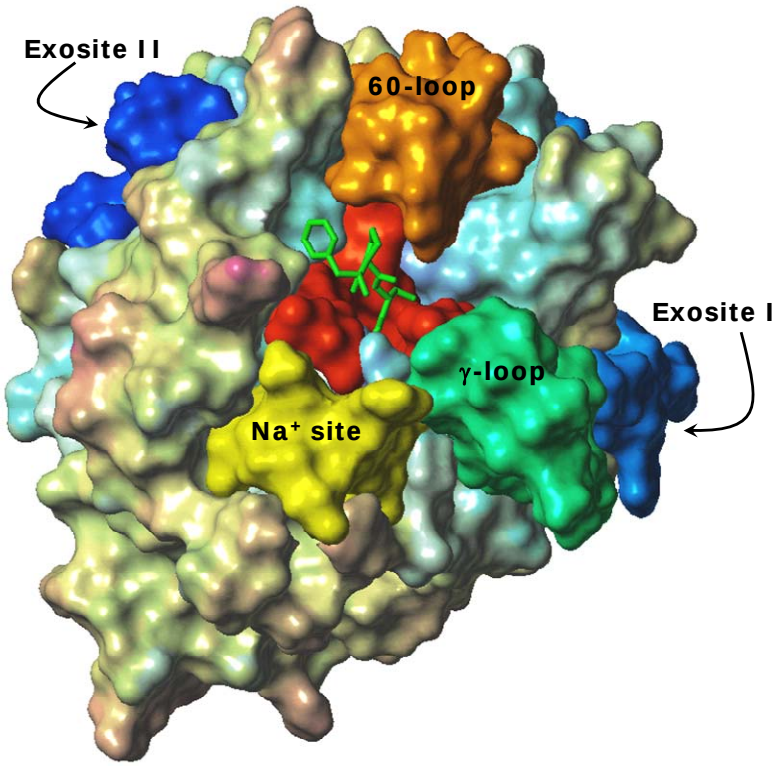


Figure 43a. Topology of thrombin. The thrombin surface is shown in a space-filling representation in the standard orientation using PDB entry 1PPB. Thrombin is bound to the active site inhibitor D-Phe-Pro-Arg-chloromethylketone (PPACK, shown in stick form). The surface shows the orientations of exosites I and II (dark blue surface), the 60-loop (orange surface) and the γ-loop (green surface), and the Na+ binding site (yellow surface). The active site is shown as a red surface.

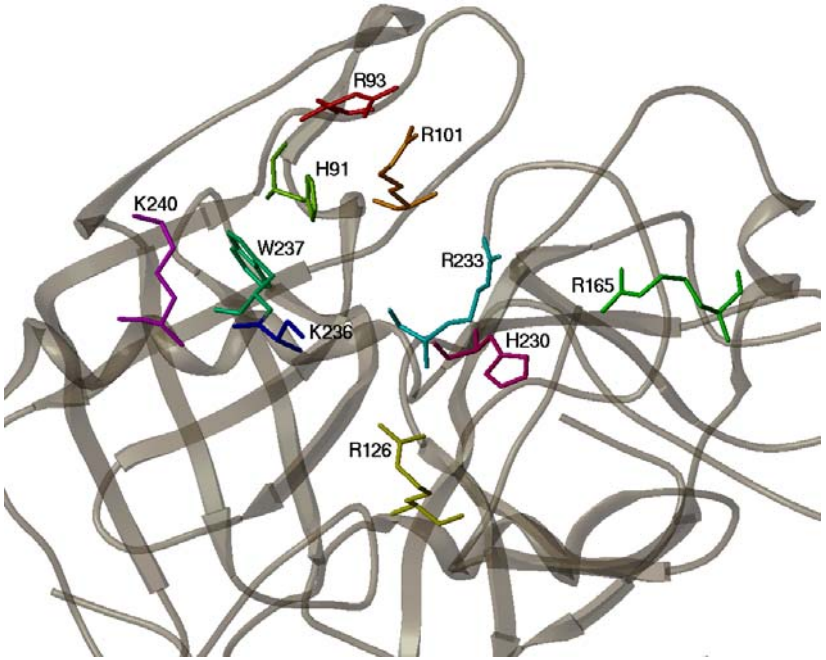


Figure 43b. Thrombin exosite-II key amino acids interacting with heparin octasaccharide. The crystal structure representation is from Protein Data Bank entry 1XMN. Shown is the important amino acids of exosite-II in different colors in AB monomer of the crystal structure.

Exclusive exosite-II ligands GAGs do not alter the catalytic efficiency (k_{CAT}/K_M) of thrombin and thus do not function as direct thrombin inhibitors (DTI).[238,240,241] This 'golden rule' is likely to change with the discovery of some chemo-enzymatically prepared lignin-based macromolecules, which have been found to reduce the k_{CAT} of small peptide hydrolysis by binding in or near exosite-II, the first in this class of DTIs.[242,243]

As discussed earlier heparin interaction at thrombin exosite-II is considered to be non-specific interaction. *In vitro*, both high- and low-affinity heparins bind to thrombin with an apparent K_D of 0.7–2 μM, suggesting that the DEFGH structure is not critical for thrombin binding.[244] The μM affinity of heparin for thrombin arises from the presence of an electropositive domain called exosite-II.

Equilibrium binding studies of heparin binding to thrombin by Olson et al. clearly showed that heparin interaction is a non-specific electrostatic interaction.[227] Their study showed that the strong dependence of the thrombin-heparin binding interaction on NaCl concentration, its minimal dependence on temperature, and the increase in apparent binding affinity with increasing heparin oligosaccharide chain length were best accounted for by a non-specific electrostatic association of thrombin with 5 to 6 anionic residues contained in a 3-disaccharide binding site of heparin. This interaction was characterized by an intrinsic dissociation constant ($K_{D,obs}$) of 6-10 μM at physiological ionic strength (Table 7).[227]

Table 7. Dissociation constants for thrombin-oligosaccharide interactions (Olson et al.).

Oligosaccharide chain length	Non-specific binding model, $K_{D,obs}$ μM
3-disaccharides (6-mer)	7.7±0.9
4-disaccharides (4-mer)	6.5±0.6
5-disaccharides (10-mer)	7.2±0.6
7-disaccharides (14-mer)	5.4±0.4
9-disaccharides (18-mer)	6.3±0.3
~13-disaccharides (26-mer)	10±1

The non-specific binding model of thrombin-heparin interaction is also supported by the crystal structure study of thrombin bound to heparin octasaccharide.[228] The crystal structure of the α-thrombin and octasaccharide co-complex of 1.85 Å crystallographic resolution (PDB entry 1XMN) is an asymmetric unit, consists of four thrombin monomers denoted AB, CD, EF and GH, to reflect the two chains of the human α-thrombin monomer (Figure 44). A heparin octasaccharide is sandwiched between two thrombin monomers, so that the asymmetric unit comprises two nearly equivalent thrombin dimers AB-GH and CD-EF. It should be noted that in the crystal structure 1XMN, only 6 residues of the octasaccharide is resolved in the AB-GH dimer and 5 residues in the CD-EF dimer, and are shown to bind in both directions; one running from the reducing end to the non-reducing end (AB-GH) and the other running from the non-reducing end to the reducing end (CD-EF) (Figure 44). Although the orientation of heparin is different in the two dimers, the amino acids involved are from the same basic patch known as exosite-II.

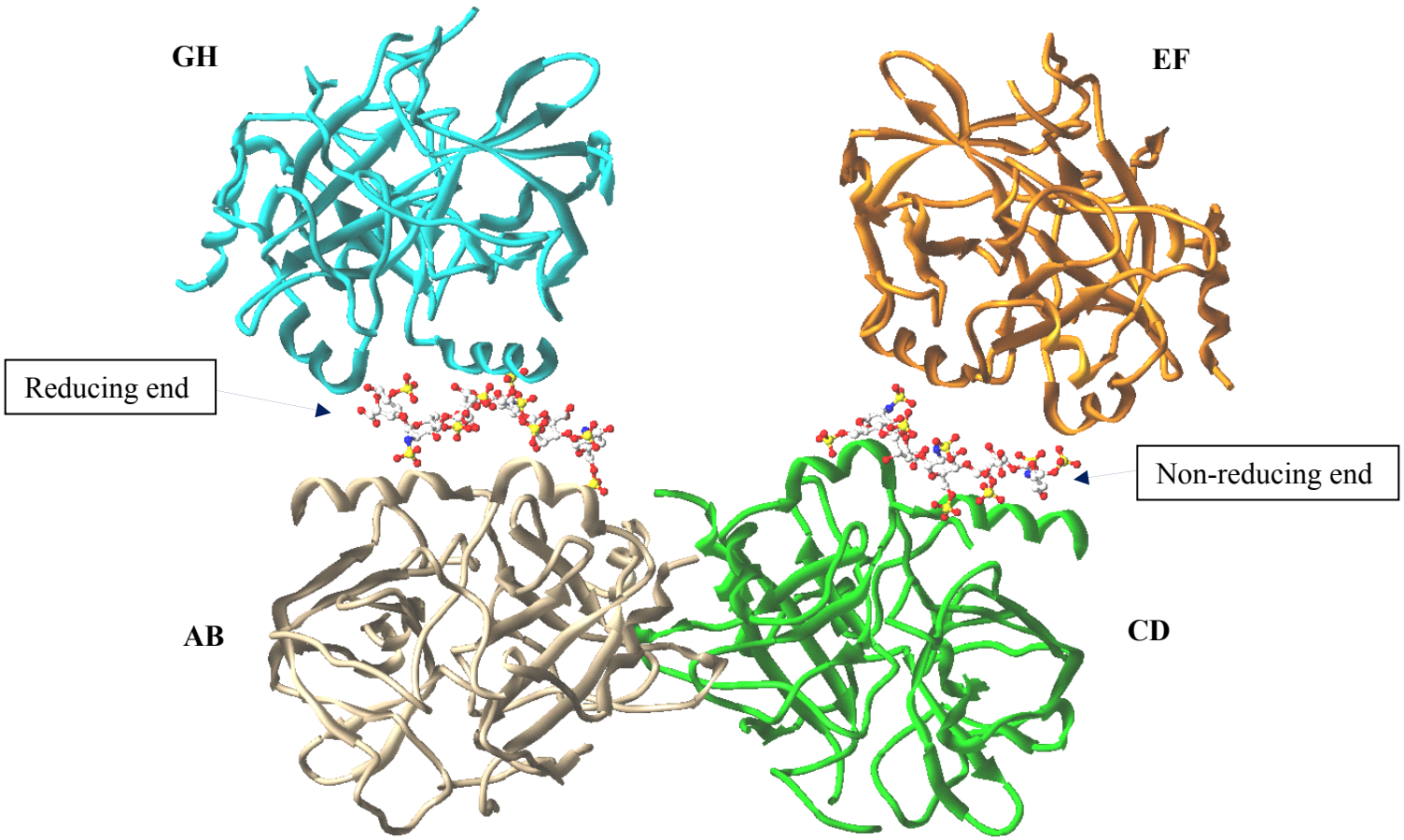


Figure 44. Ribbon representation of thrombin bound to heparin as observed in PDB structure 1XMN. The asymmetric unit consists of four thrombin monomers crystallized into two nearly equivalent dimers AB-GH and CD-EF. The ribbon in gray corresponds to monomer AB, green to CD, orange to EF and cyan to GH.

As discussed earlier, although electropositive residues line the HBS both in antithrombin and thrombin, the antithrombin-heparin interaction is highly specific but the thrombin-heparin interaction appears to be non-specific. We reasoned that although arginine and lysine are the common residues in heparin binding site at thrombin and antithrombin, there must be some critical differences in 3D geometry and topology, flexibility of the side chains and their neighboring amino acids, and/or the symmetry/asymmetry in the binding site. Those differences should determine the nature of the heparin–protein interaction. We have used molecular modeling methodologies to compare and contrast the nature of this unequivocally complex phenomenon of specificity/non-specificity.

3.4.1. Questions Remaining Unanswered

The HBS of antithrombin and thrombin are lined with Arg and Lys residues. Whereas the antithrombin-heparin interaction is specific, the thrombin-heparin interaction is considered to be non-specific. Thus, the questions that arise include: What is the structural and molecular basis for the specificity of the antithrombin-heparin interaction and the non-specificity of the thrombin-heparin interaction? Is it possible to design ligands that are specific for thrombin exosite-II?

3.4.2. Molecular Modeling: Specificity (Antithrombin) versus Non-specificity (Thrombin)

A. Accessibility: Exposed Versus Buried Nature of the Flexible Side Chains

The interaction of heparin with thrombin and antithrombin is primarily a charge-based interaction on the surface of the protein. The important amino acids involved in the heparin binding site of antithrombin are Arg46, Arg47, Lys114, Lys125, Arg129 and Arg132 (Figure 45a). Thrombin exosite-II consists of the basic amino acids Arg93, Arg101, Arg126, Arg165, Arg233, Lys236 and Lys240 (Figure 46a).

As a first step toward identification of the differences in the residues responsible for specific/non-specific interaction, the topology of the residues, in terms of the extent of the surface exposure or the buried nature of the critical amino acids quantitatively, was studied. The crystal structure of thrombin co-complexed with heparin used for this study was the AB monomer of 1XMN, and for antithrombin, it was 1TB6, a ternary complex of thrombin, antithrombin and heparin. Surface areas were calculated using the Fast Connolly Surface in the MOLCAD module of SYBYL for each amino acid either within the context of the protein crystal structure (exposed residue surface area in Table 8) or by itself (total residue surface area in Table 8).

The surface area calculation indicates that in thrombin most of the basic amino acids are more than ~60% surface exposed, except Arg101. In antithrombin, Lys114, Lys125 and Arg129, the key residues considered to be more important in the recognition phase of the antithrombin binding to pentasaccharide, are only 43-50% exposed (Figures 45b and 46b, Table 8). More surface exposure in thrombin allows flexible basic side chains to adapt to structure of the ligand counterpart. However, less surface exposure in antithrombin limits the access, requiring some kind of specific geometry in the ligand for interaction.

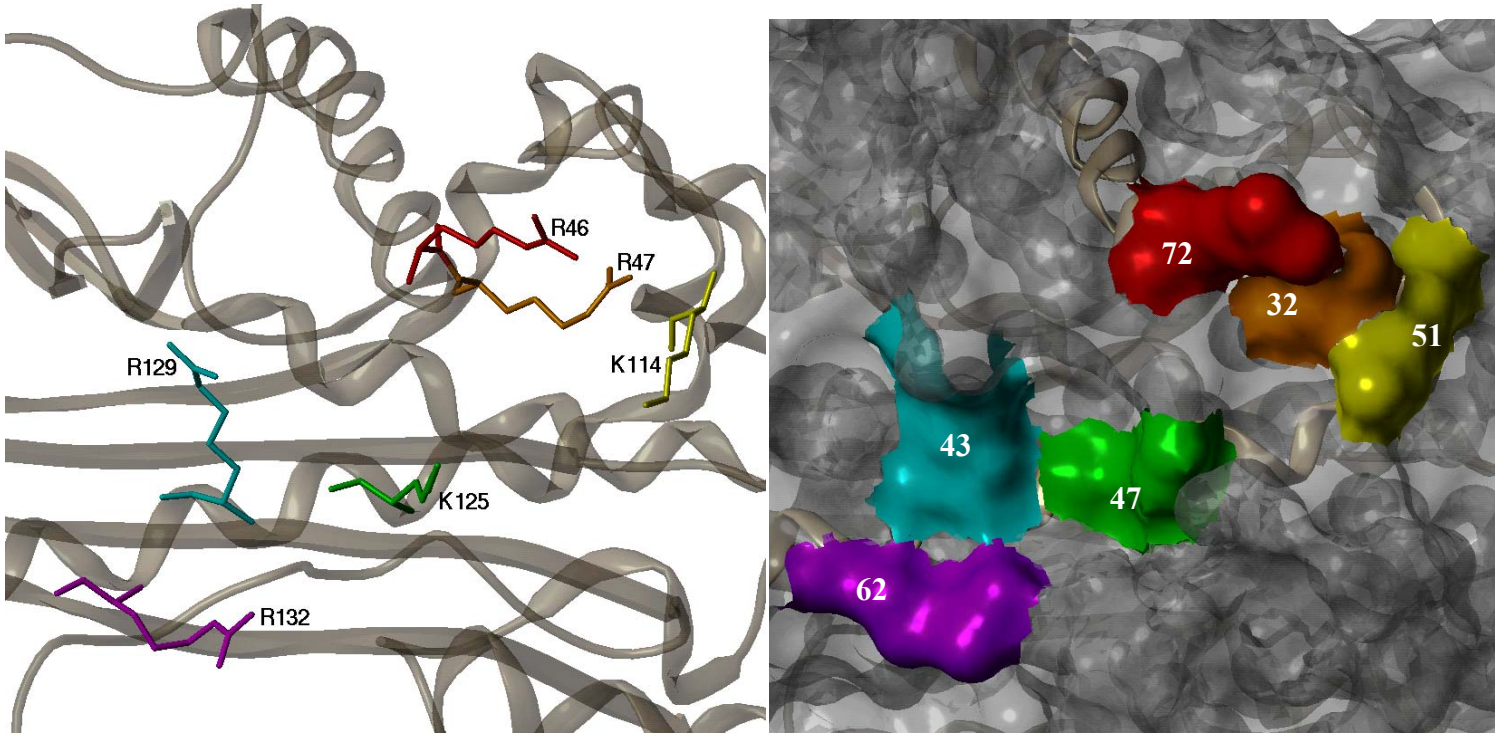


Figure 45a. Antithrombin: Key heparin binding site residues (PDB entry 1TB6); **Figure 45b.** Surface area exposure map shown for critical amino acids in antithrombin, the heparin binding site (Red: Arg46, Orange: Arg47, Yellow: Lys114, Green: Lys125, Cyan: Arg129 and Violet: Arg132)**.** The number on the colored surface is the percent surface exposure of the individual amino acids (Note: Arg132 is not a part of PBS, a subsite within the heparin binding site).

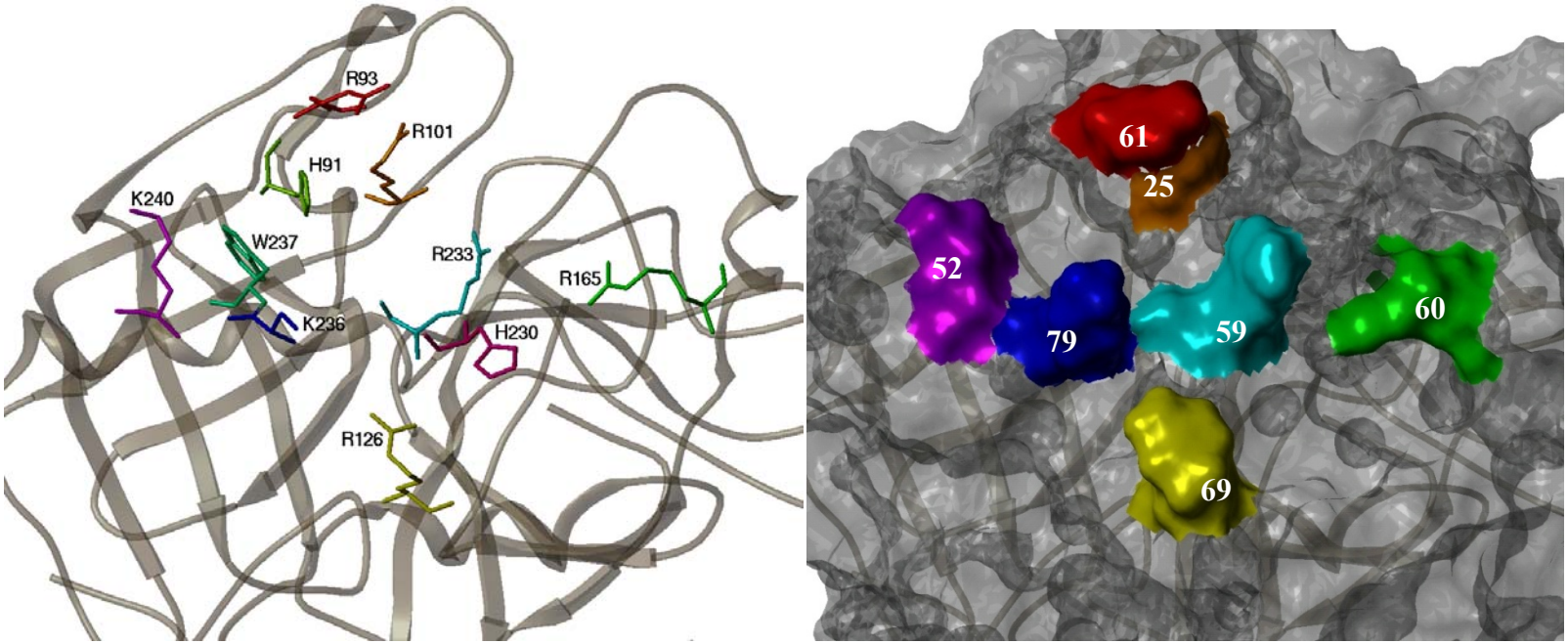


Figure 46a. Thrombin exosite-II: Key heparin binding residues (PDB entry 1XMN); **Figure 46b.** Surface area exposure map shown for critical amino acids in thrombin exosite-II (Red: Arg93, Orange: Arg101, Green: Arg165, Cyan: Arg233, Yellow: Arg126, Blue: Lys 236 and Violet: Lys240). The number on the colored surface is the percent surface exposure of the individual amino acids.

Table 8. Surface area (SA) contribution and radius of gyration (R_g) analyses of the important heparin-binding amino acids in antithrombin and thrombin.

ANTITHROMBIN

Amino acid	Exposed Residue SA	Total Residue SA	% Exposure	No. of observations	R_g	H-Bond partner (s)
R46	122.17	170.82	72	9	3.08	—
R47	54.51	169.37	32	13	0.32	S112, T115
W49	45.59	180.40	25	13	0.86	Aromatic*
K114	77.63	152.24	51	13	0.75	P12, F122[#]
K125	71.52	153.62	47	10	1.87	N45
R129	73.39	169.28	43	12	0.63	T44, E414
R132	101.84	163.08	62	8	3.46	—

THROMBIN

Amino acid	Exposed Residue SA	Total Residue SA	% Exposure	No. of observations	R_g	H-Bond partner (s)
H91	23.45	135.48	17	11	0.50	Aromatic*
R93	102.90	168.59	61	11	2.52	—
R101	42.70	169.43	25	11	0.77	D100
R126	114.37	164.69	69	10	3.10	E127[†]
R165	100.75	169.16	60	11	0.52	M180
H230	24.96	134.89	19	11	0.29	Aromatic*
R233	100.46	171.67	59	11	2.20	—
K236	122.33	154.62	79	7	3.29	—
W237	44.54	179.33	25	11	0.32	Aromatic*
K240	80.04	152.70	52	8	1.81	Q244

*W49, H91, H230, and W237 are in fixed position due to the burial of hydrophobic aromatic ring.

[#] Lys114 is held in place in antithrombin due to the hydrophobic influence of Phe122 and Pro12, which thermodynamically favors Lys144 to stay close to the protein rather than being solvent-exposed.

[†] E127 is adjacent to R126 in thrombin which could theoretically be an H-bonding partner but does not due to the helical turn; they face away from each other.

B. Flexibility: Calculation of Radius of Gyration (R_g)

The flexibility of the long basic amino acid side chains is expected to be another contributing factor in determining specificity/non-specificity. The radius of gyration can be used as a measure of the flexibility in basic amino acids in heparin binding sites. The radius of gyration R_g is the root-mean-square distance (RMSD) of the points from their center of mass (COM), and can be used as a measure of the variability in the position of the basic side chain functionalities in the GAG binding sites (more details in experimental section).

In order to calculate the radius of gyration, we collected from the Protein Data Bank, thrombin and antithrombin crystal structures co-complexed with heparin or heparin-based ligands (Table 9). All antithrombins were aligned to the reference antithrombin monomer from 1TB6 and thrombin monomers were aligned to the AB monomer of 1XMN. The alignments were performed using the Fit Monomers facility of SYBYL using the residues depicted in figures 45a and 46a. When crystal structures were aligned, we found that some amino acid residues are spatially conserved and some are spatially divergent. Calculation of the radius of gyration identified the residues that are spatially conserved in a quantitative manner.

In antithrombin, the pentasaccharide binding site amino acids Arg47 (R_g = 0.3 Å), Lys114 (R_g = 0.8 Å), Arg129 (R_g = 0.6 Å) are highly spatially conserved in a series of antithrombin crystal structures (Figure 47a). Lys125 exhibits modest spatial conservation (R_g = 1.9 Å). Arg46 (R_g = 3.1 Å) and Arg132 (R_g = 3.5 Å, the EHBS amino acid) show a very low degree of spatial conservation.

In thrombin, the Arg93 (R_g = 2.5 Å), Arg126 (R_g = 3.1 Å), Lys233(R_g = 2.2 Å), Lys236 (R_g = 3.3 Å) and Lys240 (R_g 1.8Å) positions are highly variable in a series of thrombin structures indicating that they have high degree of flexibility in contrast to antithrombin heparin binding site basic amino acids (Figure 48a). Interestingly, **Arg101 (R_g = 0.8 Å) and Arg 165 (R_g = 0.5 Å) are highly spatially conserved.** Although Arg233 has R_g = 2.2 Å, it can still be considered to be spatially conserved as (Figure 48a) it is localized in only two groups indicating a bimodal distribution with a high degree of spatial conservation within the two groups.

Although antithrombin and thrombin HBS are lined with Arg and Lys, based on our study, their relative surface exposure and flexibility does not appear to be the same. So we reasoned that there must be some differences in their surrounding amino acids. To find out why some amino acids are spatially conserved and others are not, we closely examined the environment or neighboring amino acids to see if there is anything structurally unique that makes the long rotatable side chains of the key amino acids stay in a fixed conformation. The ones that are spatially conserved had H-bonding partners within H-bond distances (Figures 47b). This hydrogen bonding network anchors the side chain to a particular orientation, which in turn makes some of the binding site basic residues uniquely positioned to engage the ligands with certain structural features, but not others.

Table 9. Crystal structures considered in the thrombin exosite-II and antithrombin PBS basic residue analyses.

PDB ID	Chain					
	T	AT	Description	R[a] (Å)	Missing residues	Ref.
1XMN	AB		Thrombin-Heparin	1.85	K236	(*228*)
	CD				K236, K240	
	EF				R126	
	GH				K236	
3B9F	LH		Thrombin-Protein C Inhibitor-Heparin	1.60	K236, K240	(*245*)
1E0F	AD		Thrombin-Haemadin	3.10		(*246*)
	BE					
	CF					
1JMO*	LH		Thrombin-Heparin Cofactor II	2.20		(*247*)
1TB6	LH		Antithrombin-Thrombin-Heparin	2.50	K240	(*248*)
		I				
2B5T	AB		Antithrombin-Thrombin-Heparin Mimetic (non-productive)	2.10		(*249*)
	CD					
		I			R132	
1SR5		A	Antithrombin-Anhydrothrombin-Heparin (mimetic)	3.27		(*250*)
1T1F*		A	Antithrombin (native)	2.75	R47, K114, K125	(*249*)
		B			R47, K114, K125	
		C			R47, K114, K125	
1AZX		I	Antithrombin (active)-Pentasaccharide	2.90		(*217*)
		L	Antithrombin (latent)-Pentasaccharide			
1E03		I	α-Antithrombin-Pentasaccharide	2.90	K125	(*251*)
		L				
1NQ9		I	Antithrombin-Heparin	2.60	R46, K125, R132	(*250*)
		L			R46, R132	
2GD4		I	Antithrombin-S195A Factor Xa-Heparin	3.30		(*253*)
		C				
3EVJ		I	Antithrombin (Intermediate State)-Natural Pentasaccharide	3.00	R46, K125, R132	(*254*)
		L			R46, R129, R132	

[a]Crystallographic resolution. *1T1F is not included in the calculation of radius of gyration (R_g), an outlier that has incompletely built important amino acids including R47, K114 and K125 and is not an activated form of antithrombin.

*1JMO is not included in the calculation of radius of gyration, an outlier that is not bound to GAG.

In the case of antithrombin, Arg47 (Hbond partners Ser112 and Thr115), Lys125 (Hbond partner Asn45) and Arg129 (Hbond partners T44 and Glu414) have H-bonding partners within H-bond distances. Lys114 is held in place in antithrombin not because of H-bond partners but due to the hydrophobic influence of Phe122 and Pro12, which thermodynamically favors Lys144 to stay closer to the protein rather than solvent exposed (Figure 47b). The same is true with thrombin, the residues that are in fixed orientation like R101 (Hbond partner Asp100), R165 (Hbond partner Met180) and K240 (Hbond partner Gln244) have H-bonding partners within H-bond distances (Figure 48b).

In summary, the heparin binding site amino acids in antithrombin are spatially conserved and contribute to specificity, while in thrombin many of them are highly flexible, which appears lead to considerable non-specificity.

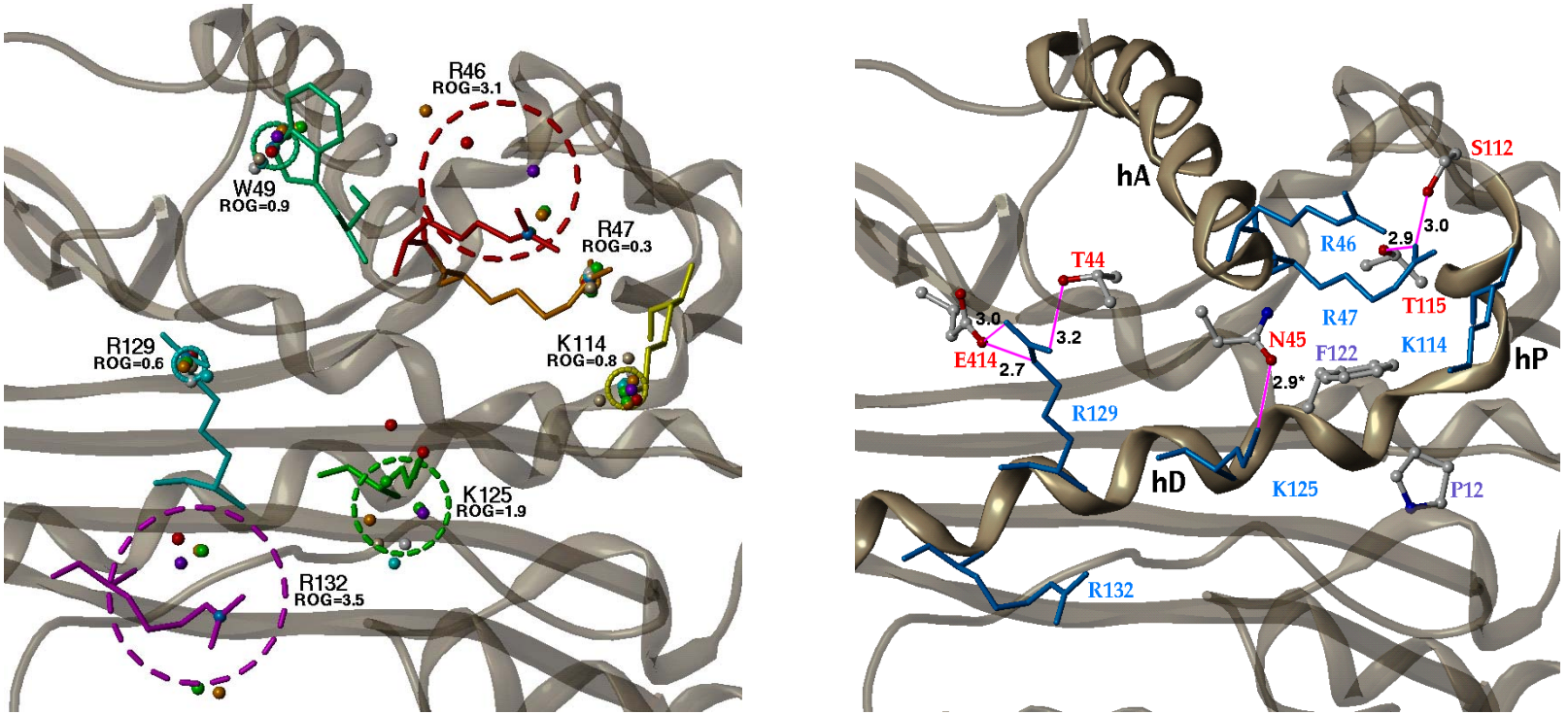


Figure 47a. The side chain conformation variation of the primary basic and other residues in antithrombin pentasaccharide binding site (PBS) for a series of antithrombin crystal structures. The reference structure shown in this picture is from PDB entry '1tb6'. The radius of gyration (ROG) is specified for each residue in angstroms and is shown as a dashed line. Sampled points are shown as small spheres. Color code: blue = 1TB6, green = 1AZX, orange = 1E03, cyan = 1NQ9, red = 1SR5, white = 2B5T, purple = 2GD4, tan = 3EVJ.

Figure 47b. H-bonding partners anchoring key side chains in antithrombin. The crystal structure The line in magenta shows the distance between the hydrogen bonding atoms of the amino acids with distance in Å.*In Asn45, the position of Oxygen and Nitrogen is interchanged from the crystal structure to show the possible H-bond between Asn45 and Lys125.

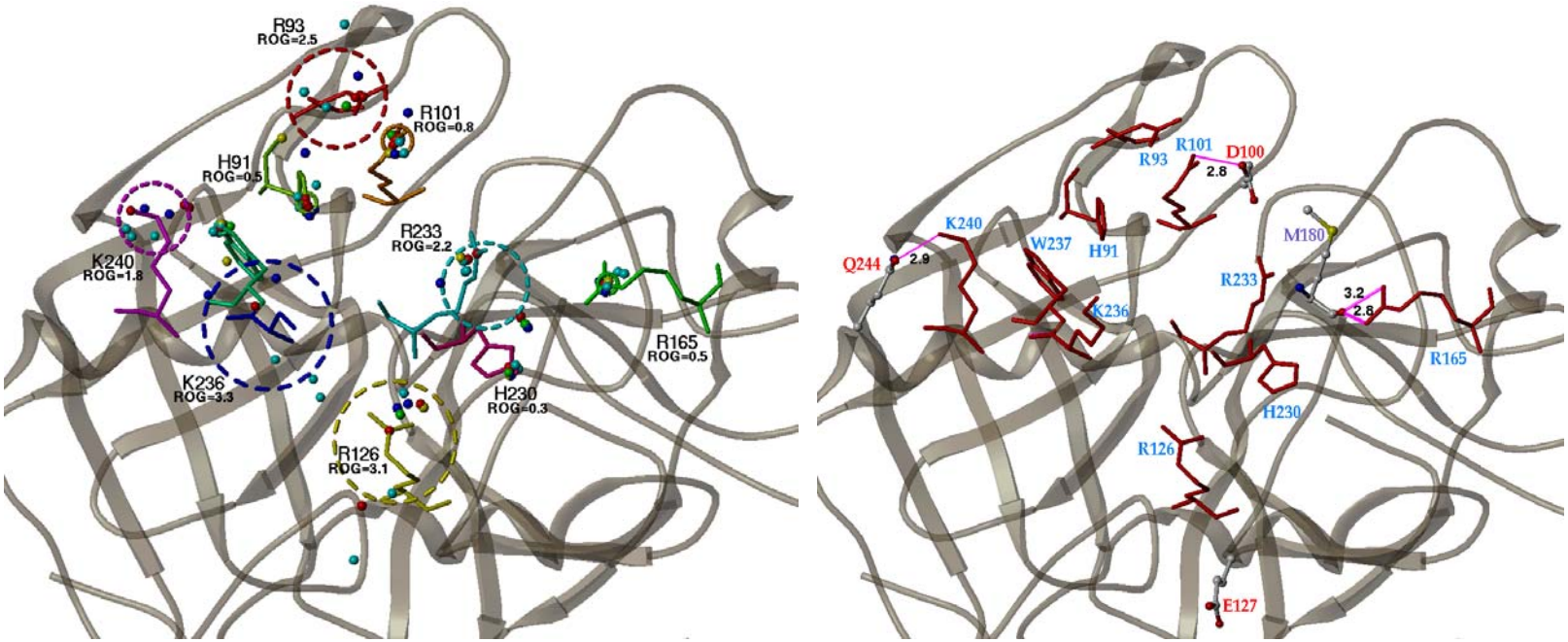


Figure 48a. The side chain conformation variation of the primary basic and other residues in thrombin Exosite II for a series of thrombin crystal structures. The reference structure shown in this picture is from PDB entry '1xmn'. The radius of gyration (ROG) is specified for each residue in angstroms and is shown as a dashed line. Sampled points are shown as small spheres. Color code: red = 1XMN, yellow = 1TB6, green = 3B9F, cyan = 1E0F, blue = 2B5T.

Figure 48b. H-bonding partners anchoring key side chains in thrombin (PDB entry '1xmn'). The line in magenta shows the distance between the hydrogen bonding atoms of the amino acids with distance in Å.

C. Symmetry/Asymmetry

Ionic sulfate-Lys and sulfate-Arg interactions may be approximately treated as equivalent and non-directional. Thus, the spatial arrangement of the basic amino acid residues in the heparin binding sites becomes more important in addition to surface exposure and flexibility in determining specificity. Ligands that recognize a binding site of high symmetry may exhibit more binding modes than those that recognize a lower-symmetry binding site.

In antithrombin, there are three important amino acids involved in heparin pentasaccharide binding including Lys114, Lys125 and Arg129. In thrombin, there are at least seven basic amino acids involved in heparin octasaccharide binding including Arg93, Arg101, Arg126, Arg165, Arg233, Lys236 and Lys240. Figure 49a-b shows some examples of how different geometry can impact the ways in which the ligand can interact with the target protein.

Analysis of the relative geometrical position and shape of the antithrombin binding site shows that it has an element of asymmetry in the binding site (Figure 50a). Lys114, Arg125 and Arg129 forms a scalene triangle that restricts the number of ways in which ligand can interact which favors the specific interaction with pentasaccharide (Note also that Lys125 is somewhat flexible, allowing the three important residues of the PBS to match the isosceles triangle).

In case of thrombin, we see an approximate element of symmetry that explains the apparent contradiction that even with more interacting points instead of providing specificity, thrombin exhibits non-specific interaction with heparin. When roughly divided

by placing a line between R101 and R126, exosite-II shows 2 equivalent points of interaction on either side in a linear fashion (Figure 50b). On the right side of the vertical are R233 and R165 and on the left side are K236 and K240. On the top is R101 and on the bottom is R126. This resembles the square/cross model symmetry with the distribution of positively charged amino acids in a symmetric fashion (Figure 49a). This symmetrical distribution of basic amino acids in exosite-II favors the non-specific interaction of heparin oligomers.

Even though thrombin exosite-II appears to be a target for non-specific heparin interaction, we have identified some spatially conserved amino acids in the upper-right-hand corner of thrombin exosite-II including Arg101, Arg165 and Arg233 that may represent a locally specific sub-pocket within exosite-II. This finding suggests that it may be possible to design specific sequences that recognize the spatially conserved residues in exosite-II. This would likely change the spectrum of future developments in thrombin exosite-II- related research.

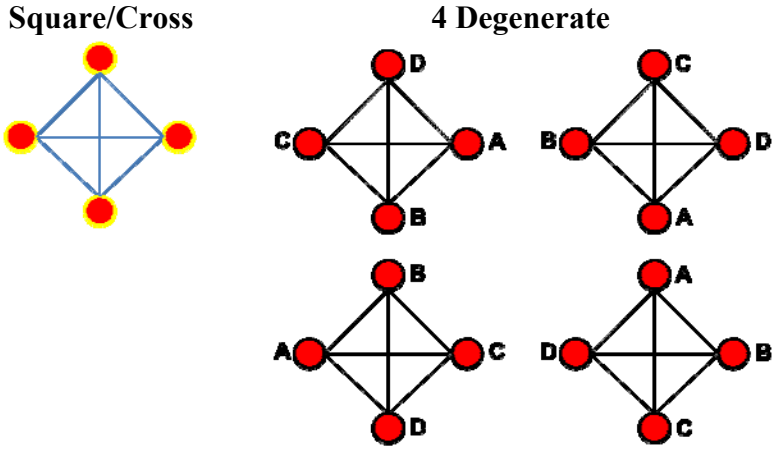


Figure 49a. Symmetric four-point receptor geometry (square/cross) with multiple degenerate binding modes for ligands.

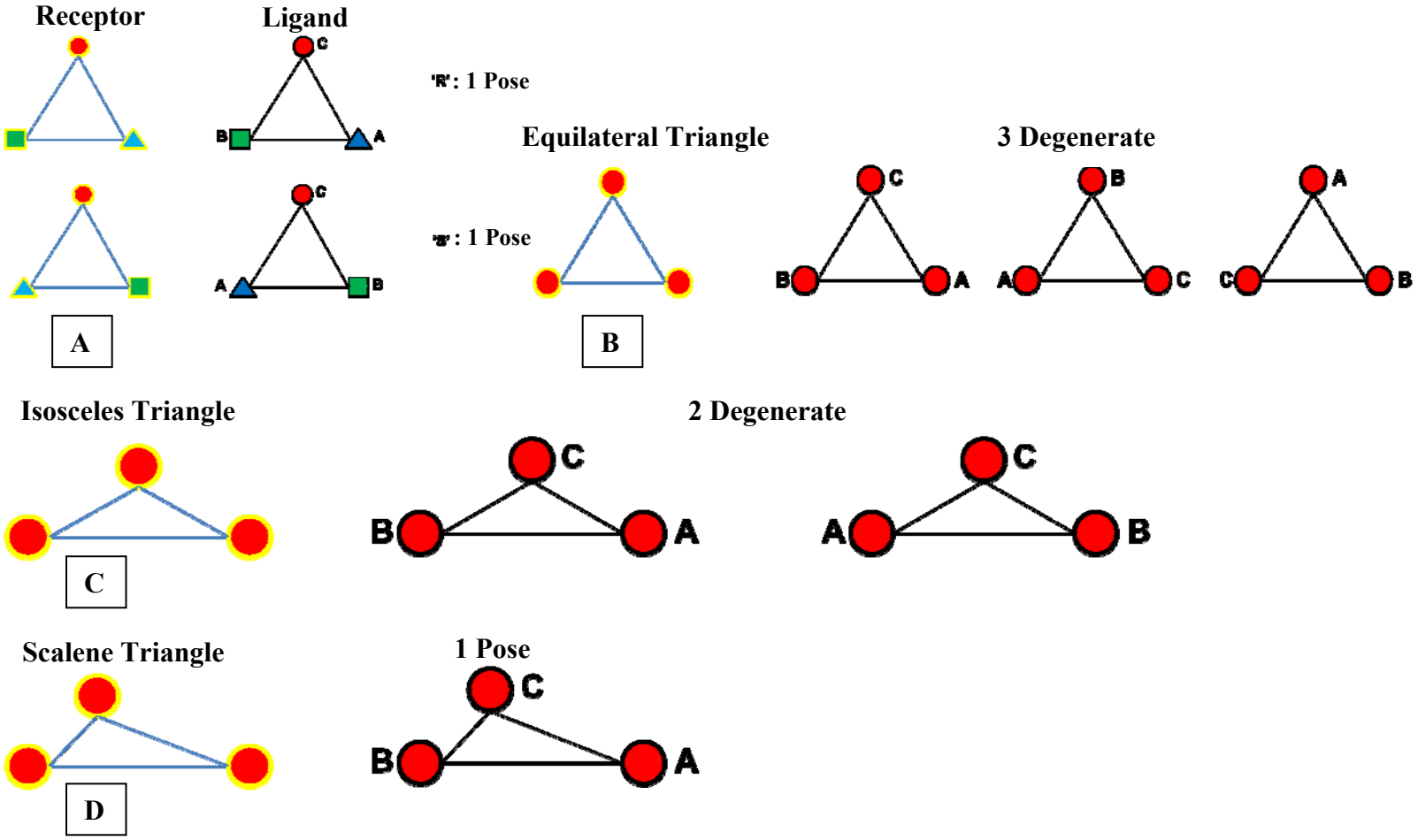


Figure 49b. A. Equilateral triangle receptor ('three-point pharmacophore') showing limited binding mode for the typical *R*- and *S*- isomer of tetrahedral ligands. Three substituents (represented by different shapes) are non-equivalent. **B**. Lys-sulfate or Arg-sulfate kind of non-directional interactions exhibit multiple binding modes in equilateral triangle receptor geometry. **C**. Isosceles and **D**. Scalene triangles show single binding mode.

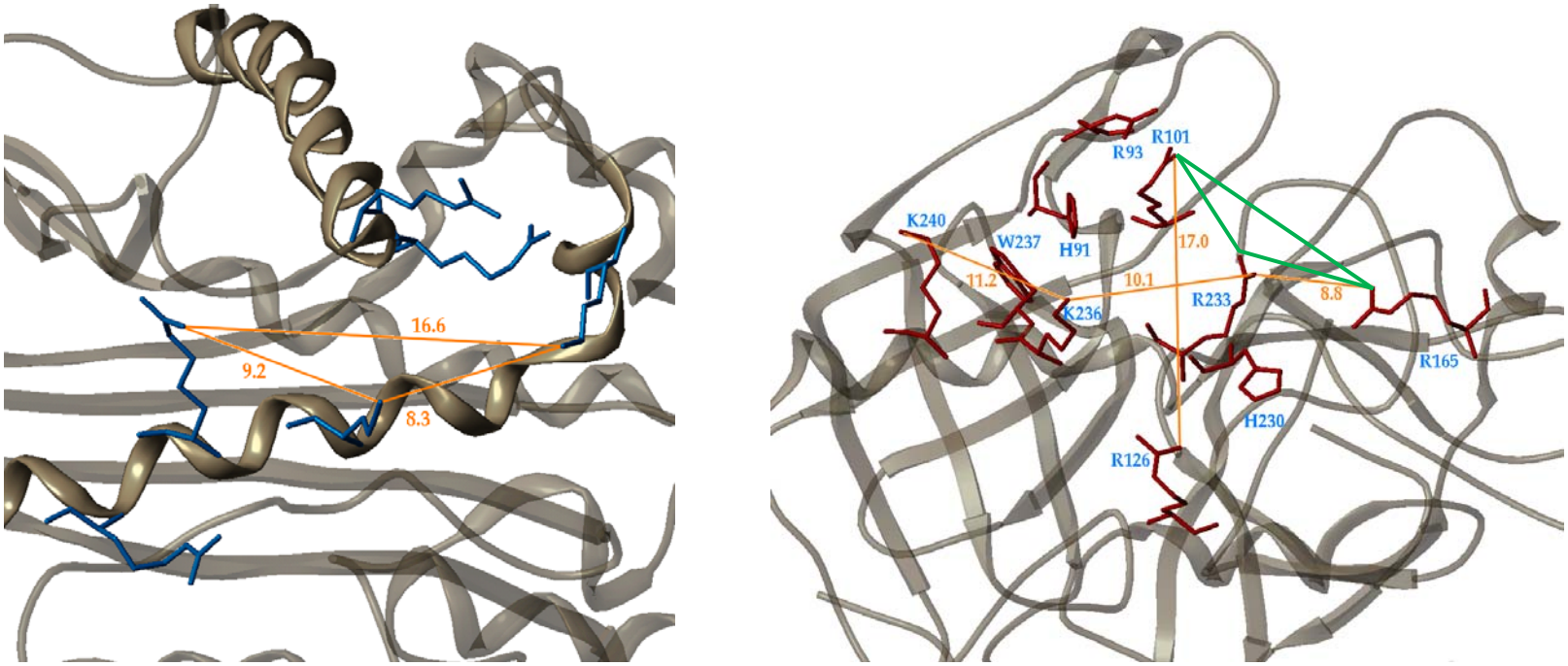


Figure 50a. Asymmetry in antithrombin pentasaccharide binding site (PBS). Scalene triangle geometry is present between three important pentasaccharide binding amino acids Lys114, Lys125 and Arg129 responsible for specificity.

Figure 50b. Element of symmetry in thrombin exosite-II. A vertical line between Arg101 and Arg126 divides exosite-II approximately into two equal parts. Lys236 and Lys240 are on the left side of the horizontal axis. Arg233 and Arg165 are on the right side of the horizontal axis. This element of symmetry favors non-specific interaction. The isosceles triangle (Green lines) between the spatially conserved Arg101, Arg165 and Arg233 on the right side of the symmetry represents a local spatially conserved asymmetry.

3.4.3. Molecular Docking of Library of Heparin Octasaccharide Sequences

A. GOLD Predicts the Binding Geometry of Natural Octasaccharide to Within 3 Å

The crystal structure of α-thrombin and octasaccharide co-complex of 1.85 Å crystallographic resolution from PDB entry 1XMN was used for molecular docking. The crystal structure asymmetric unit consists of four thrombin monomers denoted AB, CD, EF and GH, to reflect the two chains (heavy and light) of the human α-thrombin monomer. A heparin octasaccharide is sandwiched between two thrombin monomers, so that the asymmetric unit comprises two nearly equivalent thrombin dimers AB-GH and CD-EF. It should be noted that in the crystal structure 1XMN, only 6 residues of the octasaccharide are resolved on the AB-GH dimer and 5 residues on the CD-EF dimer. These are shown to bind in both directions: one running from the reducing end to the non-reducing end (AB-GH) and the other running from the non-reducing end to the reducing end (CD-EF) (Figure 44). Since the AB-GH dimer in the crystal structure represents the maximal primary interactions between thrombin and heparin this dimer was considered for docking experiments.[228]

In order to test whether the GOLD docking program is able to dock the crystal structure octasaccharide sequence reasonably within the predefined binding site as in the crystal structure, the natural octasaccharide was docked onto the thrombin dimer in exosite-II. On comparison of docked solution to the backbone of hexasaccharide in crystal structure, the docked solutions were found to be within 3 Å of the crystal structure. Among

the docked poses, the top 6 solutions (2 solutions × 3 docking experiments) were well within 2.5 Å in repeated experiments. Although the docked solution of octasaccharide backbone approximately matches the crystal structure, the directionality is opposite (Figure 51). This preliminary experiment shows that our protocol for docking of GAGs to the thrombin exosite-II reliably reproduces one of the two binding orientations and supports the non-specific nature of the thrombin-heparin interaction.

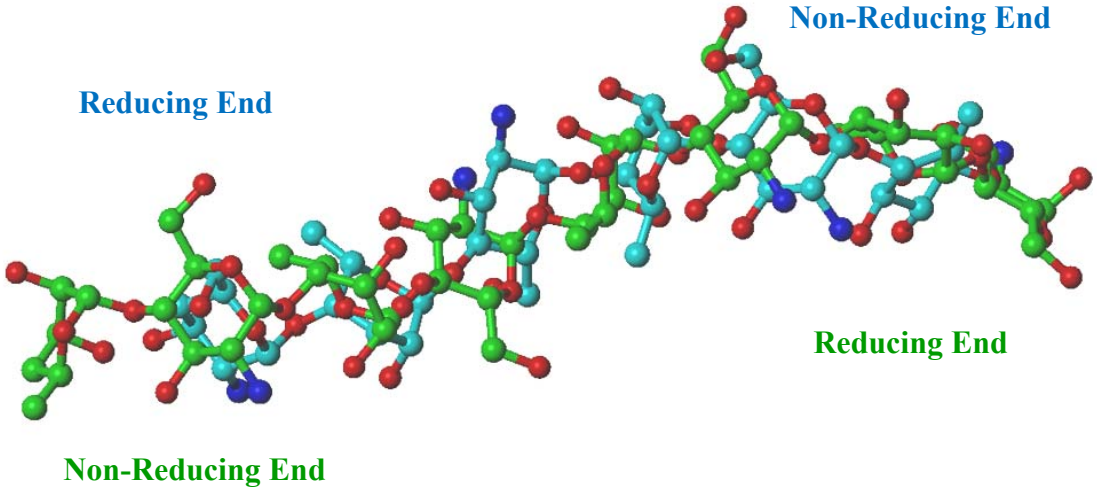
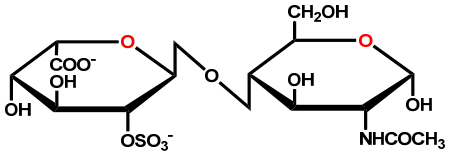


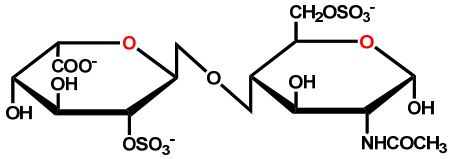
Figure 51. Comparison of the binding orientations of the docked solution of octasaccharide to the observed octasaccharide fragment in the crystal 1XMN. The octasaccharide backbone of the docked solution approximately matches the crystal structure but the directionality is opposite. In the crystal structure, only 6 residues of the octasaccharide are resolved on the AB-GH dimer and 5 residues on the CD-EF dimer, and are shown to bind in both directions: one running from non-reducing end to reducing end and the other running from reducing end to non-reducing end. (See details in experimental section). Docking predicts one of the two binding modes for the octasaccharide. Docked solution of the octasaccharide (green) and the binding mode of the 6 residues of the resolved crystal structure octasaccharide (cyan) are shown as ball-and-stick models.

B. Docking of a Library of Octasaccharide Sequences: Selected Sequences Preferentially Bind to Thrombin.

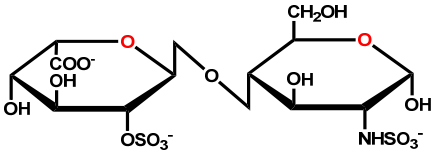
The crystal structure thrombin co-complex with octasaccharide is considered to be a reference in deciding the length of the heparin oligomers (1XMN) to be tested for specificity. As discussed earlier, 48 different disaccharide building blocks of GAGs are possible, of which only 23 are known to exist. If we use all 23 building blocks and include both $^{1}C_{4}$ and $^{2}S_{O}$ conformations for iduronic acid-containing disaccharides (23+13) to build the octasaccharide, the virtual library would result in (36×36×36×36) 1,679,616 sequences (Figure 38b). This is computationally a very challenging task. To limit our preliminary experiment, we have narrowed the number of disaccharide building blocks to 8, based on the common disaccharides present in heparin octasaccharide structure with varying sulfation pattern (Figure 52).



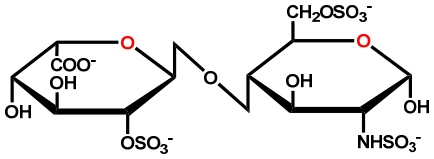
HS14a and HS14a_$^{1}C_{4}$ (Charges: 2)



HS20a and HS20a_$^{1}C_{4}$ (Charges: 3)



HS15a and HS15a_$^{1}C_{4}$ (Charges: 3)



HS21a and HS21a_$^{1}C_{4}$ (Charges: 4)

Figure 52. Building blocks for heparin oligomers used to build the octasaccharide library of 4096 sequences. Disaccharide units are named as HS+number+a where 'a' denotes an additional sulfate group at 2nd position of the iduronic acid. Since iduronic acid can exist in $^{1}C_{4}$ or $^{2}S_{O}$ forms, a suffix $^{1}C_{4}$ is added. No suffix indicates $^{2}S_{O}$ conformation.

The octasaccharide library of 4096 sequences was built in SYBYL in an automated manner using an in-house SPL (SYBYL Programming Language) and was docked to the AB-GH thrombin dimer using GOLD v3.0 (See experimental section for details). In the first phase of the docking program, GOLD filtered high-affinity sequences from the 4096 sequences, whose modified GOLDScore followed a Gaussian distribution. These sequences were then subjected to the specificity filter (Figure 53).

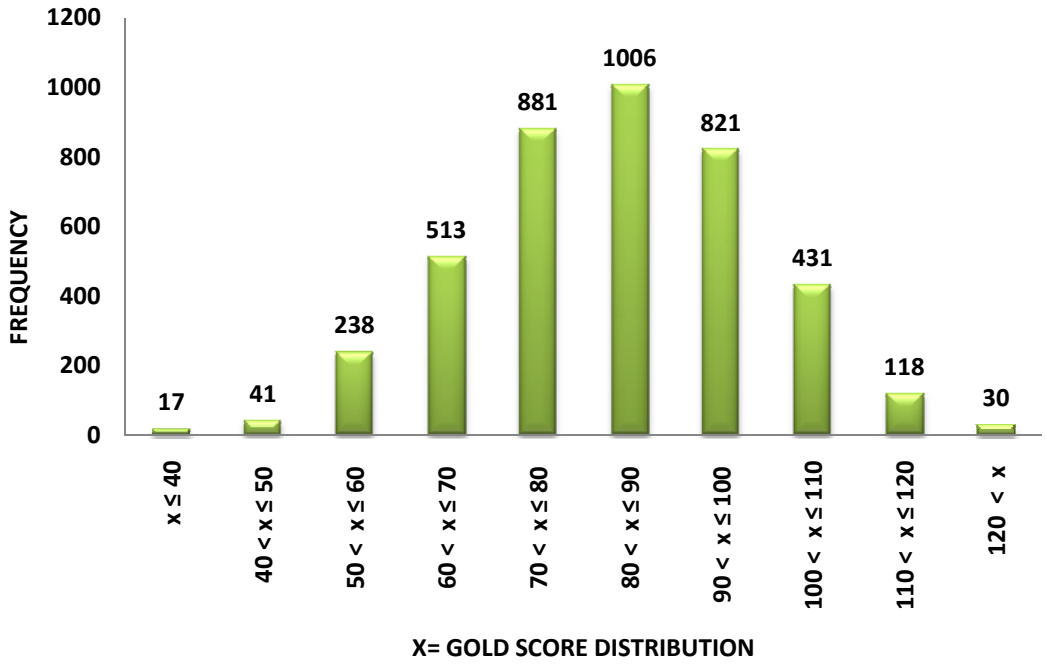


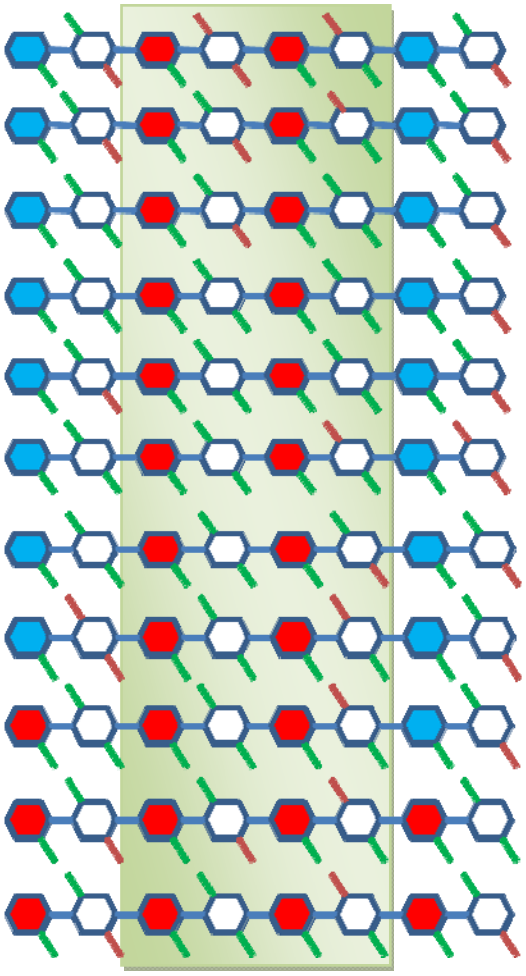
Figure 53. Histogram showing the distribution of 4096 octasaccharide sequence modified GOLDScore following the first phase of combinatorial library screening with thrombin.

The top-scoring sequences (~1%) from the first phase are then docked in triplicate under more rigorous docking conditions. Application of the specificity filter (top 2 solutions from each experiment must be within 2.5 Å) resulted in 14 hit sequences for thrombin dimer (Table 10).

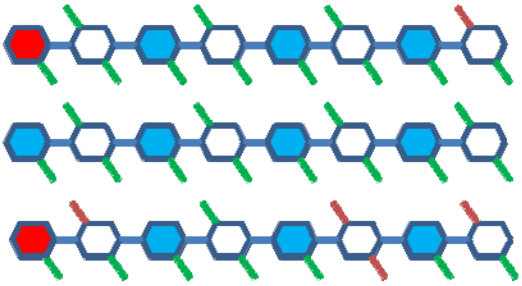
Table 10. Modified GOLDScore for hit octasaccharide sequences of a library of 4096 heparin oligomers. The score in the table represents the maximum score for 3 independent docking runs.

HIT SEQUENCES	Run 1	Run 2	Run 3	Max. score
HS20a_1C4__HS14a__HS15a__HS20a_1C4	112	129	138	138
HS20a_1C4__HS20a__HS15a__HS20a_1C4	134	110	137	137
HS21a_1C4__HS20a__HS21a__HS20a_1C4	134	114	119	134
HS21a_1C4__HS21a__HS21a__HS20a_1C4	131	118	121	131
HS21a__HS21a_1C4__HS21a_1C4__HS15a_1C4	108	128	108	128
HS20a_1C4__HS21a__HS21a__HS20a_1C4	120	126	112	126
HS21a_1C4__HS21a_1C4__HS21a_1C4__HS21a_1C4	116	101	126	126
HS21a_1C4__HS21a__HS15a__HS14a_1C4	121	124	112	124
HS21a_1C4__HS21a__HS20a__HS20a_1C4	107	122	112	122
HS15a__HS21a_1C4__HS14a_1C4__HS15a_1C4	121	93	94	121
HS21a__HS21a__HS15a__HS20a_1C4	120	121	106	121
HS14a_1C4__HS21a__HS15a__HS20a_1C4	110	117	106	117
HS20a__HS20a__HS15a__HS21a	109	103	110	110
HS20a__HS21a__HS15a__HS20a	108	107	110	110

These sequences fall into two general pattern of binding orientation, one with non-reducing end to reducing end **(Binding mode A)** and another with reducing end to non-reducing end **(Binding mode B)** (Figure 54)**.** This is fully in consistent with what is observed in the crystal structure 1XMN wherein the AB-GH and CD-EF dimers of thrombin show opposite orientations for the octasaccharide in exosite-II.



Sequences with binding mode A



Sequences with binding mode B

Red-filled hexagon: Iduronic acid with $^{2}S_{o}$ conformation; Blue-filled hexagon: Iduronic acid with $^{1}C_{4}$ conformation; Unfilled hexagon: Glucosamine; Solid green line: Presence of *N*-sulfate or *O*-sulfate; Solid red line: Absence of *N*-sulfate or *O*-sulfate.

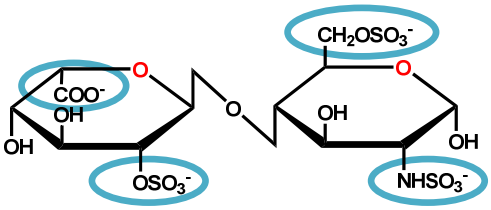
Figure 54. Cartoon representing the hit sequences identified by docking of the octasaccharide library shows 2 binding modes in thrombin exosite-II

Out of the final 14 sequences, 11 of them follow binding mode A. Although it seems that both binding geometries are possible, one binding mode is preferentially selected over the other. A comparison of the binding modes and structural features of the sequences show that sequences having the $^{1}C_{4}$ conformation for units 3 and 5 adopt binding mode B. On the other hand, 11 sequences adopt binding mode A and are uniquely different by having the $^{2}S_{O}$ conformation at units 3 and 5. This subtle difference in the

conformation reverses the direction in which octasaccharide binds to thrombin at exosite-II. Just as observed in the crystal structure, most of the octasaccharide interactions are observed with the AB monomer of thrombin and few interactions with the GH monomer (R93 and K240 only).

C. Design of Smaller Sequences for Thrombin Exosite-II

Molecular docking of the library of octasaccharides showed that the important critical interactions are from the central part of the octasaccharide while the end residues flank the thrombin exosite-II. Based on this observation, we decided to build a small library of tetra-, hexa-, and octasaccharides based on the highly-sulfated disaccharide shown in Figure 55 to test if the shorter sequences will be able recognize the same binding mode. If so, it may serve as a framework to design 'specific' sequences that recognize the spatially conserved residues in exosite-II (see Section 3.4.1).

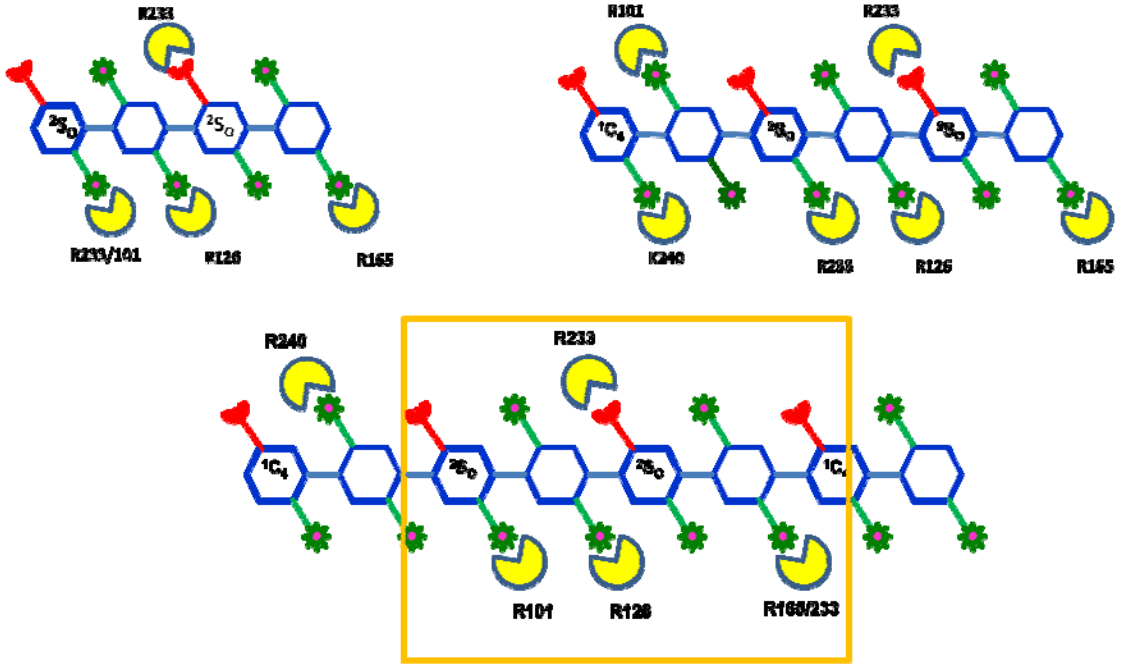


HS21a and HS21a_$^{1}C_{4}$

Figure 55. Structure of a highly-sulfated iduronic acid containing disaccharide. Both $^{1}C_{4}$ and $^{2}S_{O}$ forms are considered. Negatively charged groups are indicated with ovals.

The small library consisting of 4 tetrasaccharide, 8 hexasaccharide and 16 octasaccharide sequences was docked in triplicate to exosite-II of the AB-GH dimer of

thrombin crystal structure 1XMN. The 2 top-scoring solutions of each hit tetra-, hexa-, and octasaccharides were within 2.5 Å in triplicate experiments. The analysis of binding modes shows that, regardless of the size of the sequences, there exists a common binding motif for these sequences where certain interactions are always consistent in each case (Figures 56 and 57).



SEQUENCE	NO. OF CHARGES	GOLD SCORE
TETRA_PER_SO_3	8	118
HEXA_PER_SO_3	12	155
OCTA_PER_SO_3	16	145

Figure 56. Cartoon showing the hit tetra-, hexa-, and octasaccharides and their interaction with thrombin exosite-II amino acids. The tetrasaccharide sequence is a substructure contained within the octasaccharides and maintains the same interaction profile at thrombin exosite-II regardless of the size of the sequence. From this, a core tetrasaccharide pharmacophore is elucidated. The table lists the hit sequences, the number of charges and their corresponding modified GOLDScores. A green star represents *O*-sulfate or *N*-sulfate, and a red cup represents carboxylate.

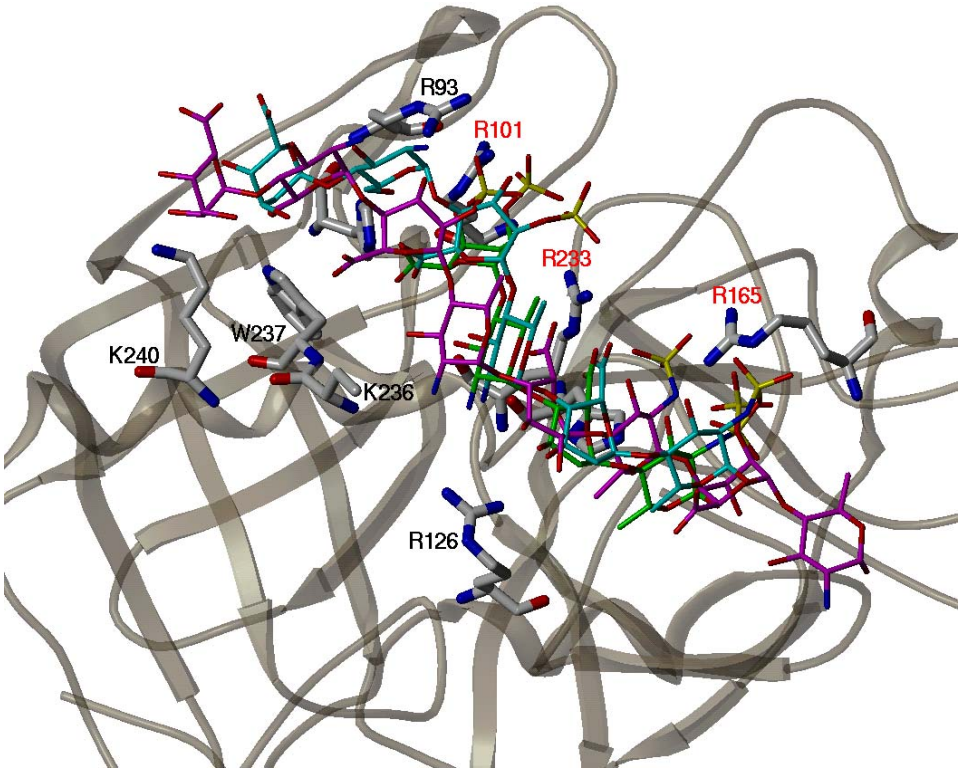


Figure 57. Docking poses of tetra- (green), hexa- (cyan) and octasaccharide (magenta) sequences in thrombin exosite-II. All of them have similar binding mode running from non-reducing end to reducing end. The ribbon representation of the AB monomer of the crystal structure 1XMN is displayed (monomer GH not displayed for clarity). Thrombin exosite-II amino acids are shown in capped stick representation.

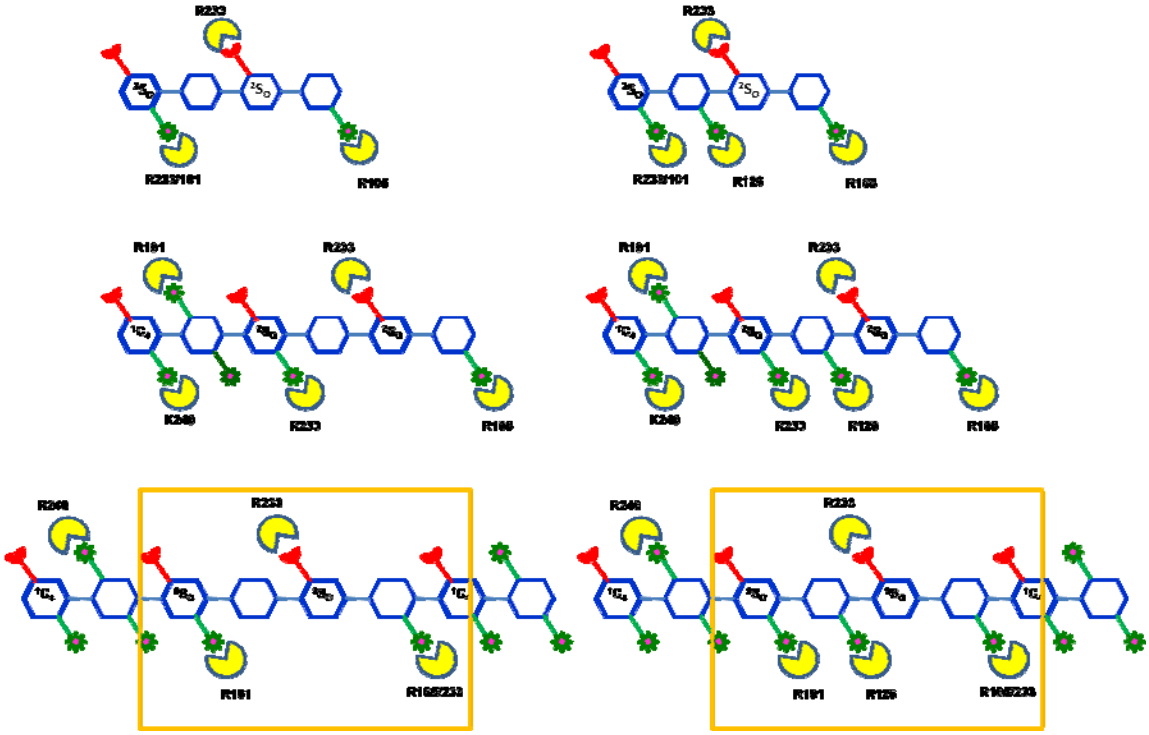
A closer look at these sequences and their interaction with thrombin exosite-II amino acids revealed that it may be possible to remove certain non-interacting redundant sulfate groups in the central core tetrasaccharide pharmacophore. But is it possible to remove these functionalities and still retain the same binding mode? To study the impact of the absence of these non-interacting groups in determining the binding mode, we have removed certain non-interacting sulfate groups in these hit sequences in the tetrasaccharide

region and docked them in triplicate to thrombin exosite-II (Figure 58). The result showed that, these sequences recognize the same amino acids and were docked consistently in the same binding mode even after removing some charges, which confirmed our optimism that designed sequences would specifically recognize spatially conserved amino acids in thrombin exosite-II (Figures 58 and 59). It is important to recall that this interaction has been considered to be a non-specific interaction.

Table 11. Modified GOLDScore for the sequences docked consistently in thrombin exosite-II.

HIT SEQUENCES	MODIFIED GOLDScore	AVERAGE*
TETRASACCHARIDE		
HS21a__HS21a_$^{1}C_4$	108	113.0
HS21a_$^{1}C_4$__HS21a	118	
HEXASACCHARIDE		
HS21a__HS21a__HS21a	142	
HS21a__HS21a__HS21a_$^{1}C_4$	135	
HS21a__HS21a_$^{1}C_4$__HS21a_$^{1}C_4$	134	141.5
HS21a_$^{1}C_4$__HS21a_HS21a	155	
OCTASACCHARIDE		
HS21a__HS21a__HS21a__HS21a	123	
HS21a__HS21a__HS21a__HS21a_$^{1}C_4$	132	
HS21a__HS21a_$^{1}C_4$__HS21a_$^{1}C_4$__HS21a_$^{1}C_4$	125	
HS21a_1C4__HS21a__HS21a__HS21a	132	126.0
HS21a_$^{1}C_4$__HS21a__HS21a__HS21a_$^{1}C_4$	145	
HS21a_$^{1}C_4$__HS21a_$^{1}C_4$__HS21a__HS21a	111	

*The GOLD score average for the subset of sequences of a given length.



HIT SEQUENCES	MODIFIED GOLDScore
TETRASACCHARIDE	
HS14a__HS15	100
HS15a__HS15	106
HEXASACCHARIDE	
HS21a_$^{1}C_4$__HS14a__HS15	111
HS21a_$^{1}C_4$__HS15a__HS15	127
OCTASACCHARIDE	
HS21a_$^{1}C_4$__HS14a__HS15__HS21a_$^{1}C_4$	112
HS21a_$^{1}C_4$__HS15a__HS15__HS21a_$^{1}C_4$	126

Figure 58. Cartoon showing the hit tetra-, hexa-, and octasaccharides with fewer charges and their interaction with thrombin exosite-II amino acids. The tetrasaccharide sequence is also a part of hexa-, and octasaccharide which is maintaining the same interaction profile at thrombin exosite-II even after removing some of the redundant charges from the sequences. The accompanying table lists the sequences and their modified GOLDScores.

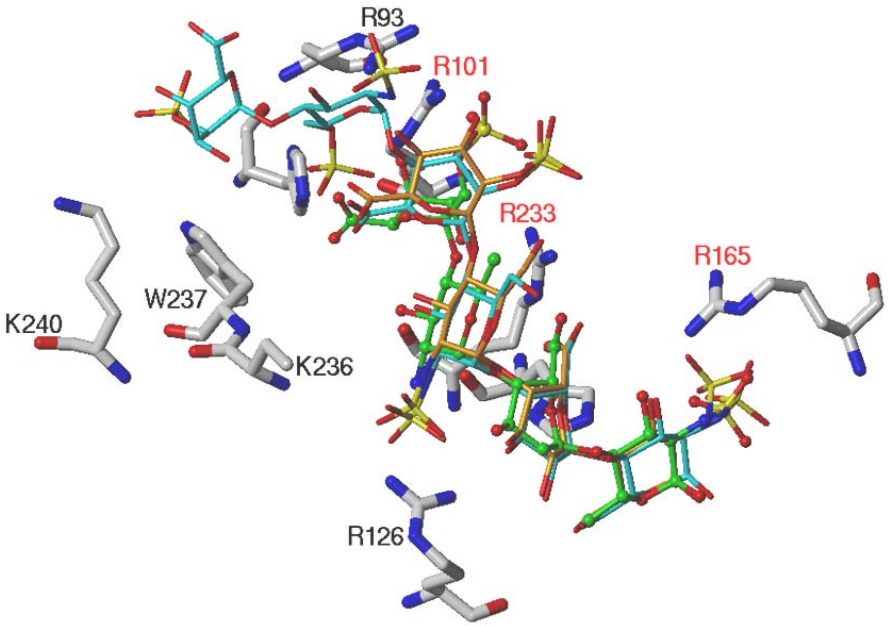


Figure 59. Docking poses of tetra- (green), hexa- (cyan) and octasaccharide (not shown for clarity) sequences in thrombin exosite-II. These sequences are stripped of some redundant charges and still maintain the same binding mode just like their highly-sulfated counterparts. All of them have a similar binding mode running from non-reducing end to reducing end. Important thrombin exosite-II amino acids in the AB monomer of the crystal structure 1XMN are displayed in capped stick representation. These are the sequences with only the "minimally required" sulfate groups (in the conserved tetrasaccharide region), displayed without the ribbon.

These sequences with only the "minimally required" sulfate groups not only docked consistently in repeated experiments but also recognized the spatially conserved Arg101, Arg165, and Arg233, which approximate isosceles triangle geometry in thrombin exosite-II. This indicates that an optimally designed heparin sequences may bind to thrombin exosite-II in a specific manner (Figure 59).

D. Designing 'High Affinity' and 'High Specificity' Tetrasaccharide Sequences as Thrombin Exosite-II Modulators

Although it is presumed that the heparin-thrombin interaction is non-specific, our molecular modeling study showed that there is a region within exosite-II with significant spatial conservation including amino acids Arg101, Arg165 and Arg233. These amino acids assume isosceles triangle geometry with a local asymmetry in a symmetrical thrombin exosite-II. Docking experiments showed that a unique tetrasaccharide sequence was docked consistently in the same binding mode in exosite-II recognizing the above mentioned amino acids. The binding mode for this unique tetrasaccharide is consistent in multiple docking experiments.

This observation of specific binding mode of tetrasaccharide sequence is not changed even after removing the non-interacting groups (redundant charges) on the tetrasaccharide. When the binding modes of tetrasaccharide sequences were analyzed, we found that the 2-*O*-sulfate in the non-reducing end iduronic acid was not close enough to effectively H-bond with Arg101, but instead faced Arg233. Based on distance and geometry calculations, we hypothesized that a 3-*O*-sulfate in the same ring would be optimal to gain interaction with Arg101 and additionally to Arg93. To test this hypothesis we designed the following tetrasaccharide sequences where in one case we simply substituted 2-*O*-sulfate by 3-*O*-sulfate and in another case added 3-*O*-sulfate in addition to the existing 2-*O*-sulfate in the non-reducing end (Figure 60).

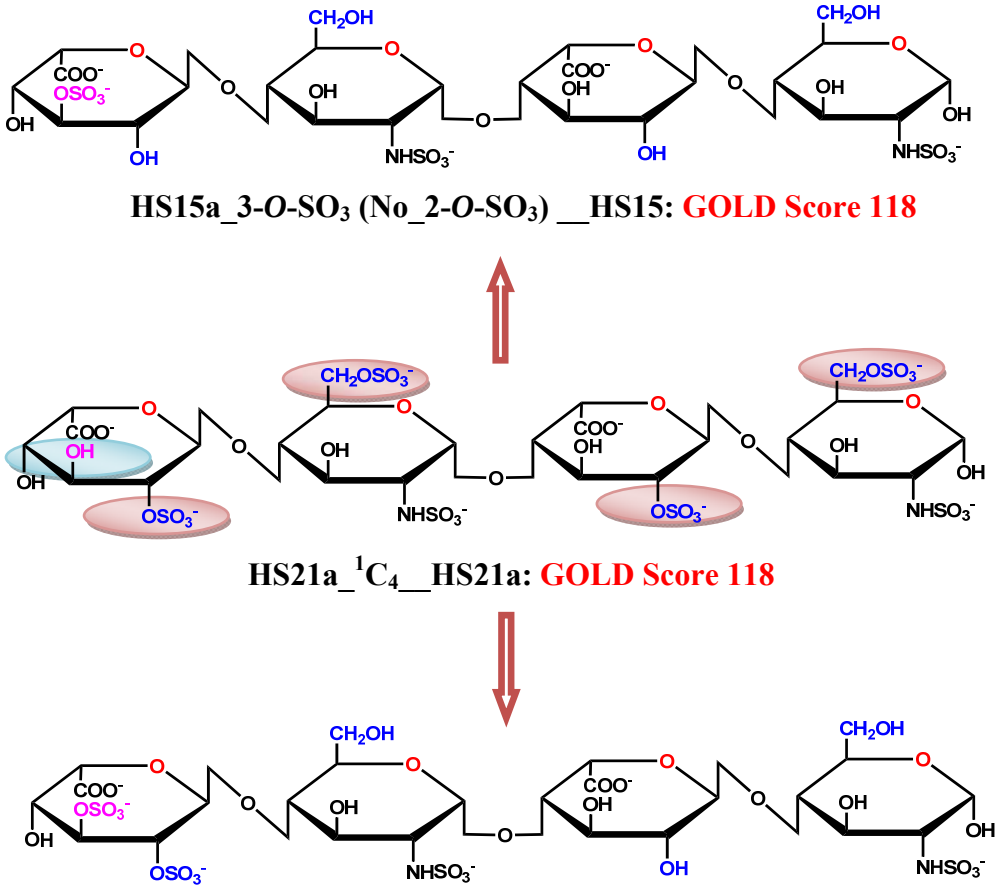


Figure 60. 'High affinity' and 'high specificity' tetrasaccharide sequences. The sequence in the middle is the highly-sulfated sequence with 8 charges that served as a pharmacophore in designing sequences with fewer charges. Appropriately placing the charges in the tetrasaccharide framework yields sequences of equal or higher GOLD score. The circled positions were modified in generating new sequences.

Docking of these designed sequences with fewer charges resulted in equal or higher GOLD scores in repeated experiments and each time when they were docked they docked in the same binding mode with less than 1 Å RMSD (Figure 61). As shown in Figures 60 and 62 the initial hit tetrasaccharide has 8 charges and has a GOLD score of 118. The sequences we have designed are the same size but with 5 or 6 charges; these have equal or

higher GOLD score (127) and predicted to recognize the same binding mode including the spatially-conserved R101, R165, and R233 and also make accessory interactions with R126 and R93, making a strong case for 'specific' exosite-II modulators.

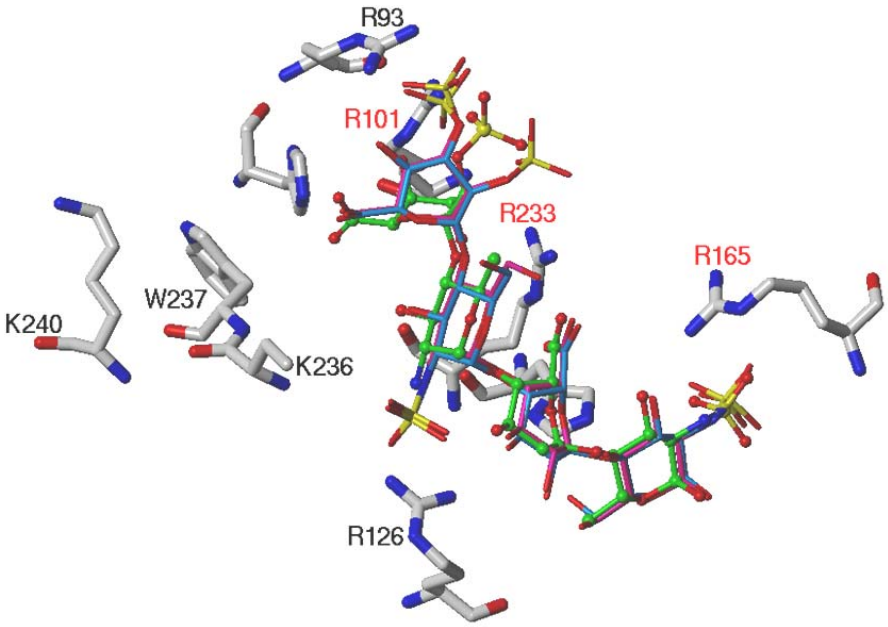


Figure 61. Docking poses of 'high affinity' and 'high specificity' tetrasaccharide sequences shown in green (ball-and-stick) is the highly-sulfated reference structure. Blue and magenta (capped stick) are the sequences with fewer charges. Sequences are docked to within 1 Å RMSD. Important thrombin exosite-II amino acids in the AB monomer of the crystal structure 1XMN are displayed in capped stick without the backbone ribbon.

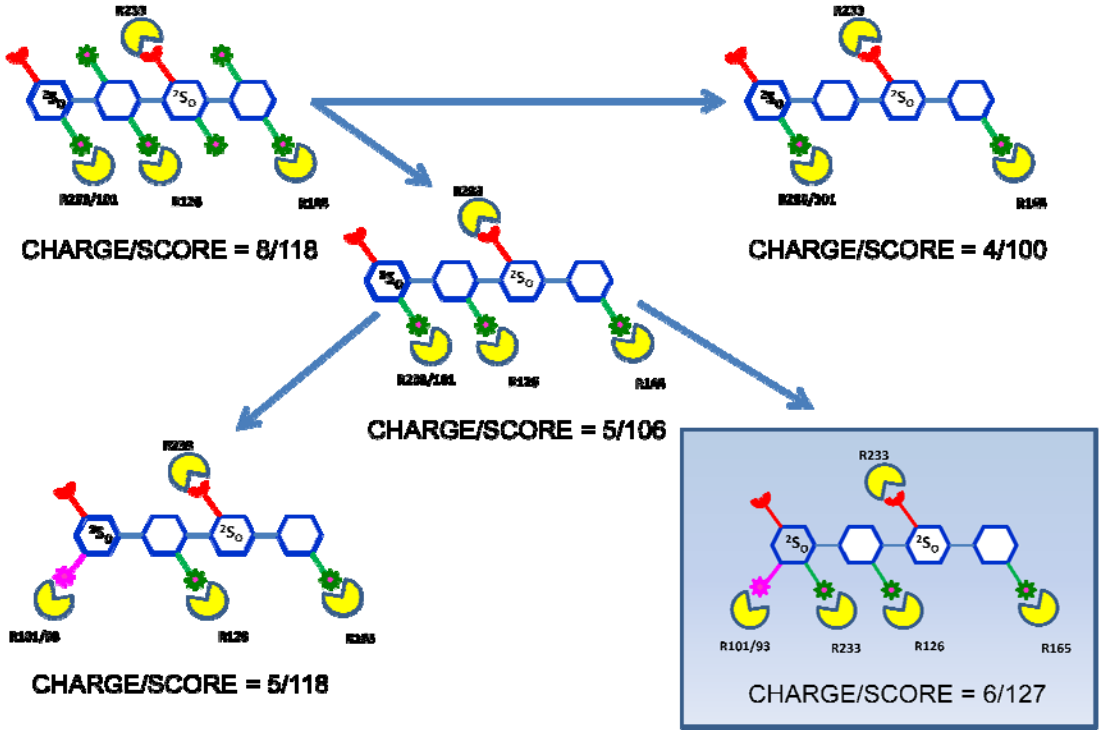
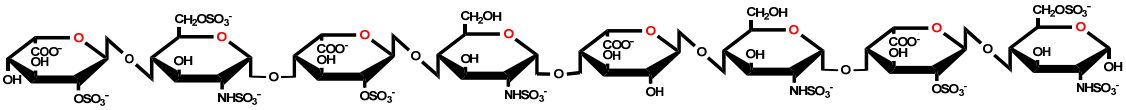
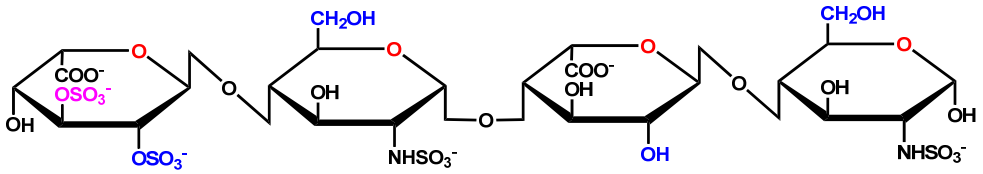


Figure 62. Cartoon structures representing the 'high affinity' and 'high specificity' sequences and their binding amino acids in thrombin exosite-II

The newly-designed 'best hit' HS15a_3-O-SO_3_HS15 sequence is 50% smaller in size and has 62.5% fewer charged functionalities but has a nearly equal GOLD score in comparison to the hit octasaccharide sequences that have 13-16 charges (Figure 63). Not only is this tetrasaccharide smaller in size with fewer charges, but has a better probability of making 'specific' interactions at thrombin exosite-II.



HS21a_$^1C_4$__HS15a__HS15__HS21a_$^1C_4$: GOLD Score 126



HS15a_3-O-SO_3__HS15: GOLD Score 127

Figure 63. Structure of hit octasaccharide and tetrasaccharide sequences.

The average GOLD score for the fully-sulfated hit octasaccharide sequences is 126. The GOLD score for the designed tetrasaccharide with only 6 charges is 127. This shows that we have designed tetrasaccharide sequences with the essential pharmacophore to recognize the important amino acids in thrombin exosite-II in a 'specific' manner. Our optimism is partly based on the reproducibility of the docking results predicting the same binding mode in repeated experiments. This design challenges the much-believed concept of non-specific interaction of heparin to thrombin exosite-II. Since heparin is larger and has many more charges, it can interact many different ways at thrombin exosite-II. This presumably would not be true for the small designed tetrasaccharides.

3.5. Summary and Conclusions

Heparin, a clinically used anticoagulant exhibits its effect by interacting with antithrombin and thrombin. Heparin also interacts with many other proteins besides antithrombin and thrombin, a property which is responsible for its side effects. However, a

heparin pentasaccharide sequence specifically recognizes antithrombin and this interaction has been studied extensively. Although the antithrombin–heparin interaction is considered a specific interaction, the architecture responsible for specificity in antithrombin remains unclear.

Our molecular modeling studies of antithrombin crystal structures explain some of the factors that are responsible for specificity. Even though, one would expect long side chain amino acids, such as Lys and Arg, that form the heparin binding site (HBS) in the surface of the antithrombin to be highly surface exposed, it was found that most of the critical amino acids in HBS are instead ~60% buried. In addition to the buried nature, they are also held in place by neighboring H-bonding partners such Asp and Glu, which restricts the flexibility of the long side chains. In some cases, like Lys114, the hydrophobic environment keeps the long side chains close to the protein rather than solvent exposed. Calculation of radii of gyration indicated that most of the critical amino acids in HBS are spatially conserved. Analysis of the relative geometry revealed an important feature where Lys114, Lys125 and Arg129 form a non-equilateral scalene triangle, which in principle would favor specific interaction and require a complementary match from the ligand to be recognized. Since pentasaccharide is a molecule with appropriately distributed charge functionalities that meet the 3D geometrical requirements of antithrombin HBS, the antithrombin-pentasaccharide interaction becomes specific rather than just ionic driven interactions between negatively charged ligand and positively charged protein.

At the same time the heparin-thrombin interaction is presumed to be non-specific even though the type of interaction is same as antithrombin. In contrast to antithrombin,

we found that most of the basic long side chain amino acids in thrombin exosite-II are more than ~60% surface exposed and are highly flexible. Also the important amino acids in thrombin exosite-II form an approximate symmetric model which favors non-specific interaction.

Calculation of radii of gyration showed that within this non-specific favored thrombin exosite-II, some amino acids are spatially conserved including Arg101, Arg165 and Arg233, forming a local asymmetric center. Using molecular docking experiments, we have designed novel tetrasaccharide sequences to specifically recognize these amino acids. The newly designed 'best hit' HS15a_3-O-SO_3__HS15 sequence is 50% smaller in size and contains 62.5% fewer charged functionalities but has a nearly equal GOLD score in comparison to the hit octasaccharide sequences that have 13 to 16 charges. This tetrasaccharide is not only smaller in size with fewer charges but has better probability of making 'specific' interactions at thrombin exosite-II based on our repeated docking experiments.

The design of novel 'specific' tetrasaccharide sequences challenges the existing idea that the thrombin–heparin interaction is non-specific and leads to the possibility of designing and synthesizing of heparin-based specific thrombin exosite-II modulators. Since we have already shown in our lab that it is possible to replace the saccharide backbone by a non-saccharide skeleton in designing non-saccharide antithrombin activators, in principle it is also possible to design non-saccharide mimics based on this novel sequence for thrombin exosite-II.

3.6. Computational Methods

Software/Hardware: SYBYL 7.1 (Tripos Associates, St. Louis, MO) was used for molecular visualization, for minimization, and for adding hydrogens to protein structures from the Protein Data Bank. All modeling was performed on an IRIX 6.5-based SGI Tezro graphical workstation. GOLD, version 3.0, was used for docking experiments. Heparin oligomeric sequences were built combinatorially in an automated manner using in-house SPL (SYBYL Programming Language) scripts.

Energy Minimizations: Heparin oligomers and thrombin structures were energy-minimized to get optimal geometric conformation. Except where stated, energy minimization was performed using the Tripos Force Field 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 5000 iterations subject to a termination gradient of 0.05 kcal/(mol-Å).

Protein Coordinates: The coordinates for the thrombin dimer ABGH were extracted from the crystal structure of the thrombin-heparin complex (PDB entry 1XMN). Hydrogen atoms were added in SYBYL 7.1, and the structure was minimized with fixed heavy-atom coordinates using the Tripos force field for 1000 iterations subject to a termination gradient of 0.05 kcal/(mol-Å).

Coordinates for Natural Heparin Octasaccharide: In the 1XMN thrombin-heparin crystal structure, thrombin was crystallized as a tetramer with 1 heparin sequence for each dimeric partner. In total, for 4 thrombin units there are 2 heparin sequences. Though they

have used octasaccharide for crystallization, in one dimer the heparin sequence was resolved only up to 6 saccharide units (AB-GH) and 5 units in the other dimer (CD-EF). In order to build the natural octasaccharide, the crystal structure-resolved hexasaccharide was extracted from the crystal structure and 1 saccharide unit has been added to both ends of the hexasaccharide. In SYBYL, the atom type of sulfur and oxygen atoms in SO_3 groups were modified to S.o2 and O.co2, respectively, and the bond type between these atoms were modified to aromatic bond. Hydrogen atoms, absent in the PDB structure, were added in SYBYL, and the resultant structure was minimized to optimize the geometry of hydrogen atoms only (no change in non-H atoms) at an average $\Phi_H\Psi_H$ values for inter-glycosidic torsion angles using the same protocol for building glycosaminoglycan (GAG) sequences as reported previously.[109]

Coordinates for Heparin Oligomers: The coordinates for the heparin oligomeric sequences were generated using a series of SPL scripts and a set of 8 disaccharide building blocks. Although the number of possible Heparin-Like GAG [UAp (1→4) GlcNp] disaccharides is 48, only 23 have been experimentally observed. On the basis of natural octasaccharide sequence, we restricted our library to include only IdoAp sequences that contain IdoAp2S and 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 of octasaccharide consists of 8 IdoAp-containing disaccharide building blocks (Figure 52). ^2S_O-IdoAp-containing disaccharides were generated using the GH residues from the 1TB6 co-crystal structure as template,[220,248] while the template for the $^1C_4$-IdoAp disaccharides was obtained from the 1BFC

structure.[255] Appropriate side-chain modifications were made to generate the 8 building blocks. Each disaccharide was minimized at the average $\Phi_H\Psi_H$ value subject to a restraining force constant of 0.01 kcal-mol^{-1}-deg^{-2}. The 8 disaccharides were then used to build a combinatorial HS octasaccharide library using an SPL script, following which each sequence was minimized with 10000 iterations as described above in an automated manner. Thus, the HS combinatorial library contained 8×8×8×8 = 4096 octasaccharide sequences.

Crystal Structures for Specific and Non-specific Interaction:

To explore the specific/nonspecific interaction of heparin to antithrombin and thrombin, we have used the reported crystal structures of antithrombin and thrombin co-crystallized with heparin fragments. The list of antithrombin and thrombin crystal structures used for this study with their PDB entry and reference is reported in Table 9.

Theoretical Background for Calculation of Radius of Gyration:

The radius of gyration is often used as a measure of the compactness of a group or cluster of points. To measure the radius of gyration, first the center of mass (COM, Equation 1) of the set of *n* points with masses m_i is calculated:

$$COM = \left(\frac{\sum_{i=1}^{n} m_i x_i}{\sum_{i=1}^{n} m_i}, \frac{\sum_{i=1}^{n} m_i y_i}{\sum_{i=1}^{n} m_i}, \frac{\sum_{i=1}^{n} m_i z_i}{\sum_{i=1}^{n} m_i}\right) = (x_{COM}, y_{COM}, z_{COM}) \quad (1)$$

The COM is the point in 3D space where all of the masses are perfectly balanced. If all of the masses are equal, which is true in our case (we are using either the C^{ζ} carbon atom [for the Arg residues] or the N^{ζ} nitrogen atom [for the Lys residues]), the COM is just the average position of the *n* individual point masses (Equation 2):

$$COM = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n}, \frac{\sum_{i=1}^{n} z_i}{n}\right) = (x_{COM}, y_{COM}, z_{COM}) \quad (2)$$

The distance r between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) is given by Equation 3:

$$r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (3)$$

The moment of inertia I of the set of masses rotating about the COM is the product of the mass and the square of the distance from the COM for each point (Equation 4):

$$I = \sum_{i=1}^{n} m_i r_i^2 = \sum_{i=1}^{n} m_i \left[(x_i - x_{COM})^2 + (y_i - y_{COM})^2 + (z_i - z_{COM})^2\right] \quad (4)$$

I is dependent on the number of points, their individual masses, and their distances from the COM. If the masses are all equal to m, then Equation 4 may be simplified to Equation 5:

$$I = m \sum_{i=1}^{n} (x_i - x_{COM})^2 + (y_i - y_{COM})^2 + (z_i - z_{COM})^2 \quad (5)$$

Multiplying the right-hand side of Equation 5 by unity (n/n) gives Equation 6:

$$I = nm \frac{\sum_{i=1}^{n} (x_i - x_{COM})^2 + (y_i - y_{COM})^2 + (z_i - z_{COM})^2}{n} \quad (6)$$

The total mass of the points is nm. Now, imagine that all of this mass is distributed evenly in a thin layer on the surface of a sphere, such that the moment of inertia I is the same as that for the individual points. The radius of gyration R_g is the radius of this sphere, where M is the total mass of the system:

$$I = MR_g^2 \quad (7)$$

Rearranging Equation 7 and solving for R_g:

$$R_g = \sqrt{\frac{I}{M}} \tag{8}$$

Substitution of Equation 6 for I, nm for M and simplification yields:

$$R_g = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{COM})^2 + (y_i - y_{COM})^2 + (z_i - z_{COM})^2}{n}} \tag{9}$$

Thus, when all of the masses are equal, R_g is the root-mean-square distance (RMSD) of the points from their COM (Equation 9), since in general

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} d^2}{n}} \tag{10}$$

Where d is a general distance measurement between two arbitrary points. The two points could represent (for example) a lysine side chain nitrogen atom and the COM as in Equation 9. Alternatively, the two points could represent the position of a particular atom in two different docked poses of the same ligand, as we do when we analyze the similarity of docked GAG backbone positions.

Docking of Heparin Oligomers:

Docking of saccharide ligands onto the AB-GH dimer of thrombin (PDB ID = 1XMN) was performed with GOLD v.3.0. The binding site in thrombin was defined based on the crystal structure ligand resolved 6 unit 'hexasaccharide' of the octasaccharide as a reference with any amino acid residues within and around the ligand for 18 to 20 Å in the protein. This definition of the binding site covers all important known exosite II residues including H91, R93, R101, R126, R165, R233, K236, K240, and W237.

GOLD 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 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. Unless specified otherwise, default parameters were employed during the GOLD docking runs.

For the smaller set of tetra-, hexa-, and octasaccharide sequences**,** docking was performed using no speed-up and a genetic algorithmic search with a default automatic setting to determine the appropriate iterations by GOLD based on the number of rotatable bonds in the sequence. In this search, GOLD 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 toward 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 are 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 six solutions.

When docking the octasaccharide combinatorial library made from 8 disaccharide building blocks, a two-step docking protocol was utilized. The first step consists of screening all possible sequences using 30000 GA iterations (7 to 8× speed up) and GOLD score evaluation of only the top-ranked solution. This step identified the most promising sequences (top 1%) that have a relatively high GOLD score. The second step consisted of docking these most interesting sequences in triplicate for 300,000 iterations. 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.[256] This modified GOLDScore, which utilizes hydrogen-bonding and van der Waals interactions (eq. 1), was used to rank the final docked solutions.

$$GOLDScore_{mod} = HB_{ext} + 1.375 \times VDW_{ext} \quad (11)$$

Where, HB_{ext} and VDW_{ext} are the "external" (non-bonded interactions taking place between the ligand and receptor) hydrogen bonding and van der Waals terms, respectively. Unless otherwise noted, the terms 'GOLD Score' and 'modified GOLDScore' in this work both refer to Equation 11.

CHAPTER 4
CONCLUSIONS

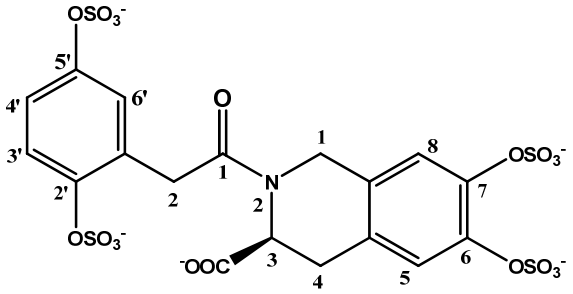
Numerous attempts have been made to design or discover new molecules that activate antithrombin.[158,160,207] However, each of these searches has relied on utilizing a saccharide scaffold as a mimic of heparin. Implicit in these designs was the expectation that a saccharide scaffold was necessary to induce antithrombin activation. Our research group has challenged the assumption that the saccharide-based skeleton is essential for the activity of specific glycosaminoglycan (GAG) sequences. We hypothesized that specific GAG sequences can be replaced by non-saccharide skeletons which may provide several advantages over the GAG skeleton. These include 1) ease of chemical synthesis; 2) likelihood of oral delivery due to enhanced hydrophobic character; 3) opportunity to gain additional non-ionic binding energy; 4) enhanced specificity for the target protein; and 5) the ability to modulate responses in either an agonist or an antagonist manner.

Desai and co-workers have designed small, non-saccharide sulfated flavans based on the trisaccharide DEF as a template using hydropathic interaction (HINT)[210] analysis. The reported sulfated flavans were found to be weak actvators of antithrombin. To improve on the antithrombin activation potential of these organic activators, a tetrahydroisoquinoline-based bicyclic-unicyclic sulfated activator IAS_5 was designed using a pharmacophore-based approach by Raghuraman et al.[209] These studies showed that IAS_5 activates antithrombin nearly 30-fold, an increase of nearly 2- to 3-fold higher than the first-generation rationally designed agents.[134,135] However, the designed molecules were

found to be weak activators of antithrombin due to their binding to the extended heparin binding site (EHBS) instead of the pentasaccharide binding site (PBS) of antithrombin.

To design better non-saccharide antithrombin activators, a virtual screening-based approach, which categorizes the ligands as either PBS- or EHBS-binding molecules in native and activated antithrombin, was employed. Combinatorial virtual screening of 24576 molecules based on a tetrahydroisoquinoline (ISOQ) core scaffold resulted in 92 hits that were predicted to bind preferentially in the PBS of activated antithrombin with good affinity. The molecular modeling results lead to several hypotheses. Of special importance are three hypotheses. The modeling results suggest 1) an optimal linker length of the 4 to 5 atoms; 2) a 2′,5′-disulfated unicyclic ring; and 3) a 3,5,6-trisubstituted bicyclic ring are predicted to be important. The hypotheses formed the basis for the synthetic work currently in progress in the Desai laboratory.

As mentioned earlier, the synthesis of highly sulfated molecules is challenging.[209] To date, no molecule with all three features has been synthesized. Yet, the laboratory has just synthesized and tested one molecule that may serve as initial test of molecular modeling experiments.



Activator 67A2L25

The potential activator 67A2L25 was synthesized (by Mr. Al-Horani) and evaluated (by Dr. Liang) using the standard antithrombin activation protocol developed by the laboratory. Activator 67A2L25 has a 6,7-disulfated bicyclic ring (instead of 5,6-disubstitution), a two carbon linker (instead of 4 or 5 carbon linker); and a 2,5-disubstituted unicyclic ring (as desired).

Biochemical analysis of 67A2L25 shows an antithrombin activation potential in the region of 50- to 100-fold (the high variance is because of the very high salt content as impurity in the sample). The activation potential of 67A2L25 is much less than full antithrombin activation (300-fold), but significantly more than 8- to 30-fold activation achieved with previous designs.[134,135,209] The antithrombin affinity was found to be between 1 and 5 μM at pH 7.4. This is significantly higher than the affinities of all previous designs (50 to 500 μM),[171,212,214] though still less than most potent saccharide-based activators (50 to 100 nM).

Since the synthesized molecule was not originally present in the virtual library, this molecule was docked using the same docking parameters. Modeling results for the synthesized molecule shows that the GOLD score for 67A2L25 is 97.2, which is lower than the scores for the hit molecules identified in the virtual library screening and the binding modes are also divergent in triplicate docking experiments.

Several conclusions may be derived based on this single data point. 1) The activation is not full (50-100 fold in comparison to 300-fold) for compound 67A2L25 probably because of the absence of optimal structural features, as predicted by GOLD modeling; 2) the activation is much better than that realized with previous designs

probably because of the slightly longer linker and/or 2′,5′-disulfation of the unicyclic ring; 3) the antithrombin affinity of 67A2L25 is found to be higher probably because of more optimal features (slightly longer linker and/or 2,5-disulfation of unicyclic ring) than those present in previous designs; and 4) the antithrombin affinity of 67A2L25 is not as great as the most optimal saccharide activators because all of the features suggested by the docking experiments have not been introduced. Overall, the antithrombin binding and activation results obtained with 67A2L25 are consistent with the GOLD-based molecular modeling results obtained and provide stronger impetus to synthesize and test the 'high-affinity, high-specificity' molecules identified in this work.

In the second project, the specific and nonspecific interactions of heparin with antithrombin and thrombin have been studied. Heparin exhibits its anticoagulant effect by interacting with antithrombin and thrombin. Although the heparin binding sites in both antithrombin and thrombin are lined with Arg and Lys amino acid residues, the antithrombin-heparin interaction is specific, whereas the thrombin-heparin interaction is considered to be nonspecific.

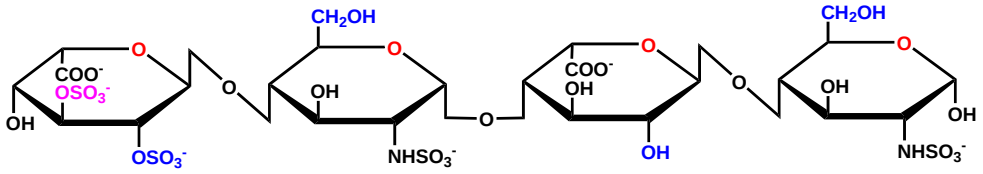
Our molecular modeling studies and the crystal structure analyses of antithrombin and thrombin explain some of the factors that are responsible for specificity. Even though one would expect long side chain amino acids such as Lys and Arg, which form the heparin binding site (HBS) on the surface of antithrombin, to be highly surface exposed, it was found that most of the critical amino acids in the HBS are instead ~60% buried. In addition to the buried nature, they are also held in place by neighboring H-bonding partners such Asp and Glu, which restricts the flexibility of the long side chains. In some

cases, like Lys114, the hydrophobic environment keeps the long side chains close to the protein rather than solvent exposed. Calculation of radii of gyration indicated that most of the critical amino acids in HBS are spatially conserved. Analysis of the relative geometry of the basic residues revealed an important feature wherein Lys114, Lys125 and Arg129 form a non-equilateral scalene triangle, which in principle would favor specific interaction and require a complementary match from the ligand to be recognized. Since pentasaccharide is a molecule with appropriately distributed charge functionalities that meet the 3-D geometrical requirements of antithrombin HBS, the antithrombin-pentasaccharide interaction becomes specific rather than simple nonspecific ionically driven interactions between negatively charged ligand and positively charged protein functional groups.

At the same time the heparin-thrombin interaction is presumed to be non-specific even though the receptor-ligand interactions are very similar to those in antithrombin. In contrast to antithrombin, we found that most of the basic long side chain amino acids in thrombin exosite-II are more than ~60% surface exposed and are highly flexible. Also, the important amino acids in the thrombin exosite-II are arranged in a highly symmetric pattern which favors non-specific interaction due to multiple or degenerate binding mode.

However, calculation of radii of gyration showed that within this non-specific favored thrombin exosite-II, Arg101, Arg165 and Arg233 are spatially conserved and form a local asymmetric center. Using molecular docking experiments, we have designed novel tetrasaccharide sequences to specifically recognize these amino acids. The newly-designed 'best hit' HS15a_3-O-SO_3__HS15 sequence is 50% smaller in size and contains 62.5%

fewer charged functionalities but has a nearly equal GOLD score in comparison to the hit octasaccharide sequences that have 13 to 16 charges. This tetrasaccharide is not only smaller in size with fewer charges but has better probability of making 'specific' interactions at thrombin exosite-II based on our repeated docking experiments.



HS15a_3-O-SO_3__HS15: GOLD Score 127

The design of novel 'specific' tetrasaccharide sequences challenges the existing idea that the thrombin–heparin interaction is non-specific and leads to the possibility of designing and synthesizing heparin-based specific thrombin exosite-II modulators. Since we have already proved in our lab that it is possible to replace the saccharide backbone by a non-saccharide skeleton in designing non-saccharide antithrombin activators, in principle it is also possible to design non-saccharide mimics based on this novel sequence for thrombin exosite-II.

Literature Cited

Literature Cited

1. Veselovsky, A. V.; Ivanov, A. S. Strategy of computer-aided drug design. *Current Drug Targets – Infectious Disorders* **2003,** *3*, 33-40.

2. Ghose, A. K.; Viswanadhan V. N.; Wendoloski, J. J. The fundamentals of pharmacophore modeling in combinatorial chemistry. *J Recept Signal Transduct Res.* **2001,** *21*, 357-375.

3. Kubinyi, H. QSAR in drug design. In *Handbook of Chemoinformatics. From Data to Knowledge. Vol. 4.* Gasteiger, J. Ed.; Wiley-VCH, Weinheim, 2003; pp 1532-1554.

4. Kubinyi, H. Comparative Molecular Field Analysis (CoMFA). In *Handbook of Chemoinformatics. From Data to Knowledge. vol. 4.* Gasteiger, J. Ed.; Wiley-VCH, Weinheim, 2003; pp 1555-1574.

5. Duchowicz, P. R.; Castro, E. A. QSPR studies on aqueous solubilities of drug-Like compounds. *Int J Mol Sci* **2009**, *10*, 2558-2577.

6. Rekker, R. F. The history of drug research: From Overton to Hansch. *Quant Struct-Act Relat* **1992**, *11*, 195-199.

7. Meyer, H. Zur Theorie Der Alkolnarkose I. Welche Eigenschaft Der Anaesthetica Bedingt Jhre Narkotische Wirkuny? *Arch Exp Pathol Pharmakol* **1899**, *42*, 109-118.

8. Overton, E. Osmotic properties of cells in the bearing on toxicology and pharmacy. *Z Physik Chemie* **1897**, *22,* 189-209.

9. Shorter, J. The prehistory of the Hammett equation. *Chem Listy* **2000**, *94*, 210-214.

10. Hammett, L. P. Some relations between reaction rates and equilibrium constants. *Chem Rev* **1935**, *17*, 125-136.

11. Hammett, L. P. *Physical Organic Chemistry, 2nd Ed;* McGraw Hill: New York, 1970.

12. Hansch, C.; Leo, A.; Hoekman D. Exploring QSAR: Hydrophobic, Electronic, and Steric Constants. American Chemical Society: Washington DC 1995

13. Taylor, P. J. Hydrophobic properties of drugs. In *Comprehensive Medicinal Chemistry*. *Vol. 4. Quantitative Drug Design. The Rational Design, Mechanistic Study and Therapeutic Applications of Chemical Compounds*; Hansch, C.; Sammes, P. G.; Taylor, J. B.; Eds.; Pergamon Press: Oxford England, 1991; pp 241-294.

14. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **1959**, *14,* 1-63.

15. Haymet, A. D. J.; Silverstein, K. A. T.; Dill, K. A. Hydrophobicity reinterpreted as 'Minimization of the entropy penalty of solvation'. *Faraday Discuss* **1996**,*103,* 117-124.

16. Taft, R.W. *Steric Effects in Organic Chemistry;* John Wiley: New York, 1956.

17. Charton, M. Steric effects. I. Esterification and acid-catalyzed hydrolysis of esters. *J Am Chem Soc* **1975**, *97,* 1552-1556.

18. Charton, M. Steric effects in drug design. The upsilon steric parameter–definition and determination. *Top Curr Chem* **1983**, *114,* 57-92.

19. Hancock, C. K.; Meyers, E. A.; Yager, B. J. Quantitative separation of hyperconjugation effects from steric substituent constants. *J Am Chem Soc* **1961,** *83,* 4211-4213.

20. Tute, M. History and objectives of quantitative drug design. In *Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design. The Rational Design, Mechanistic Study and Therapeutic Applications of Chemical Compounds*; Hansch, C.; Sammes, P. G.; Taylor, J. B.; Eds.; Pergamon Press: Oxford England, 1991; pp 1- 32.

21. Verloop, A. *The STERIMOL Approach to Drug Design;* Marcel Dekker: New York, 1987.

22. Motoc, I. Molecular shape descriptors. *Top Curr Chem* **1983**, *114,* 93-105.

23. Lien, E. J.; Guo, Z. R.; Li, R. L.; Su, C. T. Use of dipole moment as a parameter in drug–receptor interaction and quantitative structure-activity relationship studies. *J Pharm Sci* **1982**, *71,* 641-655.

24. Li, W.Y.; Guo, Z. R; Lien, E. J. Examination of the interrelationship between aliphatic group dipole moment and polar substituent constants. *J Pharm Sci* **1984**, *73,* 553-558.

25. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem Rev* **1996**, *96,* 1027-1043.

26. Roothaan, C. C. J.; Sachs, L. M.; Weiss, A. W. Analytical self-consistent-field functions for the atomic configurations $1s^2$, $1s^22s$, and $1s^22s^2$. *Rev Mod Phys* **1960**, *32,* 186-194.

27. Sotomatsu, T.; Murata, Y.; Fujita, T. Correlation analysis of substituent effects on the acidity of benzoic acids by the AM1 method. *J Comput Chem* **1989**, *10,* 94-98.

28. Loew, G. H.; Burt, S. K. Quantum mechanics and the modeling of drug properties. In *Comprehensive Medicinal Chemistry. Vol.4. Quantitative Drug Design;* Ramsden, C.A.; Ed.; Pergamon Press: Oxford England, 1990; pp 105-123.

29. Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biolmol Struct* **2003,** 32, 335–373.

30. Muegge, I.; Rarey, M. Small molecule docking and scoring. Reviews in *Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B.; Eds.; John Wiley and Sons: 2001**;** *17*, pp 1-60.

31. Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R. L. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* **1982,** *161,* 269-288.

32. Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992,** *257,* 1078-1082.

33. Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J Comput Chem* ***1997*****,** *18*, 1175-1189.

34. Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* ***2001,*** *15*, 411–428.

35. Chemical computing group. MOE. 2003. Montreal, Quebec, Canada.

36. Olson, A. J.; Goodsell, D. S. Automated docking in crystallography: analysis of the substrates of aconitase. *Proteins* **1993**, *17,* 1–10.

37. Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. Fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996***, **261***, 470–489.

38. Kramer, B.; Rarey, M.; Lengauer, T. CASP2 Experiences with docking flexible ligands using FlexX. *Proteins* **1997***, 1*, 221-225.

39. Kramer, B.; Metz, G.; Rarey, M.; Lengauer, T. Ligand docking and screening with FlexX. *Med Chem Res* **1999***, 9*, 463-478.

40. Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. *Chem Bio* **1996***, 3*, 449–462.

41. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R. Development and Validation of a genetic algorithm for flexible docking. *J. Mol. Biol* **1997***, 267*, 727-748.

42. Kearsley, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. Flexibases: A way to enhance the use of molecular docking methods. *J Comput Aided Mol Des* **1994***, 8*, 565–582.

43. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004***, 47*, 1739–1749.

44. Westhead, D. R.; Clark, D. E.; Murray, C. W. A. Comparison of heuristic search algorithms for molecular docking. *J Comput Aided Mol Des* **1997***, 11*, 209–228.

45. Opera, T. I.; Marshall, G. R. Receptor-based prediction of binding affinities. *Perspect Drug Discovery Des* **1998***, 9/10/11*, 35-61.

46. Böhm, H. J.; Stahl, M. Rapid empirical scoring functions in virtual screening applications. *Med Chem Res* **1999***, 9*,445-462.

47. Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **1999***, 42*, 5100-5109.

48. Tame, J. R. H. Scoring functions; A view from the bench. *J Comput-Aided Mol Des* **1999***, 13*, 99-108.

49. Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J Comput Chem* **1992,** *13,* 505-524.

50. Meng, E. C.; Kuntz, I. D.; Abraham, D. J.; Kellogg, G. E. Evaluating docked complexes with the hint exponential function and empirical atomic hydrophobicities. *J Comput Aided Mol Des* **1994***, 8,* 299-306.

51. Gschwend, D. A.; Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: on-the-fly optimization and degeneracy removal. *J Comput Aided Mol Des* **1996,** *10,* 123-132.

52. Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins: Struct Funct Genet* **1999,** *34,* 4-16.

53. Zou, X. Q.; Sun, Y. X.; Kuntz, I. D. Inclusion of solvation in molecular ligand binding free energy calculations using the generalized-born model. *J Am Chem Soc* **1999***, 121,* 8033-8043.

54. Böhm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *Comput Aided Mol Des* **1992,** *6,* 593-606.

55. Böhm, H. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J Comput Aided Mol Des* **1992,** *6,* 61-78.

56. Lawrence, M. C.; Davis, P. C. CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three dimensional structure. *Proteins: Struct Funct Genet* **1992,** *12,* 31-41.

57. Mizutani, M. Y.; Tomioka, N.; Itai, A. Rational search method for stable docking models of protein and ligand. *J Mol Biol* **1994**,*243,* 310-326.

58. Lamdan, Y.; Wolfson, H. J. Geometric hashing: A general and efficient model based recognition scheme. In *Proceedings of the IEEE International Conference on Computer Vision,* 1998; pp. 238-249.

59. Linnainmaa, S.; Harwood, D.; Davis, L. S. Pose determination of a three-dimensional object using triangle pairs. IEEE, Transactions on pattern analysis and machine intelligence **1998**, *10*, 634.

60. Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J Comput Aided Mol Des* **1996**, *10*, 41.

61. Havel,T. F.; Kuntz I. D.; Crippen, G. M. The combinatorial distance geometry approach to the calculation of molecular conformation. 1. A new approach to the old problem. *J Theor Biol* **1983***, 104*, 359-381.

62. Havel, T. F.; Kuntz, I. D.; Crippen, G. M. The theory and practice of distance geometry. *Bull Math Biol* **1983**, *45*, 665-720.

63. Kearsly, S. K.; Underwood, D. J.; Sheridan, R. P.; Miller, M. D. Flexibase: a way to enhance the use of molecular docking methods. *J Comput Aided Mol Des* **1994***, 8,* 565–582.

64. Murcko, M. A. Recent advances in ligand design methods. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; Wiley-VCH, New York, 1997; vol. 11, pp. 1-66.

65. Clark, D. E.; Murray, C. W.; Li, J. Current issues in *de novo* molecular design. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; Wiley-VCH, New York, 1997; vol. 11, pp. 67-125.

66. DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J Med Chem* **1986,** *29,* 2149-2153.

67. Sandak, B.; Nussinov, R.; Wolfson, H. J. A method for biomolecular structural *and* recognition docking allowing conformational flexibility. *J Comput Biol* **1998**, *5,* 631-654.

68. Leach, A. R.; Kuntz, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem* **1992**, *13,* 703-748.

69. Rarey, M.; Kramer, B.; Lengauer, T. Time- efficient docking of flexible ligands into active sites of proteins. In *Proceedings of the third international conference on intelligent systems in molecular biology*; Rawlings, C., Ed.; AAAI press, Menlo Park, CA, 1995; pp. 300-308.

70. Rarey, M.; Kramer, B.; Lengauer, T. Multiple automatic base selection-protein-ligands docking based on incremental construction without manual intervention. *J Comput Aided Mol Des* **1997**, *11*, 369-384.

71. Klebe G.; Mietzner, T. A fast and efficient method to generate biologically relevant conformations. *J Comput Aided Mol Des* **1994**, *8,* 583-606.

72. Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink-Peters, T.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge crystallographic data center: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr* **1979,** *B35*, 2331-2339.

73. Judson, R. Genetic algorithms and their use in chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1997; vol. 10, pp. 1-73.

74. Jones, G.; Willett, P.; Glen, R. C. A genetic algorithm for flexible molecular overlay and pharmacophore elucidation. *J Comput Aided Mol Des* **1995,** *9,* 532-549.

75. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* **1997***, 267*, 727-748.

76. McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J Comput Aided Mol Des* **1997,** *11*, 333-344.

77. Abagyan, R.A.; Totrov, M. M.; Kuznetsov, D. A. ICM: A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comp Chem* **1994,** *15,* 488-506.

78. Roisman, L. C.; Piehler, J.; Trosset, J. Y.; Scheraga, H. A.; Schreiber, G. Structure of the interferon-receptor complex determined by distance constraints from double-mutant cycles and flexible docking. *Proc Natl Acad Sci USA* **2001,** *98(23),* 13231-13236.

79. Murray, C. W.; Baxter, C. A.; Frenkel, A. D. The sensitivity of the results of molecular docking to induced fit effects: Application to thrombin, thermolysin and neuraminidase. *J Comput Aided Mol Des* **1999**, *13*, 547-562.

80. Goodsell, D. S.; Olson, A.J. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct Funct Genet* **1990***, 8,* 195-202.

81. Kirkpatrik, S.; Gelatt, C. D. J.; Vecchi, M. P. Optimization by simulated annealing. *Science* **1983**, *220,* 671-680.

82. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. *J Comput Aided Mol Des* **1996,** *10,* 293-304.

83. Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J Mol Recognit* **1996,** *9*, 1-5.

84. Lybrand, T. P. Computer simulation of biomolecular systems using molecular dynamics and free energy perturbation methods. In *Reviews in Computational*

Chemistry; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1990; vol. 1, pp. 295-320.

85. Leach, A. R. A survey of methods for searching the conformational space of small and medium-sized molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1991; vol. 2, pp. 1-55.
86. Dinur, U.; Hagler, A. T. New approaches to empirical force fields. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1991; vol. 2, pp. 99-164.
87. Straatsma, T. P. Free energy by molecular simulation. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1996; vol. 9, pp. 81-127.
88. Pettersson, I.; Liljefors, T. Molecular mechanics calculated conformational energies of organic molecules. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, New York, 1996; vol. 9, pp. 167-189.
89. Meirovitch, H. Calculation of the free energy and the entropy of macromolecular systems by computer simulation. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers, New York, 1998; vol. 12, pp. 1-74.
90. Spyrakis, F.; Cozzini, P.; Kellogg, G. Docking and Scoring in Drug Discovery. In Burgers Medicinal Chemistry, Drug Discovery and Development; John Wiley & sons (In press)

91. Weiner, S.; Kollman, P.; Case, D.; Singh, U.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* **1984**, *106*, 765-784.

92. Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem* **1986**, *7*, 230-252.

93. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* **1983**, *4,*187-217.

94. The SYBYL software, TI, http://www.tripos.com/ 1995; Tripos Inc.: St. Louis, MO.

95. Allinger, N. L. Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *J Am Chem Soc* **1977**, *99*, 8127-8134.

96. Allinger, N. L.; Yuh, Y.H.; Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J Am Chem Soc* **1989,** *111*, 8551-8566.

97. Bowen, J. P.; Allinger, N. L. Molecular Mechanics: The Art and Science of Parameterization, In *Rev Comput Chem*; VCH Publishers: New York, 1991; pp 81-97.

98. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews* **2004**, *3*, 935-949.

99. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to

estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* **1997, *11***, 425–445.

100. Muegge, I. A knowledge-based scoring function for protein–ligand interactions: probing the reference state. *Perspect Drug Discov Des* **2000***, 20*, 99–114.

101. Muegge, I. Effect of ligand volume correction on PMF scoring. *J Comput Chem* **2001***, 22*, 418–425.

102. Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* **1999***, 42*, 791–804.

103. Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J Mol Biol* **2000***, 295*, 337–356.

104. DeWitte, R. S.; Shakhnovich, E. I. SMoG: *de novo* design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J Am Chem Soc* **1996***, 118*, 11733–11744.

105. Krovat, E.; Steindl, T.; Langer, T. Recent advances in docking and scoring. Current computer-aided drug design **2005**, *1,* 93-102.

106. Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **1999***, 42*, 5100-5109.

107. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* **2002,** *16*, 11–26.

108. GOLD™ user manual GOLD v3.0

109. 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.

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

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

112. Jackson, R. L.; Busch, S.J.; Cardin, A.D. Glycosaminoglycans: molecular properties, protein interactions, and role in physiological processes. *Physiol Rev* **1991**, *71*, 481-539.

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

114. Sasisekharan, R.; Venkataraman, G. Heparin and heparan sulfate: biosynthesis, structure and function. *Curr Opin Chem Biol* **2000,** *4*, 626-631.

115. Capila, I.; Linhardt, R.J. Heparin–protein interactions. *Angew Chem Int Ed Engl* **2002**, *41*, 390-412.

116. Gandhi, N. S.; Coombe, D. R.; Mancera R. L. Platelet endothelial cell adhesion molecule 1 (PECAM-1) and its interactions with glycosaminoglycans: 1. Molecular modeling studies. *Biochemistry* **2008,** *47*, 4851-4862.

117. Lortat-Jacob, H.; Grosdidier, A.; Imberty, A. Structural diversity of heparan sulfate binding domains in chemokines. *Proc Natl Acad Sci USA* **2002,** *99*, 1229-1234.

118. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem* **1985**, *28*, 849-857.

119. Bitomsky, W.; Wade, 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.

120. Ricard-Blum, S.; Feraud, O.; Lortat-Jacob, H.; Rencurosi, A.; Fukai, N.; Dkhissi, F. Vittet, D.; Imberty, A.; Olsen B. R.; van der Rest M. Characterization of endostatin binding to heparin and heparan sulfate by surface plasmon resonance and molecular modeling: role of divalent cations. *J Biol Chem* **2004**, *279,* 2927-2936.

121. Mulloy, B.; Forster, M. J. Application of drug discovery software to the identification of heparin-binding sites on protein surfaces: a computational survey of the 4-helix cytokines. *Molecular Simulation* **2008,** *34*, 481-489.

122. Fallahi, A; Kroll B.; Warner, L. R.; Oxford, R. J.; Irwin, K. M.; Mercer, L. M. et al. Structural model of the amino propeptide of collagen XI {alpha} 1 chain with similarity to the LNS domains. *Protein Sci* **2005**, *14,* 1526-37.

123. Pita Samuel Silva da Roch, Vinício Tácio , Fernandes Amorim , Caffarena Ernesto Raul , Geraldo P.P. Studies of molecular docking between fibroblast growth factor and heparin using generalized simulated annealing. *Int J Quantum Chem* **2008**, *108*, 2608-14.

124. Forster, M.; Mulloy, B. Computational approaches to the identification of heparin-binding sites on the surfaces of proteins. *Biochem Soc Trans* **2006**, *34,* 431-4.

125. Kern, A.; Schmidt, K.; Leder, C.; Muller, O. J.; Wobus, C. E.; Bettinger, K. et al. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *J Virol* **2003**, *77,* 11072-81.

126. Lam, K.; Rao, V. S.; Qasba, P. K. Molecular modeling studies on binding of bFGF to heparin and its receptor FGFR1. *J Biomol Struct Dyn* **1998,** *15,* 1009-27.

127. Sadir, R.; Baleux, F.; Grosdidier, A.; Imberty, A.; Lortat-Jacob, H. Characterization of the stromal cell-derived factor-1a-heparin complex. *J Biol Chem* **2001***, 276*, 8288-96.

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

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

130. Mikhailov, D.; Mayo, K. H.; Vlahov, I. R.; Toida, T.; Pervin, A.; Linhardt, R. J. NMR solution conformation of heparin-derived tetrasaccharide. *Biochem J* **1996**, *318*, 93-102.

131. Desai, U. R.; Wang, H. M.; Kelly, T. R.; Linhardt, R. J. Structure elucidation of a novel acidic tetrasaccharide and hexasaccharide derived from a chemically modified heparin. *Carbohydrate Res* **1993**, *241*, 249-259.

132. Ferro, D. R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **1990**, *195*, 157-167.

133. Verli, H.; Guimarães, J. A. Insights into the induced fit mechanism in antithrombin–heparin interaction using molecular dynamics simulations. *J Mol Graph Model* **2005***, 24,* 203–212.

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

135. 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.

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

137. Gettins, P. G. W.; Patston, A.; Olson, S. T. Serpins: Structure, function and biology. R G Landes Company: New York, 1996.

138. Gettins, P. G. W. Serpin structure, mechanism and function. *Chem Rev* **2002**, *102*, 4751–4803.

139. Schreuder, H. A.; de Boer, B.; Dijkema, R.; Mulders, J.; Theunissen, H. J. M.; Grootenhuis, P. D. J.; Hol, W. G. J. The intact and cleaved human antithrombin III complex as a model for serpin-proteinase interactions. *Nat Struct Biol* **1994**, *1,* 48–54.

140. Carrell, R. W.; Stein, P. E.; Fermi, G.; Wardell, M. R. Biological implications of a 3 Å structure of dimeric antithrombin. *Structure* **1994**, 2, 257–270.

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

142. Wright, H. T.; Scarsdale, J. N. Structural basis of serpin inhibitor activity. *Proteins: Struct Funct Gen* **1995**, *22*, 210–225.

143. Wright, H. T. The structural puzzle of how serpin serine proteinase inhibitors work. *Bio Essays* **1996,** *18*, 453–464.

144. Baglin, T.; Carrell, R. W.; Church, F. C.; Esmon, C.T.; Huntington, J. A. Crystal structures of native and thrombin complexed heparin cofactor II reveal a multi-step allosteric mechanism. *Proc Natl Acad Sci USA* **2002**, *99*, 11079–11084.

145. Bruch, M.; Weiss, V.; Engel, J. Plasma serine proteinase inhibitors (serpins) exhibit major conformational changes and a large increase in conformational stability upon cleavage at their reactive sites. *J Biol Chem* **1988**, *263*, 16626–16630.

146. Kaslik, G.; Kardos, J.; Szabo´, E.; Szila´gyi, L.; Za´vodsky, P.; Westler, W. M.; Markley, J. L.; Gra´f, L. Effects of serpin binding on target proteinase: Global stabilization, localized increased structural flexibility, and conserved hydrogen bonding at the active site. *Biochemistry* **1997**, *36*, 5455–5464.

147. Huntington, J. A.; Read, R. J.; Carrell, R. W. Structure of a serpin–protease complex shows inhibition by deformation. *Nature* **2000**, *407*, 923–926.

148. Plotnick, M. I.; Mayne, L.; Schechter, N. M.; Rubin, H. Distortion of the active site of chymotrypsin complexed with a serpin. *Biochemistry* **1996,** *35*, 7586–7590.

149. Calugaru, S. V.; Swanson, R.; Olson, S. T. The pH dependence of serpin-proteinase complex dissociation reveals a mechanism of complex stabilization involving inactive and active conformational states of the proteinase which are perturbable by calcium. *J Biol Chem* **2001**, *276*, 32446–32455.

150. Björk, I.; Olson, S. T. Antithrombin: A bloody important serpin. In: *Chemistry and Biology of Serpines*; Church, F. C.; Cunningham, D. D.; Ginsburg, D.; Hoffman, M.; Tollefsen, D. M.; Stone, S. R., Eds.; Plenum Press: New York, 1997; pp 17–33.

151. Olson, S. T.; Shore J. D. Demonstration of a two-step reaction mechanism for inhibition of α-thrombin by antithrombin III and identification of the step affected by heparin. *J Biol Chem* **1982**, *257*, 14891–14895.

152. Latallo, Z. S.; Jackson, C. M. Reaction of thrombins with human antithrombin III. II. Dependence of rate of inhibition on molecular form and origin of thrombin. *Thromb Res* **1986**, 43, 523–537.

153. Wong, R. F.; Windwer, S. R.; Feinman, R. D. Interaction of thrombin and antithrombin. Reaction observed by intrinsic fluorescence measurements. *Biochemistry* 1983, 22, 3994–3999.

154. Craig, P. A.; Olson, S. T.; Shore, J. D. Transient kinetics of heparin-catalyzed protease inactivation by antithrombin III. Characterization of assembly, product formation, and heparin dissociation steps in the factor Xa reaction. *J Biol Chem* **1989**, *264,* 5452–5461.

155. Jordan, R. E.; Oosta, G. M.; Gardner, W. T.; Rosenberg, R. D. The kinetics of haemostatic enzyme-antithrombin interactions in the presence of low molecular weight heparin. *J Biol Chem* **1980**, *255*, 10081–10090.

156. Griffith, M. J. Kinetics of the heparin-enhanced antithrombin III/thrombin reaction. Evidence for a template model for the mechanism of action of heparin. *J Biol Chem* **1982,** *257,* 7360–7365.

157. Olson, S. T.; Björk, I. Predominant contribution of surface approximation to the mechanism of heparin acceleration of the antithrombin-thrombin reaction. Elucidation from salt concentration effects. *J Biol Chem* **1991**, *266,* 6353–6364.

158. Olson, S. T.; Björk, I.; Sheffer, R.; Craig, P. A.; Shore, J. D.; Choay, J. 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.

159. Duchaussoy, P.; Jaurand, G.; Driguez, P-A.; Lederman, I.; Ceccato, M-L.; Gourvenec, F.; Strassel, J-M.; Sizun, P.; Petitou, M.; Herbert, J-M. Assessment through chemical synthesis of the size of the heparin sequence involved in thrombin inhibition. *Carbohydr Res* **1999**, *317*, 85–99.

160. Desai, U.R. Antithrombin activation and designing novel heparin mimics. In *Chemistry and Biology of Heparin and Heparan Sulfate*; Garg, H. G.; Linhardt, R. J.; Hales, C. A., Ed.; Elsevier, New York, 2005; pp. 483-512.

161. ChuangY-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.

162. Petitou, M.; Herault, J-P.; Bernat, A.; Driguez, P-A.; Duchaussoy, P.; Lormeau, J-C.; Herbert ,J-M. Synthesis of thrombin-inhibiting heparin mimetics without side effects. *Nature* **1999**, *398*, 417–422.

163. Rezaie, A. R. Calcium enhances heparin catalysis of the antithrombin-factor Xa reaction by a template mechanism. *J Biol Chem* **1998**, *273*, 16824–16827.

164. Jin, L.; Abrahams, J-P.; Skinner, R.; Petitou, M.; Pike, R. N.; Carrell, R. W. The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci USA* **1997**, *94,* 14683–14688.

165. Ersdal-Badju, E.; Lu, A.; Zuo, Y.; Picard, V.; Bock, S. C. Identification of the antithrombin III heparin binding site. *J Biol Chem* **1997**, *272,* 19393–19400.

166. Desai, U. R.; Swanson, R. S.; Bock, S. C.; Björk, I.; Olson, S. T. The role of arginine 129 in heparin binding and activation of antithrombin. *J Biol Chem* **2000**, *275*, 18976–18984.

167. Schedin-Weiss, S.; Desai, U. R.; Bock, S. C.; Gettins, P. G. W.; Olson, S. T.; Björk, I. The importance of lysine 125 for heparin binding and activation of antithrombin. *Biochemistry* **2002**, *41*, 4779–4788.

168. Arocas,V.; Bock, S. C.; Raja, S.; Olson, S. T.; Björk, 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.

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

170. Arocas, V.; Turk, B.; Bock, S. C.; Olson, S. T.; Björk, 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.

171. Desai, U. R. New antithrombin–based anticoagulants. *Med Res Rev* **2004**, *24*, 151-181.

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

173. 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.

174. Warkentin, T. E.; Greinacher, A. Heparin-induced thrombocytopenia: recognition, treatment, and prevention: the seventh ACCP conference on antithrombotic and thrombolytic therapy. *Chest* **2004**, *126*, 311S-337S.

175. Weitz, J. I. Low-molecular-weight heparins. *New Engl J Med* **1997**, *337,* 688–698.

176. Hirsh, J.; Levine, M. N. Low molecular weight heparins. *Blood* **1992**, *79*, 1–17.

177. Cosmi, B.; Hirsh, J. Low molecular weight heparins. *Curr Opin Cardiol* **1994***, 9,* 612–618.

178. Hirsh, J.; Warkentin, T. E.; Shaughnessy, S. G.; Anand, S. S.; Halperin, J. L.; Raschke, R.;Granger, C.; Ohman, E. M.; Dalen, J. E. Heparin and low molecular weight heparins: Mechanisms of action, pharmacokinetics, dosing, monitoring, efficacy and safety. *Chest* **2001**,*119*, 64S-94S.

179. Hirsh, J.;Warkentin, T. E.; Raschke, R.; Granger, C.; Ohman, E. M.; Dalen, J. E. Heparin and low-molecular-weight heparin: Mechanisms of action, pharmacokinetics, dosing considerations, monitoring, efficacy and safety. *Chest* **1998**, *114(Suppl 5),* 489S–510S.

180. Cohen, M. The role of low-molecular-weight heparins in arterial diseases: Optimizing antithrombotic therapy. *Thromb Res* **2000**, *100*, 131–139.

181. Boneu, B. Low molecular weight heparins: Are they superior to unfractionated heparins to prevent and to treat deep vein thrombosis? *Thromb Res* **2000**, *100,* V113–V120.

182. Nurmohamed, M. T.; ten Cate, H.; ten Cate, J. W. Low molecular weight heparin(oid)s: Clinical investigations and practical recommendations. *Drugs* **1997,** *53*, 736–751.

183. Thomas, D. P. Does low molecular weight heparin cause less bleeding? *Thromb Haemost* **1997**, *78*, 1422–1425.

184. Leizorovicz, A.; Haugh, M. C.; Chapuis, F. R.; Samama, M. M.; Boissel, J. P. Low molecular weight heparin in prevention of perioperative thrombosis. *Br Med J* **1992,** *305*, 913–920.

185. Turpie, A. G. Anticoagulants in acute coronary syndromes. *Am J Cardiol* **1999**, *84*, 2M–6M.

186. Bick, R. L.; Fareed, J. Low molecular weight heparins: Differences and similarities in approved preparations in the United States. *Clin App Thromb Hemost* **1999**, *5(Suppl 1),* S63–S66.

187. United States Food and Drug Administration: Information on heparin sodium injection. Available on the FDA website at: www.fda.gov/cder/drug/infopage/heparin/default.htm (Accessed on October 15, 2008).

188. United States Food and Drug Administration. Medwatch. Available at: www.fda.gov/medwatch/safety/2008/heparin_recall.htm (Accessed on October 15, 2008).

189. Guerrini, M.; Beccati, D.; Shriver, Z.; Naggi, A.; Viswanathan, K.; Bisio, A.; Capila, I.; Lansing, J. C.; Guglieri, S.; Fraser, B.; Al-Hakim, A.; Gunay, N. S.; Zhang, Z.; Robinson, L.; Buhse, L.; Nasr, M.; Woodcock, J.; Langer, R.; Venkataraman, G.; Linhardt, R. J.; Casu, B.; Torri, G.; Sasisekharan, R. Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. *Nat Biotechnol* **2008***, 26*, 669-675.

190. Kishimoto, T. K.; Viswanathan, K.; Ganguly, T.; Elankumaran, S.; Smith, S.; Pelzer, K.; Lansing, J. C.; Sriranganathan, N.; Zhao, G.; Galcheva-Gargova, Z.; Al-Hakim, A.; Bailey, G. S.; Fraser, B.; Roy, S.; Rogers-Cotrone, T.; Buhse, L.; Whary, M.; Fox, J.; Nasr, M.; Dal Pan, G. J.; Shriver, Z.; Langer, R. S.; Venkataraman, G.;

Austen, K. F.; Woodcock, J.; Sasisekharan, R. Contaminated heparin associated with adverse clinical events and activation of the contact system. *N Engl J Med* **2008**, *358*, 2457-2467.

191. Zhang, Z.; Weïwer, M.; Li, B.; Kemp, M. M.; Daman, T. H.; Linhardt, R. J. Oversulfated chondroitin sulfate: impact of a heparin impurity, associated with adverse clinical events, on low-molecular-weight heparin preparation. *J Med Chem* **2008**, *51*, 5498-5501.

192. Blick, S. K.; Orman, J. S.; Wagstaff, A. J.; Scott, L. J. Spotlight on fondaparinux sodium in acute coronary syndromes. *Bio Drugs* **2008**, *22,* 413-415.

193. Yusuf, S.; Mehta, S. R.; Chrolavicius, S.; Afzal, R.; Pogue, J.; Granger, C. B.; Budaj, A.; Peters, R.J.; Bassand, J. P.; Wallentin, L.; Joyner, C.; Fox, K. A. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *New Eng J Med* **2006***, 354*, 1464-1476.

194. Chung, T. L.; Holton, L. H 3rd, Silverman, R. P. The effect of fondaparinux versus enoxaparin in the survival of a congested skin flap in a rabbit model. *Ann Plast Surg* **2006**, *56,* 312-315.

195. Samama, M. M.; Gerotziafas, G. T. Evaluation of the pharmacological properties and clinical results of the synthetic pentasaccharide (fondaparinux). *Thromb Res* **2003***, 109*, 1-11.

196. Turpie, A. G.; Eriksson, B. I.; Lassen, M. R.; Bauer, K. A. Fondaparinux, the first selective factor Xa inhibitor. *Curr Opin Hematol* **2003**, *10*, 327-332.

197. Gallus, A. S.; Coghlan, D. W. Heparin pentasaccharide. *Curr Opin Hematol* **2002**, *9*, 422-429.

198. de Kort, M.; Buijsman, R. C.; van Boeckel, C. A. Synthetic heparin derivatives as new anticoagulant drugs. *Drug Discov Today* **2005**, *10*, 769-779.

199. Walenga, J. M.; Jeske, W. P.; Fareed, J. Short- and long-acting synthetic pentasaccharides as antithrombotic agents. *Expert Opin Investig Drugs* **2005**, *14,* 847-858.

200. Gerotziafas, G. T.; Samama, M. M. Heterogeneity of synthetic factor Xa inhibitors. *Curr Pharm Des* **2005**, *11*, 3855-3876.

201. Herbert, J. M.; Herault, J. P.; Bernat, A.; van Amsterdam, R. G.; Lormeau, J. C.; Petitou, M.; van Boeckel, C.; Hoffmann, P.; Meuleman, D. G. Biochemical and pharmacological properties of SANORG 34006, a potent and long-acting synthetic pentasaccharide. *Blood* **1998**, *91,* 4197-4205.

202. 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 USA* **1980**, *77*, 6551–6555.

203. Atha, D. H.; Stephens, A. W.; Rimon, A.; Rosenberg, R. D. Sequence variation in heparin octasaccharides with high affinity from antithrombin III. *Biochemistry* **1984**, *23*, 5801–5812.

204. Lindahl, U.; Thunberg, L.; Backström, G.; Riesenfeld, J.; Nordling, K.; Bjork, I. Extension and structural variability of the antithrombin-binding sequence in heparin. *J Biol Chem* **1984**, *259*, 12368–12376.

205. Atha, D. H.; Lormeau, J-C.; Petitou, M.; Rosenberg, R. D.; Choay, J. Contribution of monosaccharide residues in heparin binding to antithrombin III. *Biochemistry* **1985**, *24*, 6723–6729.

206. Desai, U. R.; Petitou, M.; Björk, 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

207. 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.

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

209. Raghuraman, A.; Liang, A.; Krishnasamy, C.; Lauck, T.; Gunnarsson, G. T.; Desai, U. R. On designing non-saccharide, allosteric activators of antithrombin. *Eur J Med Chem* **2009,** *44*, 2626-2631.

210. 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.

211. 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.

212. 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.

213. Krishnasamy, C.; Gunnarsson, G. T.; Desai, U. R. Antithrombin activation with designed small organic activators: The design of a bicyclic-unicyclic isoquinoline based activator. *Abstracts of Papers, 230th ACS National Meeting,* Washington, DC, United States, Aug. 28-Sept. 1, **2005**.

214. Liang, A.; Raghuraman, A.; Desai, U. R. Capillary electrophoretic study of small, highly sulfated, non-sugar molecules interacting with antithrombin. *Electrophoresis* **2009**, *30*, 1544-1551.

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

216. Skinner, R.; Chang, W. S.; Jin, L.; Pei, X.; Huntington, J. A.; Abrahams, J. P.; Carrell, R. W.; Lomas, D. A. Implications for function and therapy of a 2.9 Å structure of binary-complexed antithrombin. *J Mol Biol* **1998***, 283*, 9-14.

217. Jin, L.; Abrahams, J. P.; Skinner, R.; Petitou, M.; Pike, R. N.; Carrell, R. W. The anticoagulant activation of antithrombin by heparin. *Proc Natl Acad Sci USA* **1997***, 94*, 14683-14688.

218. 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.

219. 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.

220. Esko, J. D.; Selleck, S. B. Order out of chaos: Assembly of ligand binding sites in heparan sulfate. *Annu. Re*V. *Biochem.* **2002**, *71*, 435-471.

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

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

223. 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.

224. 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.

225. J. R. Fromm, R. E. Hileman, E. E. O. Caldwell, J. M. Weiler, R. J. Linhardt, Arch. Biochem. Biophys. 1995, 323, 279-287.

226. 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.

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

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

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

230. Meagher, J. L.; Huntington, J. A.; Gettins, P. G. W. Role of arginine 132 and lysine 133 in heparin binding to and activation of antithrombin. *J Biol Chem* **1996**, *271*, 29353-29358.

231. Schedin-Weiss, S.; Desai UR, Bock SC, Olson ST, Björk 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.

232. Jairajpuri, M. A.; Lu, A.; Desai, U. R.; Olson, S. T.; Björk, I.; Bock, S. C. Antithrombin III phenylalanines 122 and 121 contribute to its high affinity for heparin and conformational activation. *J Biol Chem* **2003***, 278*, 14941-14950.

233. 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.

234. Grootenhuis, P. D. J.; van Boeckel, C. A. A. Constructing a molecular model of the interaction between antithrombin III and a potent heparin analogue. *J Am Chem Soc* **1991***, 113*, 2743.

235. Bode, W.; Turk, D.; Karshikov, A. The refined 1.9-A X-ray crystal structure of D-Phe-Pro-Arg chloromethylketone-inhibited human alpha-thrombin: structure analysis, overall structure, electrostatic properties, detailed active-site geometry, and structure-function relationships. *Protein Sci* **1992**, *1*, 426-471.

236. Ye, J.; Liu, L. W.; Esmon, C. T.; Johnson, A. E. The fifth and sixth growth factor-like domains of thrombomodulin bind to the anion-binding exosite of thrombin and alter its specificity. *J Biol Chem* **1992**, *267*, 11023-11028.

237. Huntington, J. A. Molecular recognition mechanism of thrombin. *J Thromb Haemost* **2005**, *3*, 1861-1872.

238. Hogg, P. J.; Jackson, C. M. Fibrin monomer protects thrombin from inactivation by heparin-antithrombin III. Implications for heparin efficacy. *Proc Natl Acad Sci USA* **1989,** *86*, 3619-3623.

239. Vacca, J. P. New advances in the discovery of thrombin and factor Xa inhibitors. *Curr Opin Chem Biol* **2000**, *4*, 394-400.

240. Hogg, P. J.; Jackson, C. M.; Labanowski, J. K.; Bock, P. E. Binding of fibrin monomer and heparin to thrombin in a ternary complex alters the environment of the thrombin catalytic site, reduces affinity for hirudin, and inhibits cleavage of fibrinogen. *J Biol Chem* **1996***, 271*, 26088-26095.

241. Hogg, P. J.; Jackson, C. M. Formation of a ternary complex between thrombin, fibrin monomer, and heparin influences the action of thrombin on its substrates. *J Biol Chem* **1990**, *265*, 248-255.

242. Henry, B. L.; Monien, B. H.; Bock, P. E.; Desai, U. R. A novel allosteric pathway of thrombin inhibition. Exosite II mediated potent inhibition of thrombin by chemo-enzymatic, sulfated dehydropolymers of 4-hydroxycinnamic acids. *J Biol Chem* **2007**, *282*, 31891-31899.

243. Monien, B. H.; Henry, B. L.; Raghuraman, A.; Hindle, M.; Desai, U. R. Novel chemo-enzymatic oligomers of cinnamic acids as direct and indirect inhibitors of coagulation proteinases. *Bioorg Med Chem* **2006**, *14*, 7988-7998.

244. Streusand, V. J.; Björk, I.; Gettins, P. G.; Petitou, M.; Olson, S. T. Mechanism of acceleration of antithrombin-proteinase reactions by low affinity heparin. Role of the antithrombin binding pentasaccharide in heparin rate enhancement. *J Biol Chem* **1995,** *270*, 9043-9051.

245. Li, W.; Adams, T. E.; Nangalia, J.; Esmon, C. T.; Huntington, J. A. Molecular basis of thrombin recognition by protein C inhibitor revealed by the 1.6-Å structure of the heparin-bridged complex. *Proc Natl Acad Sci USA* **2008***, 105*, 4661-4666.

246. Richardson, J. L.; Kröger, B.; Hoeffken, W.; Sadler, J. E.; Pereira, P.; Huber, R.; Bode, W.; Fuentes-Prior, P. Crystal structure of the human α-thrombin-haemadin complex: An exosite II-binding inhibitor. *EMBO J* **2000***, 19*, 5650-5660.

247. 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 USA* **2002**, *99*, 11079-11084.

248. Li, W.; Johnson, D. J. D.; 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.

249. Johnson, D. J. D.; Langdown, J.; Li, W.; Luis, S. A.; Baglin, T. P.; Huntington, J. A. Crystal structure of monomeric native antithrombin reveals a novel reactive center loop conformation. *J Biol Chem* **2006**, *281*, 35478-35486.

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

251. McCoy, A. J.; Pei, X. Y.; Skinner, R.; Abrahams, J.-P.; Carrell, R. W. Structure of β-antithrombin and the effect of glycosylation on antithrombin's heparin affinity and activity. *J Mol Biol* **2003**, *326*, 823-833.

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

253. Johnson, D. J. D.; Li, W.; Adams, T. E.; Huntington, J. A. Antithrombin-S195A factor Xa-heparin structure reveals the allosteric mechanism of antithrombin activation. *EMBO J* **2006**, *25*, 2029-2037.

254. Langdown, J.; Belzar, K. J.; Savory, W. J.; Baglin, T. P.; Huntington, J. A. The critical role of hinge-region expulsion in the induced-fit heparin binding mechanism of antithrombin. *J Mol Biol* **2009***, 386*, 1278-1289.

255. 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.

256. http://www.ccdc.cam.ac.uk/products/life_sciences/gold/

APPENDIX A

ABREVIATIONS

AT: Antithrombin

HBS: Heparin binding site

PBS: Pentasaccharide Binding Site

EHBS: Extended Heparin Binding Site

GAG: Glycosaminoglycan

DEFGH (H_5): Natural sequence-specific heparin pentasaccharide

LMWHs: Low Molecular Weight Heparins

ISOQ: Tetrahydroisoquinoline

QSAR: Quantitative Structure-Activity Relationship

QSPR: Quantitative Structure-Property Relationship

VITA

Chandravel Krishnasamy was born on 11 May, 1976 in Eluthur village, Tamil Nadu, INDIA. He is a citizen of India. He obtained his Bachelor of Pharmacy degree from the K.M. College of Pharmacy in Madurai, Tamil Nadu in 1997. He also received his Master of Pharmacy degree from Shri G. S. Institute Technology and Sciences in Indore, Madhya Pradesh 1999. He worked as a Lecturer at College of Pharmacy, IPS Academy, Indore, between 1999 and 2002. He began graduate studies in Department of Medicinal Chemistry at Virginia Commonwealth University, in Richmond, USA in Janaury 2003.