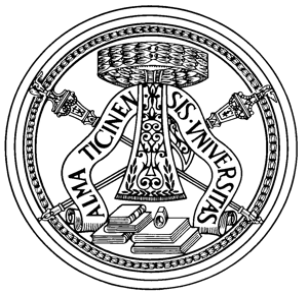


UNIVERSITÁ DEGLI STUDI DI PAVIA

DOTTORATO IN SCIENZE CHIMICHE E FARMACEUTICHE

XXXII CICLO

Coordinatore: Chiar.mo Prof. Mauro Freccero

APPLICATION OF NMR SPECTROSCOPY TO STUDY PROTEIN-LIGAND INTERACTION OF HEPARAN SULFATE OLIGOSACCHARIDES, OBTAINED BY CHEMOENZYMATIC SYTHESIS

Academic Tutor:

Prof. Lino Colombo

Industrial Tutor:

Dr. Marco Guerrini

Industrial co-Tutor:

Stefano Elli PhD

Tesi di dottorato di

Eduardo Stancanelli

Anno accademico 2018/2019

Index

A Mio Padre…Per sempre

1. Introduction

1.1. General Background

Heparin (Hep) and heparan sulfate (HS) are linear, acidic, structurally related polysaccharides belonging to the glycosaminoglycan (GAG) family. Heparin and HS differ in their cellular origin and location. Heparin is synthesised as a polymer (80 KDa) by mast cells, where it is found in intracellular secretory granules. In contrast, HS is a ubiquitous component of all cells and tissues being present on cell surfaces and in the extracellular matrix[1]. The structural diversity of heparin and heparan sulfate (HS) is central for the varied range of physiological processes that these molecules modulate. HS works as co-receptor of various trans-membrane tyrosine kinase receptors as well as to localize proteins at the cell surfaces and thus promoting protein-receptor interactions. Exogenously administrated heparins are isolated from animal organs (predominantly porcine intestinal mucosa) and are currently used as anticoagulant and antithrombotic drugs. The molecular basis for the anticoagulant activity of the heparin was explained in the early 1980s when a pentasaccharide sequence within heparin chains was identified as being crucial for binding and activating antithrombin, leading to accelerated inhibition of the coagulation cascade[2,3]. Over the last decades, heparin and heparan sulfate have been shown to interact with many important proteins thereby regulating a wide range of biological activities including cell proliferation, inflammation, angiogenesis, viral infectivity and development. Understanding the structure-activity relationship of these molecules might lead to design new HEP-like potential drugs.

1.2. Glycosaminoglycans

GAGs mainly consist of linear carbohydrate chains covalently linked, made up by alternating sequences of uronic acid and glycosamine residues. In accord to the glycosamine contained in their chains, GAGs can be classified as:

- **glucosaminoglycans**, characterized by D-glucosamine (GlcN), including HEP and HS, as well as hyaluronic acid and keratan sulfate;
- **galactosaminoglycans**, characterized by D-galactosamine (GalN), including chondroitin sulfate and dermochondan sulfate (Dermochondan is the recommended term indicating the copolymeric nature of dermatan sulfate, replacing dermatan [4,5]).

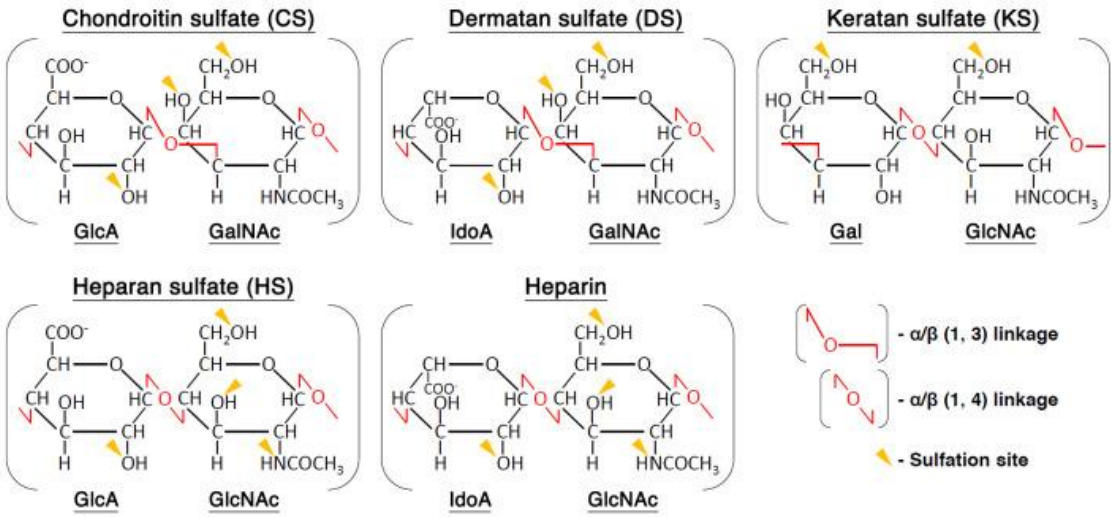


Figure 1. *Disaccharide repeating unit characterising gluco- and galactosaminoglycans.*

Structural heterogeneity of the GAGs is determined by variation of prevalent disaccharide repeating units and chain length. However, each single chain generally contains differently sulfated and acetylated glucosamines or galactosamines as well as different types of uronic acids (i.e. D-glucuronic acid (GlcA), L-iduronic acid (IdoA), 2-O-sulfated-L-iduronic acid (IdoA2S), and in lower amount

also 2-O-sulfated-D-glucuronic acid (GlcA2S)[6]. Hyaluronic acid has the simplest structure with further modification occurring in Golgi apparatus. Heparan sulfate (HS) and heparin (Hep) contain repeating disaccharide units of N-acetylglucosamine and glucuronic acid in heparan sulfate, while iduronic acid is present in heparin. Sulfation of the various hydroxyl groups or the amino group present on the glucosamine compound of HS/Hep modulate their bioactive function. Chondroitin sulfate (CS) and dermatan sulfate (DS) are similar in structural composition to HS. Their disaccharide repeat consists of N-acetylgalactosamine linked to an iduronic acid in CS and glucuronic acid in DS. Keratan sulfate (KS) contains the disaccharide repeat consisting of galactose and N-acetylglucosamine. These differences in chemical composition largely determine biological and pharmacological activities through the interaction between GAGs and different kind of proteins such as growth factors, enzymes, morphogens, cell adhesion molecules and cytokines.

1.2.1. HS/Hep Biosynthesis

HS/Hep biosynthesis occurs in the endoplasmic reticulum and proceeds through the Golgi apparatus. Although genomics has identified the enzymes involved in glycosaminoglycan synthesis of vertebrates and invertebrates, the regulation of the process is not completely understood. In the early stage xylose is transferred to a serine acceptor on the core protein. Then two galactose units and a glucuronic acid moiety are added to the newborn chain. The completed tetrasaccharide, named the linkage region, is often referred to as a stem or linker, because it is common to HS/Hep and chondroitin/dermatan sulfate synthesis. A specific copolymerase, consisting of two proteins EXT1 and EXT2, catalyzes the polymerization of repeating disaccharide units N-acetyl-D-glucosamine/D-glucuronic acid $[\beta\text{-GlcA-}(1\rightarrow4)\text{-}\alpha\text{-GlcNAc-}(1\rightarrow4)]_n$, allowing the formation of HS/Hep precursor. Before chain elongation is completed, early modification steps occur. The first is carried out by one

of the four N-deacetylases/N-sulfotransferases (NDSTs) and consists in N-deacetylation and N-sulfation explicated by a single protein. This is followed by epimerization of some of the uronic acid residues that will be converted from glucuronic to iduronic acids. Some of the iduronate is then sulfated at the 2-O position. Heparan Sulfate 6-O-sulfotransferase (HS6ST) is then responsible of the 6-O-sulfation of glucosamine residues. Finally, and rarely, the 3-O-sulfation takes place, but interestingly, seven mammalian enzymes are capable of completing this step[7], each of which in specific sulfated GAG sequences. In the case of heparin, characteristic chain modification is more extensive than in HS, for that reason the likelihood of obtaining sequences recognized by HS3ST1 and thus more 3-O-sulfation is higher.[8–10]. These processes occur in a partially stepwise fashion, where the product of a given reaction will provide the substrate for subsequent modifications. However, most of the reactions do not show full conversion, in fact, a fraction of the potential substrate residues in each step escapes modification. This partial polymer modification is at the base of the structural heterogeneity involving both non-sulfated and GAG chains with variable degree of sulfation. Starting from the linkage region towards the non-reducing end, HS/Hep chain structures present an unmodified $[GlcA\text{-}GlcNAc]_n$ domain (NA domain) followed by a NS/NA domain in which the most represented disaccharide are made up by. The trisulfated disaccharide IdoA2S-GlcN,6S is predominant in the following region, named the NS domain, following region, named NS domain, or regular region [Figure 3].

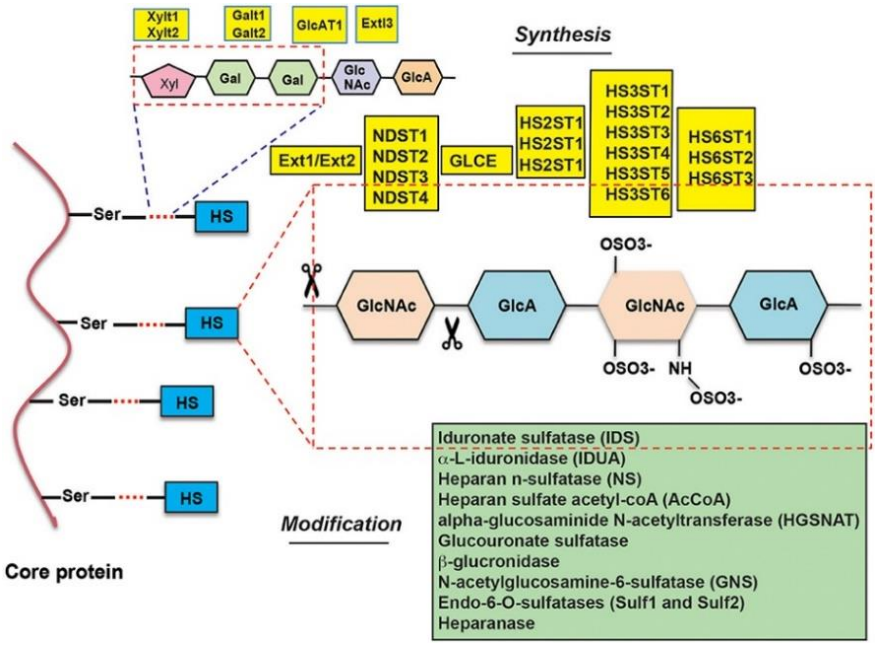


Figure 2. *Idealized representation of a heparin chain constituted of N-acetylated (NA), N-sulfated (NS), and mixed NA/NS domains. Detailed structures of disaccharide units characterizing the different domains and of the AT-binding pentasaccharide AGA*IA.*

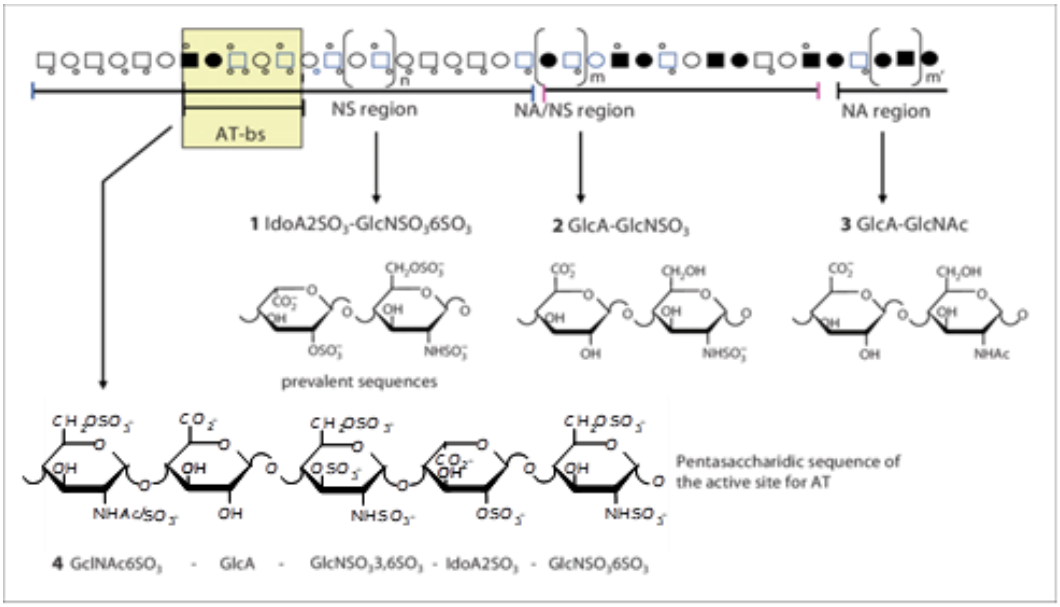


Figure 3*. Idealized representation of a heparin chain constituted of N-acetylated (NA), N-sulfated (NS), and mixed NA/NS domains. Detailed structures of disaccharide units characterizing the different domains and of the AT-binding pentasaccharide AGA*IA.*

More recently for heparin, this scheme was revisited, observing that less sulfated chains can be present in different position and length along the chain, according the organ or animal source. Some of the chains present a pentasaccharide sequence frequently labelled in literature as AGA*IA[11], characterized by the trisulfated glucosamine GlcNS3S6S residue, that is part of the sequence: GlcNAc/NS,6S-GlcA-GlcNS,3S,6S-IdoA2S-GlcNS,6S. This pentasaccharide is the minimal sequence required to binds and activate AT, and it plays a pivotal role in heparin anticoagulant/antithrombotic activity, inactivating some key factors of the coagulation system. In commercially available HEP and HS preparations, the molecular weight (MW) is centred at about 15-18 KDa and 25-30KDa, respectively, and with a polydispersity of 1.3-1.5[12]. Moreover, HEP and HS structures have different distribution of residues, the former has a more regular sequence, while the latter show longer chains and a block composition, in which regions rich in N-sulfated glucosamines (NS) alternate with sequences that contain un-modified glucosamines (NA), that are generally made of about 9-10 disaccharide units. On the contrary, Hep is richer of NS domain and has shorter NA domain (1-2 disaccharide units) compared to HS[13]. The conversion of β-D-glucuronate to α-L-iduronate by the C5 epimerase [14] may seem trivial, however it profoundly affects the conformational and dynamic properties of the resulting polymer. A full analysis of the NMR spectra of heparin and synthetic heparin mono- and oligosaccharides combined with molecular mechanics calculations clearly demonstrated that IdoA and IdoA2S residues are in an equilibrium among three conformations of comparable energy $^{1}C_{4}$, $^{4}C_{1}$ and $^{2}S_{0}$.[15,16] [Figure 4]. The population of different conformers ($^{1}C_{4}$, $^{4}C_{1}$ and $^{2}S_{0}$) could broadly differ depending on the presence of sulfate groups on the same residue or in that of the adjacent residues as well as on extrinsic factors such as ionic strength and specific counter ions (i.e. Ca^{+2})[17]. The conformation of iduronate residue induce perturbation of the 3D helix structure of HS/Hep chains. The orientations of the substituent groups

significantly differ in the two conformations, involving, i.e., different spacing between sulfate and/or carboxylate groups[18] [Figure 5].

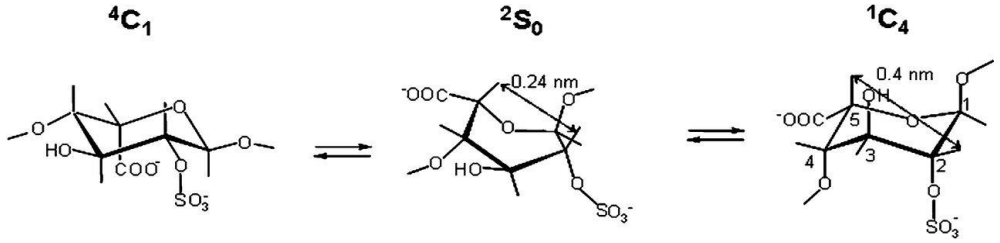


Figure 4. *The flexibility of iduronic acid, which can be found as $^1C_4$ and $^4C_1$ chairs and $^2S_0$ skew-boat conformers.*

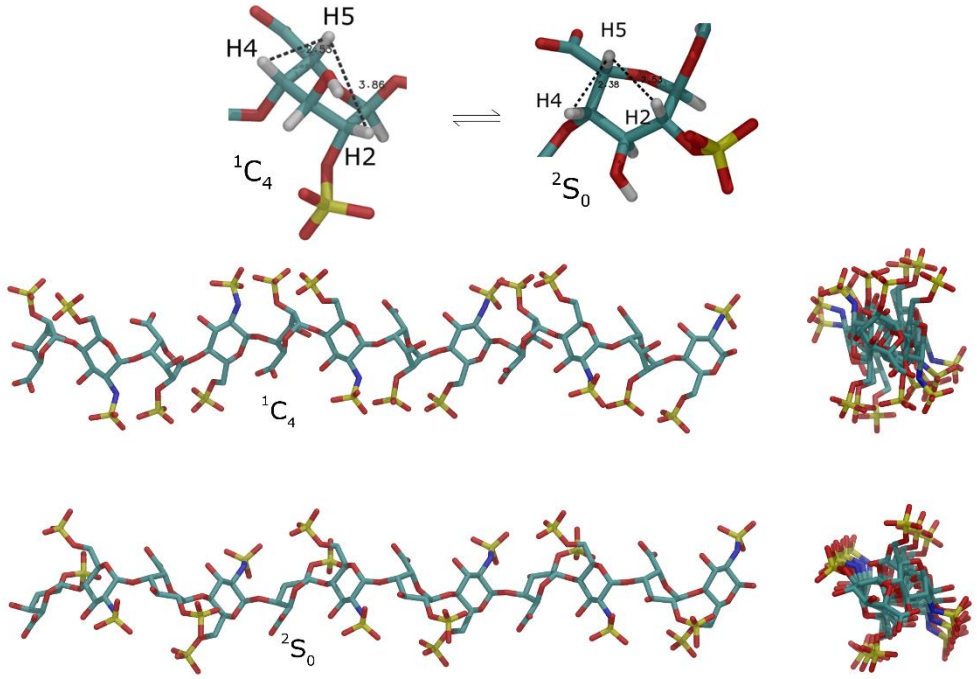


Figure 5. Different conformation of Iduronc acid residues and their influence on the orientation of Heparin chain.

^{1}H NMR spectroscopic analysis and molecular modelling showed that in the biologically active pentasaccharide AGA*IA, the iduronic acid moieties exist mainly in an equilibrium between the $^1C_4$ and $^2S_0$ conformations in which the latter is dominant[19,20]. Nonetheless, significant evidence suggests

that the $^{2}S_{0}$ form corresponds to the active conformation of L-iduronic acid when the pentasaccharide is bound to AT[21]. Indeed, synthetic analogues with an IdoA2S locked in $^{1}C_{4}$ or $^{4}C_{1}$ conformation were found unable to activate AT [22], whereas an equivalent derivative with L-iduronic acid locked in the $^{2}S_{0}$ conformation had similar potency to the parent compound AGA*IA[23]. This behaviour was perceived to have an important role on binding and associated biological properties of iduronic acid-containing GAGs. The conformational flexibility of the iduronate pyranose is supposed to facilitate the most effective docking of the anionic groups of GAGs to the appropriate basic groups of the proteins. The degree of freedoms that a HS/Hep chain applies to fit the receptor binding site of AT involve limited rotations around glycosidic linkages to slightly adjust the glycosidic backbone to the receptor site geometry, but also the selection of the most suitable conformation of iduronic acid residues. The analysis of several crystal structures of heparin oligosaccharide co-crystallised with AT showed that the linear propagation of GAG chain is interrupted by "kinks" of the heparin helix motif associated with the $^{2}S_{0}$ conformation, that an iduronate residue belonging to the AGA*IA sequence, assume in bound state with AT [24]. Such kinks play their main role in the achievement of the appropriate arrangement of GAG chains in protein binding sites.

1.2.2. Interaction with Protein

HS is crucial for embryonic development[25] and is required for the normal physiology of adult animals [Figure 6]. The expression and structure of HS change promptly during the animal development indicating a possible role in cell and system signalling[26,27].The role of HS emerges from its ability to interact and regulate the activity of a vast number of extracellular proteins including growth factors and extracellular matrix components. This role is also supported by the pervasive evidence that substoichiometric modification during the HS biosynthesis produce characteristic domains, which

vary in size and number within each chain allowing HS/Hep able to interact with many different proteins. In fact, heparin and HS endowed with the capacity to form specific complex called 'interactome' that regulate multiple functional proteins such as coagulation cofactors, chemokines, and growth factors. The interactome of a molecule is the whole set of interactions a molecule is involved in. The GAG regulation activity on some of these protein is documented, even if a greater number of cases are currently under investigation or are completely unknown. In a study of Fernig the heparin/HS interactome was built combining a literature mining effort with an affinity proteomics approach for the identification of heparin/HS-binding proteins (HBPs). Obtained data were used to generate a comprehensive list of the interactions between heparin/HS and proteins that is currently composed of a non-redundant list of 435 human proteins[28,29], but that number is still growing.. HS/Hep have been implicated in many human diseases including cardiovascular disease, chronic obstructive pulmonary disease, cancer, infectious disease, amyloidosis, and HIV/AIDS.

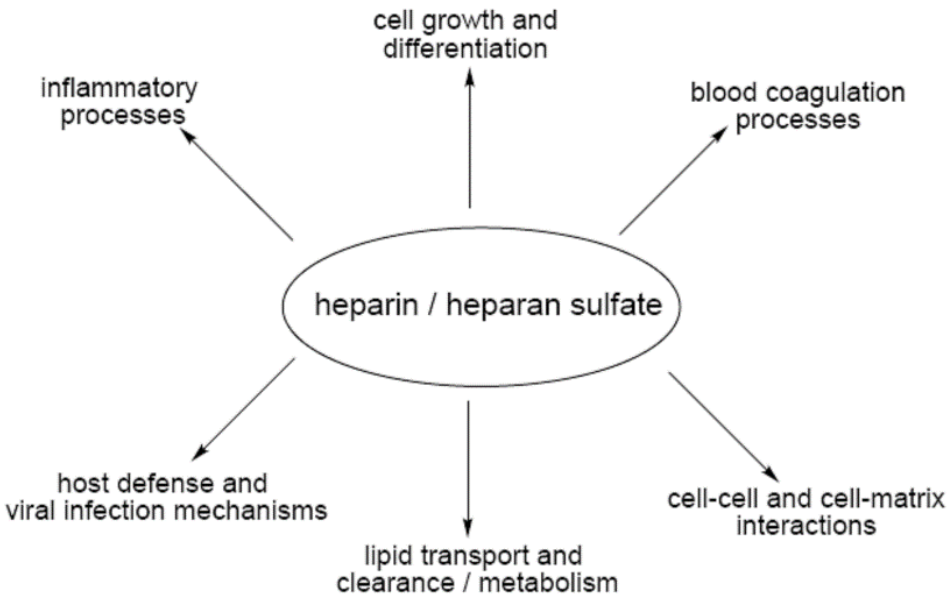


Figure 6. *Biological activities modulated by the interaction of protein with heparin and heparan sulfate.*

Biological processes may be widely influenced by this HS interactome because, while HS interacts with a large variety of proteins, these HS-binding proteins themselves interact with each other, forming a complex protein–protein interaction networks[28]. From a functional point of view, the heparin/HS interactome is strongly associated with biological processes belonging to multicellular organisms and with pathways that are crucial for the conversion of extracellular hint into intracellular signalling events and finally into a phenotypic response. These processes and pathways seems to be strongly related to complex biological phenomena of higher organisms such as development and the immune response, and they are consequently linked to pathological conditions such as cancer and neurodegenerative disorders. In this study authors collected evidences that Heparan sulfate proteoglycans are involved in all the key processes required for the establishment of neural connectivity including axon guidance, neuron-target interaction, and synapse development[30]. Understanding how interactome really works should be useful to elucidate all the mechanisms involved in the control of physiological and pathological processes in complex diseases.

1.2.3. Antithrombin Activation

The family of serine protease inhibitors known as serpins is represented in all branches of life and predominates in the higher organism, including man. The mechanism of the serpins involves a significant conformational change, known as “activation”, by which an extended loop of the protein become ready to literally ‘capture’ the target protease, that at the end of the process will be inactivated by an irreversible reaction. The mechanism of capture could be compared to a spring-loaded mousetrap: because the active state of the serpin is metastable, and the energy released upon conversion to its more stable form, is then used to trap the protease. This complex mechanism

provides more advantages in comparison to the simple lock-and-key type mechanism, in fact, the double dependence on serpin and protease conformational change is exploited for regulation, signalling and clearance. Antithrombin is the most important anticoagulant molecule in the blood of mammals. In co-operation with its co-factor, heparin, it is the most effective, well-regulated inhibitor of the coagulation factors, such as factors IXa, Xa, XIa, IIa (thrombin) and XIIa [Figure 7]. It has a well-defined structure, consisting of three β-sheets, nine α-helices and a reactive centre loop (RCL). Antithrombin circulates in a low activity state (AT latent form) until it encounters a specific sequence on heparin or heparan sulfate, which induces an activating conformational change. Heparin binding results in extensions of the D-helix, and the expulsion of the reactive center loop from the main β-sheet. The release of the RCL reorients the reactive center allowing it to capture the target protease (fIIa, IXa, and Xa) exposing part of this loop to the reactive center of the protease, (AT active form), literally biting it, and [Figure 8] prepare it for the next step. It is important to underline that the activated Antithrombin (AT*) has an affinity toward its proteases approximately 1000-fold greater compared to the latent AT.

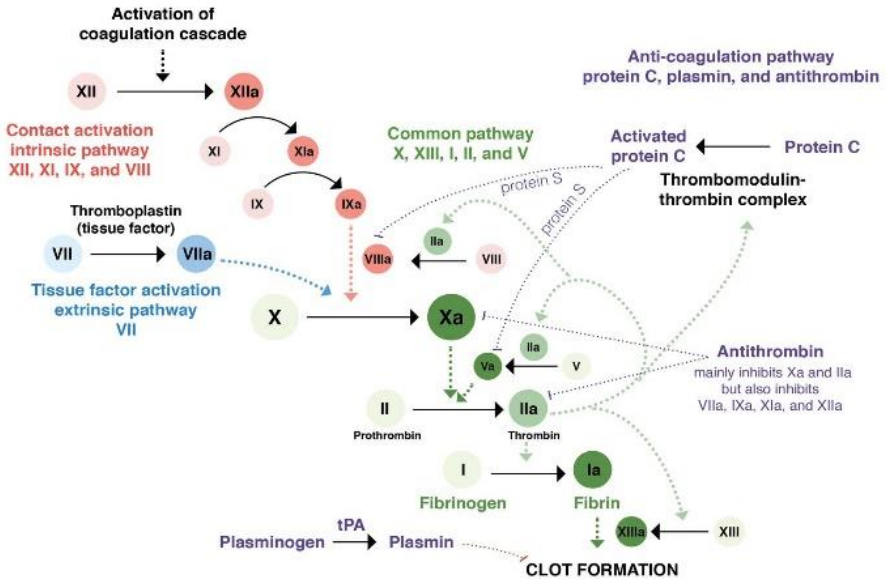


Figure 7. *The role of antithrombin in the coagulation cascade. Antithrombin is inhibitory to factors IIa (thrombin) and Xa, and to a lesser extent factors IXa, XIa and XIIa.*

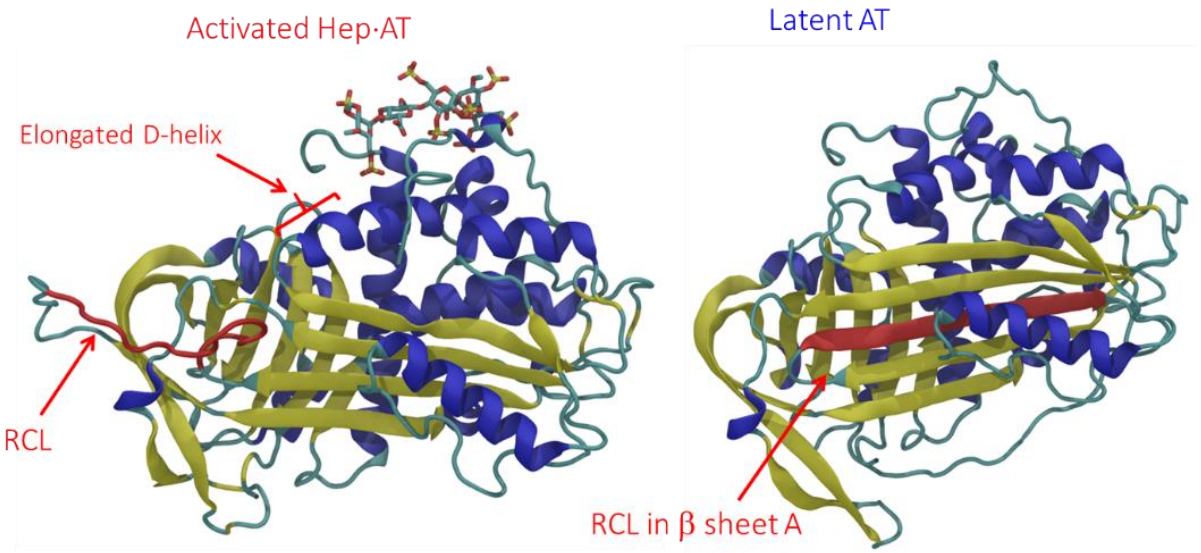


Figure 8. *Structure of antithrombin alone and in complex with heparin pentasaccharide. The antithrombin is shown alone on the right, with the reactive center loop (RCL) shown in magenta. The AT-hep pentasaccharide complex with the extruded RCL is shown on the left.*

In the late 1970s different research groups discovered, independently, that heparin heterogeneously interacts with AT. They were able to separate the high affinity fractions (HA) of heparin from the low affinity ones (LA). The anticoagulant activity is largely the prerogative of the HA fraction, which constitutes only about one third of the unfractionated heparin (UFH)[31,32]. Initially that was a landmark finding and a surprising one, since the compositional differences between HA and LA chains were subtle and not easily detectable. Throughout the years, differences in composition of both chains were progressively observed, and in the early 1980s Lindahl's group established that the Antithrombin Binding Region (ATBR) was a pentasaccharide containing the characteristic 3-O-sulfated GlcN, residue [Figure 9]. That pentasaccharide was elucidated through identification of the consensus sequences of hexa- and octasaccharides isolated by heparin degradation and affinity chromatography on immobilized AT[8,33]. The structure of ATBR [Figure 9] was confirmed through to the synthesis of the pentasaccharide sequence performed by Pierre Sinay and Choay groups, demonstrating that the glycan and particularly the 3-O-sulfate group of the central glucosamines are required for binding

and activation of AT[34]. Several analogues of the pentasaccharide in which a selected sulfate or carboxylate group is missing, were synthesized. This allowed to establish that some groups are essential for the binding and activation of AT, whereas other groups are known to only "contribute" to increase the affinity to AT and the biological activity [Figure 10].

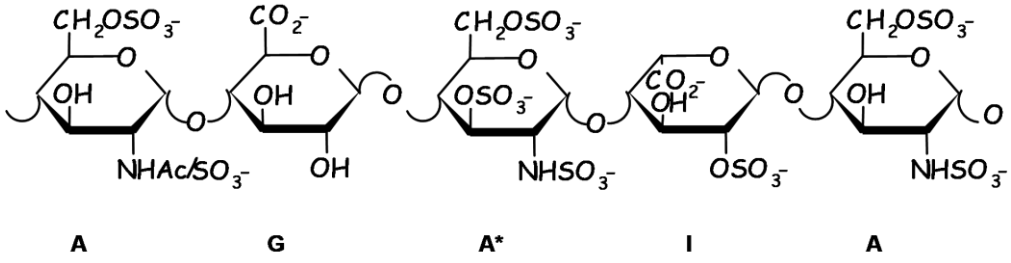


Figure 9. *Pentasaccharide sequence containing the typical 3-O-sulfated GlcN, residue was found to represent the minimum sequence that binds and activate AT.*

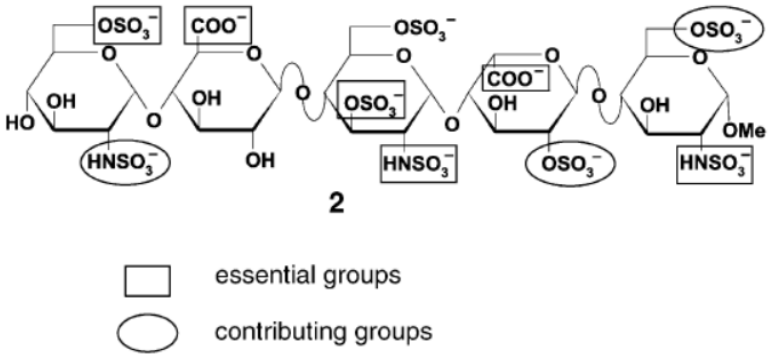


Figure 10. *Structures of ATBR pentasaccharide. Boxes highlight groups that are absolutely essential for AT activation, whereas circles identify groups that only help to increase the biological activity.*

1.2.4. Molecular Details of Antithrombin Activation

A deep understanding at the molecular level of the mechanism by which heparin binds and activates antithrombin was obtained from the analysis of the crystal structures of antithrombin alone and in complex with the AGA*IA pentasaccharide[35]. In the crystal structure of the complex we can observe

that three regions are close together in the folded protein to form the receptor binding site (RBS), which is surrounded by positively charged Arg and Lys residues, and characterized by and elongated shape, which size host approximately 5-6 Heparin residues. These regions correspond to the N-terminal end of the protein, the A helix, and the D helix along with a loop at the N-terminal side of the latter helix. The basic residues in this site that interact with the pentasaccharide are Lys 11 and Arg 13 in the N-terminal end; Arg 46 and Arg 47 in the A helix; and Lys 114, Lys 125, and Arg 129 in the D helix and its N-terminal loop. The rigid non-reducing end trisaccharide, ANAc,6S-G-ANS,3,6S-, of the AGA*IA is initially attracted to the positively charged surface of the heparin-binding site of antithrombin and interacts directly with Lys 125, but with poor complementarity of the interacting surfaces. This initial binding then induces a conformational change in antithrombin [Figure 11]. The conformational change also generates a new complementary site involving mostly Lys 114 and then Arg 47 which allows the interaction with the flexible reducing-end disaccharide, I2S-ANS,6S. As already cited, this interaction is assisted by the conformational flexibility of the Iduronic acid residue, which adopt the $^{2}S_{0}$ skew-boat conformation, enhancing the contacts between the pentasaccharide and the basic amino acid residues of the RBS (Receptor Binding Site) of AT [Figure 11]. These events lie at the basis of AT activation and are responsible for the stabilization of the high affinity, activated conformation[36,37].

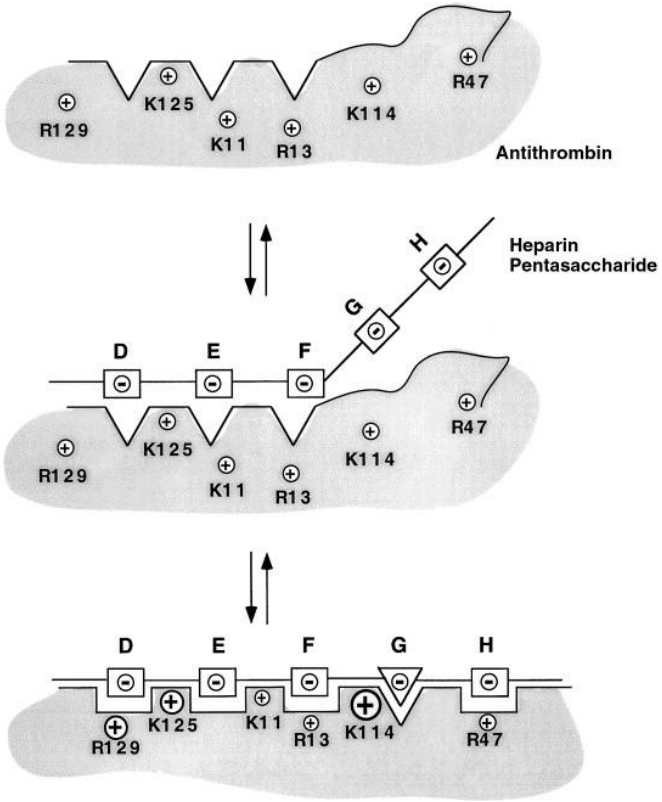


Figure 11. *Recognition phases of pentasaccharide sequence. The negatively charged reducing-end trisaccharide, ANAc,6S-G-ANS,3,6S-, is initially attracted and interacts directly with Lys 125. AT undergoes through a conformational change which allows the interaction with the flexible reducing-end disaccharide, I2S-ANS,6S*

As already mentioned, such conformational change also enhances the affinity of AT for thrombin and other serine proteases[36,38,39]. It is important to underline that, even if a short glycan such as the AGA*IA pentasaccharide, is proven to bind and activate AT, binding and inactivating the factor Xa, a heparin chain not shorter than a hexadeca-saccharide is still required to inhibitfactor IIa (thrombin) [23]. The inhibition of factor Xa is assumed to be largely responsible forthe antithrombotic effect, while inhibition of coagulation at the level of thrombin (IIa) seems to increase the risk of haemorrhage. The different mechanisms of heparin mediated acceleration of anti-Xa / anti-IIa activity were at the basis of the introduction of low-molecular weight heparins (LMWHs) as antithrombotic agents with an enhanced anti-Xa / anti-IIa activity ratio. Additionally, LMWHs were characterized by higher bioavailability in the blood stream in comparison to UFH, and for that reason the former class of

heparins nowadays are largely used in treatment of venous thromboembolism. The LMWHs have a mean molecular weight of 4.5 to 5 KDa, compared to the 15-20 KDa of UFH.

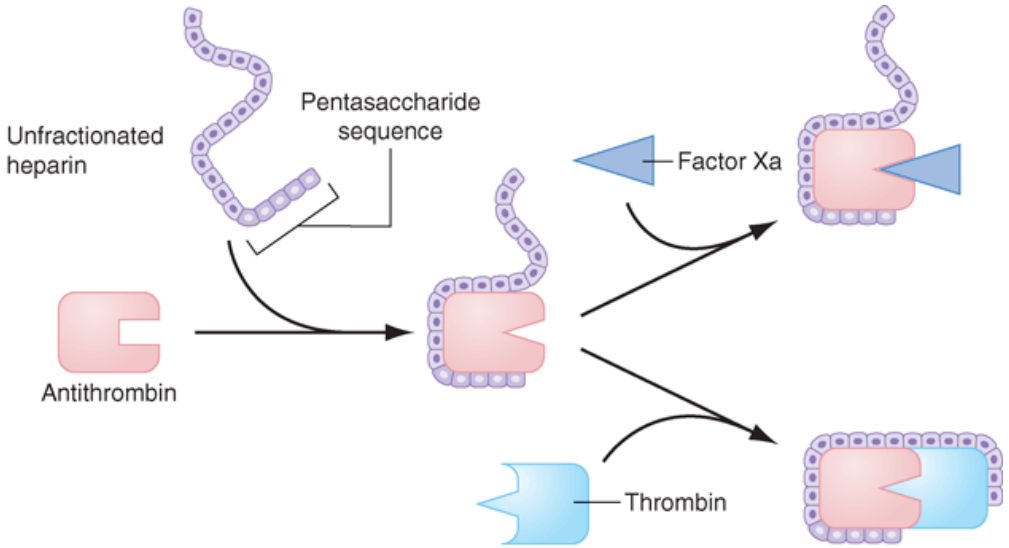


Figure 12. *Mechanism of action of Heparin. Heparin binds to antithrombin via its pentasaccharide sequence. This induces a conformational change in the reactive center loop of antithrombin that accelerates its interaction with factor Xa. Heparin simultaneously bind to antithrombin and thrombin. Only heparin chains composed of at least 18 saccharide units are of sufficient length to perform this bridging function. LMWHs are too short to bridge antithrombin to thrombin.*

The requirement for precisely targeted anticoagulant activity, that avoid the problems associated with nature-sourced heparins, led to the development of fondaparinux, a synthetic pentasaccharide derived from the antithrombin specific pentasaccharide AGA*IA. Fondaparinux (registered in USA and Europe with trade name Arixtra®), an ATBR analogue in which the reducing end was protected by a methyl group [Figure 13], is the first of a new class of selective factor Xa inhibitors. The availability of pure ATBR analogues obtained by chemical synthesis also allowed to investigate the Structure Activity Relationship (SAR), through the conformational analysis supported by X-ray diffraction, In the last twenty years, drug design has paid special attention to correlate the conformational properties of heparin like oligosaccharides containing the AGA*IA sequence, with their binding affinity and

interaction properties toward AT. To improve this knowledge, the effects of finer structural alteration of GAG oligosaccharides in their AGA*IA sequence and/or neighbouring residues, on their ability to bind and activate AT, represent a possible development for heparin-like anticoagulant drugs. The correlation between the structural variability of the heparin binding region and their interaction properties with AT was considered several years ago, when the oligosaccharide models were chemically synthesized obtained by partial enzymatic depolymerization of heparins. For example, substitution of the acetyl with a sulfate group at the N-position of the non-reducing glucosamine of the pentasaccharide increased the AT affinity by two folds, whereas lack of 6-O-sulfation in the 3-O sulfated glucosamine did not significantly affect the binding affinity toward AT[33]. However, the simplistic idea that the AT binding depends exclusively on the specific pentasaccharide, independently of the structure of the nearby residues in the heparin chain was recently refined by several studies on HS and heparin oligosaccharides, which demonstrated the active role in the binding and activation of AT, of residues extending the active sequence toward both its reducing and non-reducing sides[34,36,40]. Although the active pentasaccharide AGA*IA is taken as paradigm for a unique heparin sequence targeting AT, different binding epitope and contacts have been proposed for its interaction with AT in terms of position and conformation of sugar residues. Many AT-binding oligosaccharides have been isolated from LMWHs and ULMWHs. The "Ronzoni Institute", together with Sanofi played a pioneering role in the discovery and characterization of novel binding sequences isolated from LMWH and ULMWH[41–45]. In 2008, for example, an important study on different octasaccharides isolated from LMWH Enoxaparin was conducted. In this work four high purity octasaccharides isolated by size exclusion chromatography, high pressure liquid chromatography, and AT-affinity chromatography were selected. The conformational and AT binding properties of these octasaccharides were investigated by NMR (STD and transferred-NOESY) spectroscopy and

docking simulations. These results confirmed how residues that precede and follow the AGA*IA sequence contribute to the heparin molecular recognition by AT[40]. More recent studies suggested that negatively charged heparin chains and AT may interact in a non-canonical way[46]. In these latter studies, the possibility has been considered that sulfated residues in heparin sequences, different from that of the AT-binding site, may activate AT through nonspecific interactions.

1.2.5. Future

Unfractionated heparins (UFHs) and low molecular weight heparins (LMWHs) are complex mixtures of molecules prepared from animal tissues and used as pharmacological agents. Due to the poorly regulated supply chain of pharmaceutical heparin, in late 2007 and early 2008 the "heparin crisis" caused the death of about hundred people upon injection of contaminated heparin[47,48]. The heparin contaminant was identified as oversulfated chondroitin sulfate (OCSC) [49]. Even if the new guidelines of quality control standards for heparin were improved, another heparin crisis cannot be excluded, particularly considering the intrinsic heterogeneity of the heparin. The recent Chinese outbreak of African swine fever which resulted in the death of millions of pigs in that country has led to falling U.S. supplies of heparin, opening the possibility of a potential shortage of raw material and consequential possibility to cut heparin with not allowed other substances. In this scenario the possible introduction of new artificial heparin analogous or not previously admitted animal sources, could be a solution to overcome this problem. The first attempt in this direction corresponded to the synthesis of Fondaparinux, and other synthetic mimetics of ATBR pentasaccharide. However, even if FDA approved this pharmacophore as a drug, experience showed that it does not maintain all the clinical benefits of UFH and LMWH. Moreover, UFH remains the anticoagulant of choice for haemodialysis and open-heart surgery patients.

Progress was made in chemical synthesis to obtain new HS/Hep analogues: however, a pure chemical approach is challenging due to the complexity and countless of reaction involved. The new chemoenzymatic method recently proposed by the group of Professor J. Liu (North Carolina University UNC), overcoming the pure synthetic approaches, is capable of generating oligosaccharides with defined new structures to be used both as LMWH analogs or models to further investigate specific interactions with relevant proteins and enzymes involved in anticoagulant or other biological functions[50]. Furthermore, this strategy should also provide impetus to drug discovery for probing new Structure Activity Relationships[51]. Such approach could enable the opening of new therapeutic horizons for heparin and other glycosaminoglycans[52].

1.3. NMR GAG Characterization

The detailed analysis of GAGs is an extremely important issue because the overall sulfate content as well the sulfation pattern is likely to have a significant impact on biological effects of this class of molecule. GAG structural determination remains a challenge, mainly for the difficulty in determining their fine structure. According to our best knowledge there is not a single method available which allows the detailed characterization of all structural aspects of the intact GAG polysaccharides. Although many (chromatographic or electrophoretic) methods allow the estimation of the molecular weight, the determination of the overall sulfate contents or the determination of the building blocks, there is not a single method available able to differentiate all the isomers present in a polysaccharide mixture. Soft ionization mass spectrometry (MS) techniques such as electrospray ionization (ESI) or matrix-assisted laser desorption and ionization (MALDI) are useful methods to obtain information of GAG composition. Unfortunately, MS is exclusively applicable to GAG oligosaccharides, usually up

to 7-8 disaccharides units, while longer chains can be characterized only after chemical or enzymatic depolymerization.

Nuclear Magnetic Resonance (NMR) spectroscopy represents a very important technique in the structural characterization of GAGs. It is one of the few methods that provides information using the intact polymer without requiring any sort of chemical or enzymatic treatment. NMR can provide for both complex mixture and purified oligosaccharides information on the monosaccharide composition, presence of impurities, configuration at the anomeric center and even sequence. NMR spectroscopy is the leading technique to detect and quantify signals associated with major or minor structural sequences in heparins. 1D proton and carbon NMR spectroscopy have been used to describe and quantify the sulfation or acetylation degree present in heparin samples coming from different animal sources[53]. A more recent multidimensional approach (quantitative HSQC) involving the combined use of proton and carbon NMR, allowed the monosaccharide composition, sulfation pattern and the epimeric state configuration of the uronic acid residues to be determined[54]. Moreover, high field NMR analysis has also been applied to observe conformational properties of chain residues and glycosidic linkages. This could be achieved measuring the J-couplings constants and analyzing the inter-proton Nuclear Overhauser Effect spectra in which the transfer of nuclear spin polarization from one population of spin-active nuclei to another occurs via cross-relaxation. The NOE is particularly important in the elucidation and confirmation of the structures or configurations of organic and biological molecules. Finally, the NMR techniques, such as saturation transfer difference, water-logsy and transferred-NOEs provide information about the epitope binding and conformation of the ligand in interaction with proteins. Even if a relatively large amount of sample is required to run NMR experiments compared other techniques (i.e. LC-MS), the versatility of this approach and the

possibility to recover the analyzed sample remain essential for studying different structural and compositional aspects of glycosaminoglycan mixtures.

1.3.1. NMR Interaction with Protein

The reactions that occur within the cells constitute the result of highly regulated intermolecular recognition mechanisms. In particular, biological processes such signal transmission and cellular metabolism are mainly regulated by the binding of low molecular weight molecules (the ligands) to proteins. Protein activity can be affected by specific interactions with ligands exhibiting cooperativity through mechanisms involving protein conformational changes. From this perspective, knowledge of the structural, conformational and dynamic features of the molecular recognition processes between proteins and their ligands targets allow to rationalize binding and selectivity features, and therefore to assist in the design of new molecular probes and of novel therapeutic agents[55–57].

1.3.1.1. NMR-based Analysis of Protein-Ligand Interactions

NMR has become a powerful tool to monitor molecular interactions and to deduce features of recognition processes at different levels of complexity. NMR spectroscopy detects and reveals protein–ligand interactions with a large range of affinities (10^{-9}–10^{-3}M). Among the biophysical methods (such X-ray crystallography, Surface Plasmon Resonance, Differential Scanning Fluorimetry and so on), NMR spectroscopy has the advantage to examine the sample in a condition not too far from the physiological one. In this non-rigid environment is possible to investigate the flexibility of the proteins and the conformational changes that occur upon ligand binding. This chapter will briefly discuss some of the most important NMR techniques exploited for protein interaction studies in order to provide basic information for the comprehension of the applications

described in the present thesis.

1.3.2. Protein-Ligand Based Interactions

The study of protein-protein and protein-ligand interactions in solution has recently become possible at an atomic level by new NMR spectroscopy experiments able to identify binding events either by looking at the resonance signals of the ligand or the protein. Ideally, a combination of these two techniques allows a complete picture of the molecular recognition event involving ligand-receptor interaction to be obtained. These methodologies can be applied to the lead generation in drug discovery[55], that correspond to the identification of compounds that demonstrating specific activity against a therapeutic target, as well as to determine the structure-activity relationships (SAR). A protein and a ligand in thermodynamic equilibrium are characterized by the dissociation constant K_d, which for the simplest case of a protein with a single binding site is defined as:

$$K_d = \frac{[P] + [L]}{[PL]}$$

where [P], [L] and [PL] are the equilibrium concentrations of protein, ligand and the complex ligand-protein, respectively. K_d has the units of concentration. Therefore, a value of K_d in the mM range implies an approximately 1:1000 ratio of free to bound states in an equimolar mixture of P and L and a K_d in the µM range implies an approximately 1:1000000 ratio of these states, this last case correspond to a more stable complex, that allow a smaller concentration of free ligand and protein in the system at the equilibrium.

Protein–ligand complexes are dynamic systems and therefore the rate at which the components of the protein–ligand complex exchange between free and bound forms is central to the NMR method.

For a system in slow exchange on the chemical shift timescale (high binding affinity complex and low dissociation constant, K_d= µM or lower), resolved peaks might be expected for the free and bound

states, allowing the integration of separate resolved signals. In practice this is extremely difficult for the complexity of the crowded spectrum. For a system in fast exchange the observed NMR response of a ligand is the mole fraction weighted average of the signals in both-free and bound states:

$$\mathbf{M}_{obs} = \mathbf{M}_{L(free)} X_{L(free)} + \mathbf{M}_{L(bound)} X_{L(bound)}$$

where $\mathbf{M}_{obs}$ is any NMR observable characteristic of the equilibrium system, $X_{L(free)}$ and $X_{L(bound)}$ are the mole fractions of free and bound ligand, and $\mathbf{M}_{L(free)}$ and $\mathbf{M}_{L(bound)}$ represent the NMR properties of the ligand in its free and bound states, respectively.

Similarly, for the observation of the protein,

$$\mathbf{M}_{obs} = \mathbf{M}_{P(free)} X_{P(free)} + \mathbf{M}_{P(bound)} X_{P(bound)}$$

Where X_P and $\mathbf{M}_P$ are now the mole fraction and the NMR properties of the non-bound protein and bound protein.

The mole fractions are defined as,

$$X_{L(free)} + X_{L(bound)} = 1$$

and

$$X_{P(free)} + X_{P(bound)} = 1$$

In bound state the ligand adopts the NMR properties characteristic of the typically much larger receptor (macromolecule), that correspond to a perturbation of the chemical shift of the ligand, in term of position and shape (widening) of the signal. On the other hand, from the receptor's

perspective, the ligand perturbs the binding site microenvironment, generating an alteration of the magnetic properties of the receptor, localized in the Receptor Binding Site of the protein. In either case, the exchange modulates the NMR response of both molecules allowing the detection of the receptor-ligand interaction.

1.3.2.1. NMR-based Analysis of Protein-Ligand Interactions

Saturation transfer difference (STD) NMR spectroscopy is extensively used to obtain structural details of the interaction, thereby enabling the mapping of the ligand binding epitope, or the discrimination of ligands from non-binding molecules. STD results from the difference between ~~of~~ two acquisition experiments. The whole experiment is based on the transfer of saturation from the protein to a bound ligand. This transfer process can be obtained by selectively saturating the protein, in a frequency window, that does not contain ligand resonances. The saturation signal is generated by a sum of Gaussian soft pulses, as shown in Figure 13.

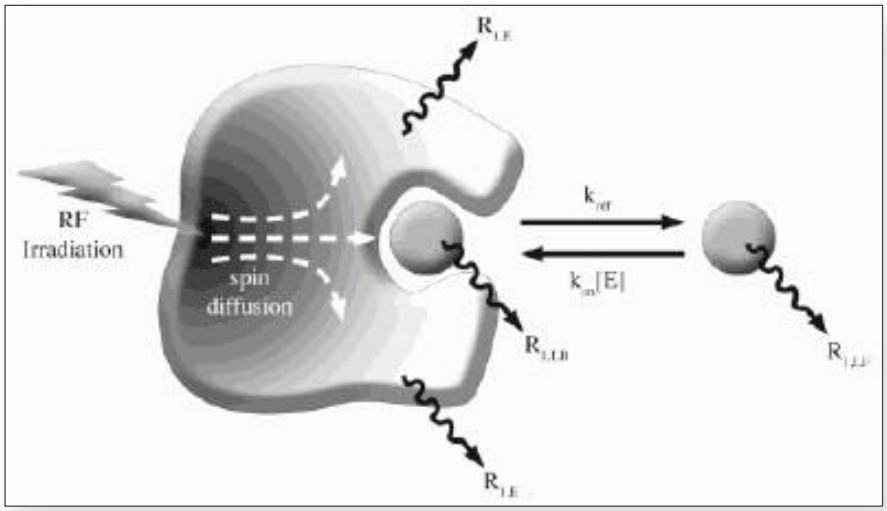


Figure 13. *Schematic view of the STD experiment. The protein protons are selectively saturated at a specific frequency, and the effect of the saturation is transmitted through the whole polypeptide. Any ligand interacting with the protein will also be affected by the saturation. During the irradiation time, the saturation is transferred to the bound ligand.*

The magnetization propagates rapidly through the entire protein mediated by spin diffusion. This process of spin-diffusion is very efficient due to the typically large molecular weight of the protein. Bound ligands acquire this saturation via ^{1}H-^{1}H inter-molecular cross-relaxation at the ligand-protein interface. Protons which are close in space to the protein surface receive the saturation from the receptor. When the ligand dissociates back into free solution, its saturation state persists, while other ligand molecules replace the previous in the receptor binding site, where they will acquire saturation. In this condition the concentration of the saturated ligand raises to enable it to be detected by the acquisition system. Then, only ligand protons that interact with the protein are saturated and show a decreased intensity in the on-resonance spectrum. As it is usually difficult to identify those attenuated signals, Mayer and Meyer suggested the acquisition of an additional off-resonance spectrum with a saturation frequency where no protein or ligand signals are present, usually around ~30 ppm [58]. The final saturation transfer difference (STD) NMR spectrum is obtained by subtracting the on-resonance from the off-resonance spectrum, which results in a spectrum showing only signals of the ligands in contact with the receptor in the lifetime of the complex. [Figure 14].

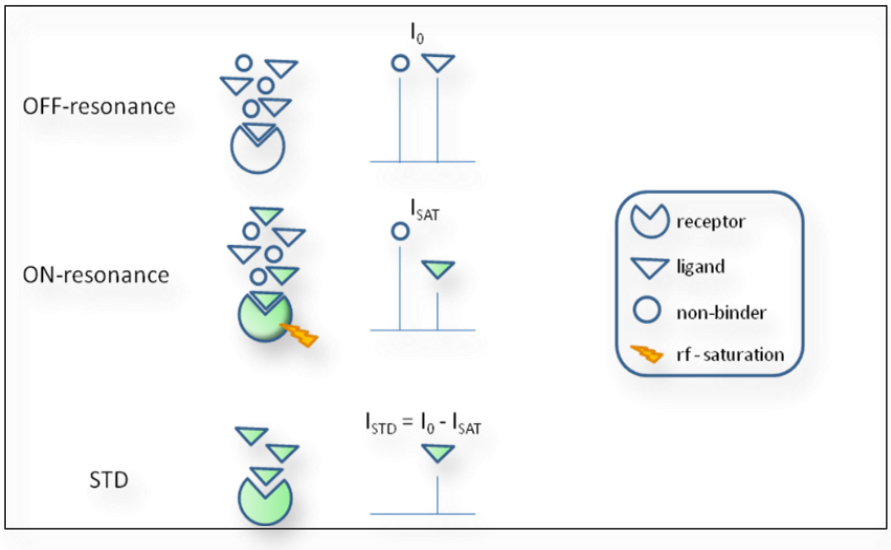


Figure 14. *STD spectrum is generated by subtraction of "on-resonance" spectrum from "off-resonance" spectrum.*

Ligands in solution lose their saturation by normal T_1/T_2 relaxation, which is in the order of one second for small molecules; For larger molecules motions become slower and and more efficient relaxation results (T_1/T_2 shortens) Thus, during the saturation time the number of ligands carrying the magnetization from saturated protein quickly increase into solution, which means that only a small amount of protein is required in this experiment. STD works properly for ligands that have binding affinity (K_d) in the $10^{-2} – 10^{-8}$ M range. High-affinity ligands that undergo slow chemical exchange typically reside longer within the protein-binding site and are thus not detectable by STD NMR spectroscopy. The intensity of the STD signals depends, mainly, on the saturation time and on the excess of ligand molecules in solution, in fact increasing the ligand concentration, or the saturation time, also the STD signal increase. In general, an irradiation time of 2 second and 100-fold excess of ligand give good results. In conclusion STD NMR is an excellent technique both to show ligand-protein binding and to determine the binding epitope of the ligand, overriding information of the drug design[55,59].

1.3.2.2. Transferred NOESY (TrNOE)

Measurements of nuclear Overhauser enhancement (NOE) provide information about distances between hydrogen atoms inside molecules and are indispensable for NMR structure determination of both small and large molecules. The sign and size of the NOE depends on the rotational correlation time and therefore on the size of the molecule.

The observation of trNOEs relies on different tumbling times τc between free and bound molecules. Low or medium-molecular weight molecules (Mw <1000--- 2000 Da) have a short correlation time τc and, therefore, such molecules exhibit positive NOEs, no NOEs, or very small negative NOEs depending on their molecular weight, shape, and the field strength. Large molecules, however, exhibit

strongly negative NOEs. When a small molecule (ligand) is bound to a large-molecular weight protein (receptor) it behaves as a part of the large molecule and adopts the corresponding NOE behaviour, showing strong negative NOEs, so called trNOEs. These trNOEs also reflect the bound conformation of the ligand. Ligand in bound state with a receptor, can thus easily be distinguished by looking at the sign and size of the observed NOEs. Therefore, the maximum enhancement for trNOEs is observed at significantly shorter mixing time τMIX in comparison to the corresponding NOE measured for isolated small molecules in solution. The estimated range of binding affinities that can be probed by trNOESY is $100nM \leq K_D \leq 1mM$[59,60]. In general, one can observe inter and intramolecular trNOEs. Intramolecular trNOEs are the key to define bound-ligand conformations, intermolecular trNOEs occur between a ligand and a receptor protein, and therefore, in principle, allow the determination of the orientation of bound ligands in protein binding pockets.

One of the major drawbacks of this type of experiment is the possible presence of the spin diffusion effect, that are typical for high molecular weight proteins[56,61]. In this case, the magnetization transfer can be mediated by other spins, including receptor nucleus, and by protons closed in space.

Thus, the trNOE method allows fast screening of putative binders compared to a specific target and, at the same time, permits the knowledge of the recognized conformation of the ligand bound to the receptor, with considerable implications for a rational structure-based drug design.

Prediction of theoretical trNOEs, as well as STD, from a 3D complex structure model can be obtained by the application of the CORCEMA (COmplete Relaxation and Conformational Exchange Matrix) approach. Finally, comparing the set of measured trNOEs for a ligand-receptor interaction system, with those simulated on the optimized 3D structure of the ligand-receptor complex by the CORCEMA approach, information of the bound conformation of the ligand can be achieved[62].

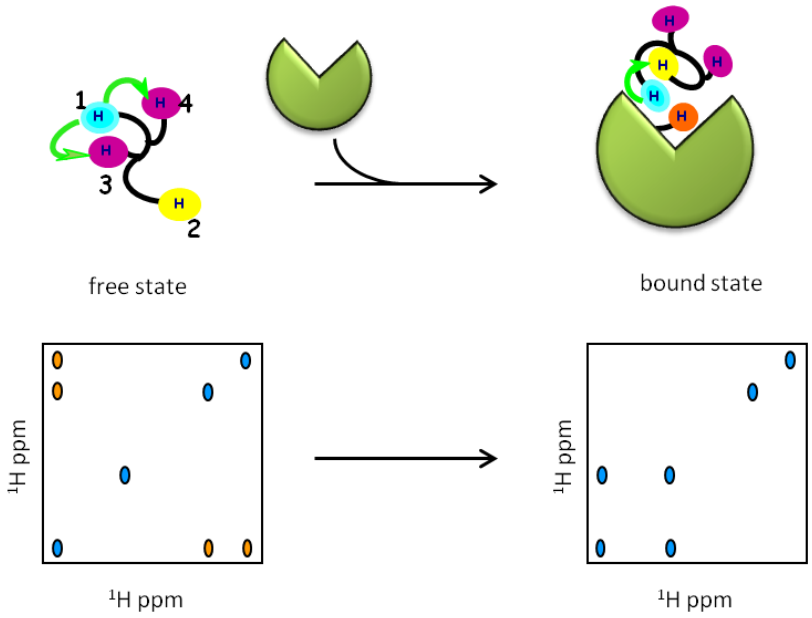


Figure 15. *Left. Schematic representation of a NOESY spectrum for a free ligand. Cross peaks and diagonal peaks have different sign. Right. Schematic representation of a trNOESY spectrum recorded for an exchanging ligand–protein system. Cross peaks and diagonal peaks have the same sign, as expected for a large molecule, thus indicating binding to the protein.*

1.4. A Simplified Introduction to Molecular Modelling

The NMR spectra of a molecule or a macromolecule in solution allow the direct observation of its magnetic properties (chemical shifts, coupling constants, magnetization relaxation, etc.) that depend on their structure, the visited conformations, and the system dynamics; for ligand-receptor or supra-molecular complexes, these properties contain information about the interactions (molecular contacts details) between two or multiple elements, for example the solvent, or a ligand in contact with a receptor. In this description a molecular 'structure' corresponds to the geometry and is completely defined by the set of the atoms, the atom type and the coordinates. Frequently, the molecules/macromolecules are observed to undergo conformational changes, corresponding to geometric changes, obtained without bond breaking, but through rotation of a set of bonds of the system. The dynamic behaviour of the molecule is the whole motion, usually decomposed in the

internal, internal (vibrational, rotational), and external (rotational and translational). Focusing on a single macromolecule/macromolecular complex, the internal coordinate: bond distances, bond angles, dihedral angles are the preferred choice of coordinate to describe the internal motions. The modelling approach allow a "bottom-up" description based on the general idea to reproduce by "numerical experiment" (computer simulation) the chemical and physical behaviour of our system being a molecule in solution or a macromolecule or a supra-molecular complex usually in condition of fixed number of particle (N), absolute temperature (T), and system volume (V) or system pressure (P); from a theoretical point of view, the simulation correspond to the sampling of the canonical ensemble NVT, or NPT of the system.

Computer simulations act as a bridge between microscopic length and time scales toward the macroscopic world of the laboratory: roughly speaking this approach provide to guess of the conformational and dynamic features of the macromolecule, the forces acting between atoms that belong to the same or to different molecules, and compare this model behaviour with the NMR experiments results. In particular, the molecular mechanic and molecular dynamic approach complement the NMR experimental description[63].

Molecular dynamics simulation consists of the numerical, step by step, solution of the classical equations of motion, which for a simple atomic system belonging to a single or several molecules may be written in accord to the Newton second law of classic mechanics

$$m_i \mathbf{r_i} = \mathbf{f_i} \qquad \mathbf{f_i} = -\frac{\vartheta}{\vartheta r_i} U \qquad (1)$$

For this purpose, we need to be able to calculate the forces f_i acting on each atom of mass m_i that is subject to an acceleration r_i. The forces are derived from an empirical potential energy U (r^N) that normally is parameterised for a particular set of molecule or macromolecule or material. The parameters of this potential energy correspond to molecular structural parameter for a particular class

of compound, such as bond equilibrium distances, angles, and torsional angles. There are also other parameters that quantify the ability of these degree of freedom to fluctuate around their equilibrium values, and represent the system energy cost to a deform it.

In an atomistic detail model $r^N = (r_1; r_2;...r_N)$ represents the complete set of 3N atomic coordinates, being N the number of atoms of the model, while in accord to the statistical mechanic the set of 3N velocities are also required to define the complete microstate of the system.

1.4.1. Non-Bond Interactions

Describing macromolecule and/or supramolecular systems two main non bonded interaction, that act between non covalently bonded atoms, are important to be considered: the "Van der Waals" forces and the electrostatic forces. The formers are generated between the external part of the electronic density of two approaching molecules, and are considered as short-range forces: i.e. forces that act for distances smaller than 4-6 Å. Electrostatic interactions are supposed to be applied between partial charges of polar groups, for example (δ+)H--- O(δ---), and also between groups owning finite charges, like ---COO(---) and $–NH_3(+)$ or other ionisable groups. This description requires to introduce a set of "partial charges" that, in the simplest case, correspond to a point-like charge on each atom of the molecule; this charge will describe the electrostatic energy experienced by an external probe charge interacting with that atom. These electrostatic interactions are also used to describe hydrogen bond interactions, in this case the interaction exist if and only if a suitable angle between donor and acceptor groups is maintained.

The Lennard-Jones[64] potential is the most commonly used form to describe empirically the Van der Waals forces:

$$v^{LJ}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad (2)$$

with two parameters that need to be defined for each pair of atom of the system: σ , the diameter, and ε, the well depth.

If electrostatic charges are present, we add the appropriate Coulomb potentials

$$\upsilon^{\text{Coulomb}}(r) = \frac{Q_1 Q_2}{4\pi r' \epsilon_0} \qquad (3)$$

where Q_1, Q_2 are the partial charges of the atom pair, and ϵ_0 is the dielectric constant of the vacuum. The parameters Q_1 and Q_2 could be estimated by a single-molecule quantum-chemical calculation, that allows to estimate the electron density throughout the molecule, which may then be modelled by a distribution of partial charges via Eq. (3), or more accurately by a distribution of electrostatic multipoles.

1.4.2. Bonding Potentials

For molecules we must also consider the intramolecular bonding interactions, an empirical energy function that describe the connectivity between covalently bond atoms, three different potential energy terms are required: the stretching terms, the bending, and the torsional angle terms, that depend from pairs, ternary, and quaternary set of consecutively linked atoms. The simplest molecular model will include terms as reported below:

$$U_{\text{intramolecular}} = \frac{1}{2}\sum_{bonds} k_{ij}^{r} \left(r_{ij} - r_{ij}^{eq}\right)^2 \qquad (4a)$$

$$+ \frac{1}{2}\sum_{bend\ angles} k_{ijk}^{r0} \left(\theta_{ijk} - \theta_{ijk}^{eq}\right)^2 \qquad (4b)$$

$$+ \frac{1}{2}\sum_{torsion\ angles} \sum k_{ijkl}^{\varphi,m} \left(1 + cos\left(m\theta_{ijkl} - \gamma_m\right)\right) \qquad (4c)$$

The "bonds" will typically involve the separation $r_{ij} = |r_i - r_j|$

Between pairs of atoms that are covalently bond in the molecular framework, and we assume in Eq. (4a) an harmonic form with specified equilibrium bond length r_{ij}^{eq} and energy deformation quantified by the "spring constant" k_{ij}^{r}. This functional form allows to describe the bond length degree of freedom as a harmonic oscillator, but in principle more complicated functional form including a higher number of parameters are possible. The "bond angles" θ_{ijk} are defined between successive bond vectors such as $r_i - r_j$ and $r_i - r_k$ and therefore involve three atom coordinates:

$$\cos\theta_{ijk} = \hat{r}_{ij}\,\hat{r}_k = \left(\hat{r}_{ij}\hat{r}_{ij}\right)^{-\frac{1}{2}}\left(\hat{r}_{jk}\hat{r}_{jk}\right)^{-\frac{1}{2}}\left(\hat{r}_{ij}\hat{r}_{jk}\right) \qquad (5)$$

where

$$\hat{r} = \frac{r}{r'}$$

Usually this bending term is taken to be quadratic in the angular displacement from the equilibrium value, as in Eq. (4b), although higher order term functions could be also used to refine the description. Even in that case θ_{ijk}^{eq} and k_{ijk} are the equilibrium bond angle and the corresponding "bending stiffness". The "torsion angles" are defined in terms of three consecutively connected bonds, hence four atomic position vectors:

$$\cos\varphi_{ijkl} = -\hat{n}_{ijk}\hat{n}_{jkl}$$

where $n_{ijk} = r_{ij}\,r_{jk}$, $n_{jkl} = r_{jk}\,r_{kl}$ and $\hat{n} = \frac{n}{n'}$

the unit normal to the plane defined by each pair of bonds. Usually the

torsional potential involves an expansion in periodic functions of the order

m=1,2,…, Eq. (4c).

A simulation package force-field will specify the precise form of Eq. (4) and the non-bond terms to

apply, and the various strength parameters k and other constants therein. We need to keep in mind how Eq. (4) is a considerable oversimplification, some aspects of the intermolecular forces cannot be described by these empirical force fields. Molecular mechanics force-fields, aimed at accurately predicting structures and properties, will include many cross terms (e.g. stretch-bend). Quantum mechanical calculations may give a guide to the "best" molecular force-field; also, comparison of simulation results with thermos-physical properties and vibration frequencies is valuable in force-field development and refinement. A separate family of force fields, such as AMBER[65], CHARMM[66] are oriented to biological macromolecule (proteins, polymers) in condensed phases; their functional forms are similar to Eq. (4), and their parameters are typically optimized for that condition, and determined by quantum chemical calculations combined with thermophysical and phase coexistence data.

1.4.3. Force Calculation

Having specified the potential energy function $U(r^N)$, the next step is to calculate the atomic forces

$$f_i = -\frac{\vartheta}{\vartheta r_i} U(r^N) \qquad (6)$$

For site-site and intra-molecular potentials this is a simple numerical or analytical exercise.

2. Aims of the Work

As already mentioned, design of new HS-like potential drugs requires investigation of oligosaccharides-protein SAR as well as generation of HS analogues having pharmaceutical interest.

This thesis deals with structural characterisation of HS oligosaccharides and conformational properties both of free state and of complexed proteins. We focused our attention on four different subjects:

1. Conformational analysis of chemoenzymatically synthetized hexasaccharides in complex with AT;

2. Conformational analysis of a novel chemoenzymatically synthetized octasaccharide in complex with AT;

3. Structural analysis of Heparan Sulphate mimetics;

4. Application of chemoenzymatic synthesis to obtain Heparan Sulphate-like oligosaccharides.

3. Hexasaccharides

3.1. Introduction

The existing evidence suggest that the electrostatic interactions between the sulfate and carboxylate groups, that periodically "decorate" HS chains, and the basic amino-acids (mainly Arg and Lys residues) at the specific binding site of AT, are the driving force of the Hep/HS-AT recognition process, at least for long distances (greater than 4-5 Å)[67]. Previously, it was observed that the removal of selected sulfate or carboxylate groups from the specific sequence AGA*IA, significantly compromised the binding affinity between the heparin like oligosaccharide and AT. This suggests that the electrostatic forces contribute significantly to molecular recognition, whereas other evidence indicates that the residues type, and the length of the sequence linked to the pentasaccharide at both non-reducing and reducing ends, additionally contribute to the recognition. These observations support the specificity of interaction between the AGA*IA sequence and the Heparin Binding Site HBS. In fact, oligosaccharide sequences (penta- or hexasaccharides) that were previously known to bind and activate AT, always include the specific sequence AGA*IA.[23,68]A recent systematic study conducted on different chemo enzymatically synthetized hexasaccharides which include variants of the AGA*IA sequence, demonstrated that while the sulfation of glucosamine units is essential to produce anticoagulant activity, the sulfation of the iduronic acid residue internal to the AGA*IA sequence does not seems to be critical. This observation arisen by the analysis of a series of AGA*IA containing hexasaccharides obtained by chemoenzymatic synthesis and differing by the degree of sulfation of the AGA*IA sequence[69]. In particular, this study was focused on the correlation between the sulfation pattern of a series of hexasaccharides including the AGA*IA sequence and the

conformation of the key iduronic acid, which was present in its 2-O sulfated and 2-O desulfated forms. [Figure 16 and Table 1].

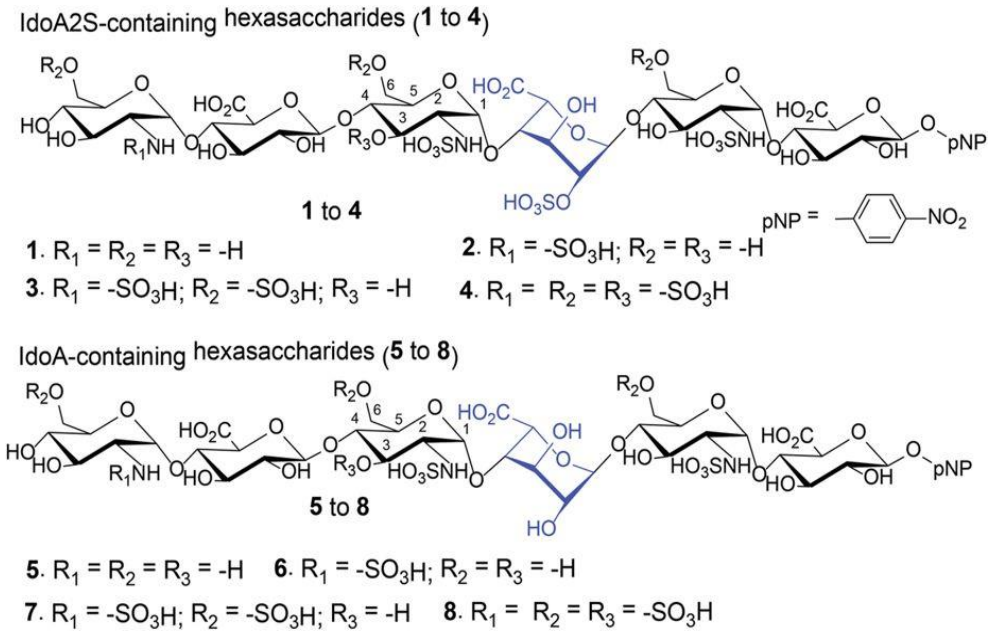


Figure 16. *Structures of hexasaccharides synthesized by Liu group using chemoenzymatic method. Both IdoA2S (in 1 to 4) and IdoA (5 to 8) residues are colored in blue.*

Table 1.

Population of conformers for IdoA2S and IdoA residues in hexasaccharides. In red the conformation of hexasaccharides chosen for our study

Hexasaccharides	**Population Off Conformers**		
	$^1C_4$	**$^2S_0$**	**$^4C_1$**
1-IdoA2S	68%	32%	-
2-IdoA2S	68%	32%	-
3-IdoA2S	47%	53%	-
4-IdoA2S	***25%***	***75%***	***-***
5-IdoA	53%	47%	-
6-IdoA	51%	49%	-
7-IdoA	68%	29%	3%
8-IdoA	***73%***	***19%***	***8%***

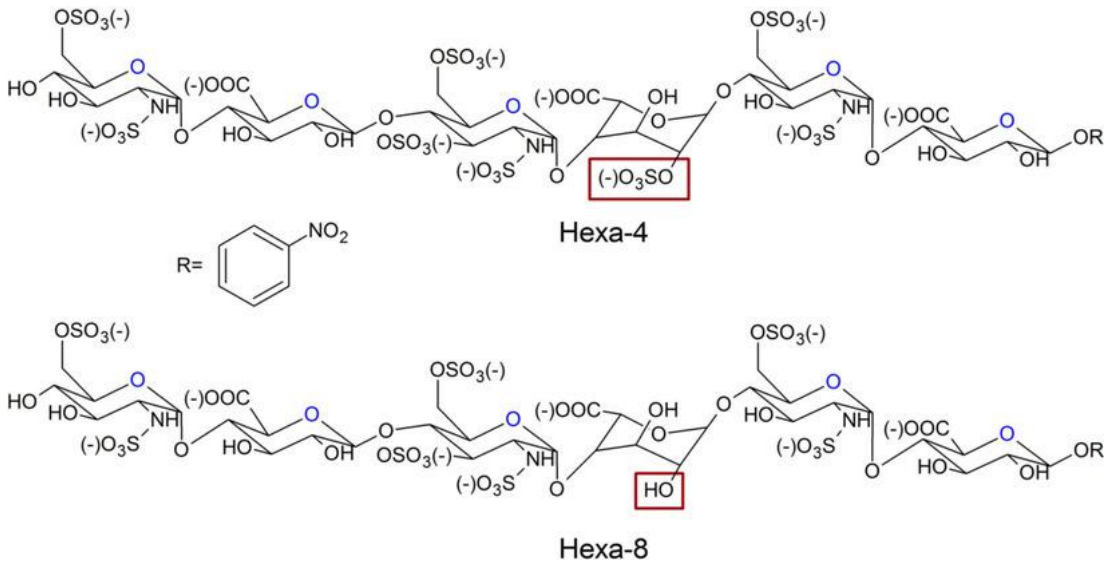


Figure 17. *Chemical structures of hexa-4 and hexa-8. The structural difference of these two compounds is the sulfation of IdoA residues, highlighted by a red box. R=4-nitrophenyl. Residues from non-reducing end: GlcN6S-GlcA-GlcN3S6S-IdoA/IdoA2S-GlcN6S''-GlcA''-R.*

As already shown, the increasing degree of sulfation of adjacent glucosamine residues, in particular the 6-O sulfation of the reducing glucosamine, affects the ring conformation of IdoA2S, shifting it toward the $^{2}S_{0}$ form[15,17,70]. In contrast, the conformation of the IdoA residue is shifted toward the chair $^{1}C_{4}$ when the 2-O-sulfo group is missing[69]. Among the eight synthesized hexasaccharides, only hexa-4 and hexa-8 produce anticoagulant activity [69][Figure 17]. This result was expected for compound 4, that include the typical AGA*IA sequence, but was not predicted for compound 8, in which the 2-O-sulfo group of IdoA is missing[69]. Hexasaccharides 4 and 8 were than selected for the investigation of the structure-activity relationship looking at their 3D structure in free solution and in the presence of AT. In this investigation several complementary approaches were applied: Isothermal Titration Calorimetry (ITC), Nuclear Magnetic Resonance (NMR), and molecular modeling.

3.2. ITC Studies of the heparin-AT interaction (thermodynamic characterization)

The interaction of the hexasaccharides with Antithrombin was investigated by microITC using Fondaparinux as reference model. In order to estimate the dissociation constant of the heparin-AT interaction, titration experiments were performed under the same conditions (see experimental conditions). The titration isotherm curves for the studied heparin-like models are shown in Figure 18. Binding constants and thermodynamic parameters were obtained by fitting the raw titration data [Figure 18 bottom] to a model involving a single binding site for Antithrombin, i.e. applying stoichiometry for the ratio ligand:receptor = 1:1. The K_d values obtained by numerical fitting, are in accordance with those for similar octasaccharides isolated from LMWH Enoxaparin in 2008 but different from those published in previous work for the same oligosaccharides[40]. As a consequence of the higher salt concentration electrostatic charged sites in the system are screened by higher

concentration of free ions (ionic atmosphere), reducing the range of distances in which the electrostatic forces apply, as a consequence decreasing the ligand-receptor affinity. For this reason K_d values measured in this study are greater than those published in a previous work[69].

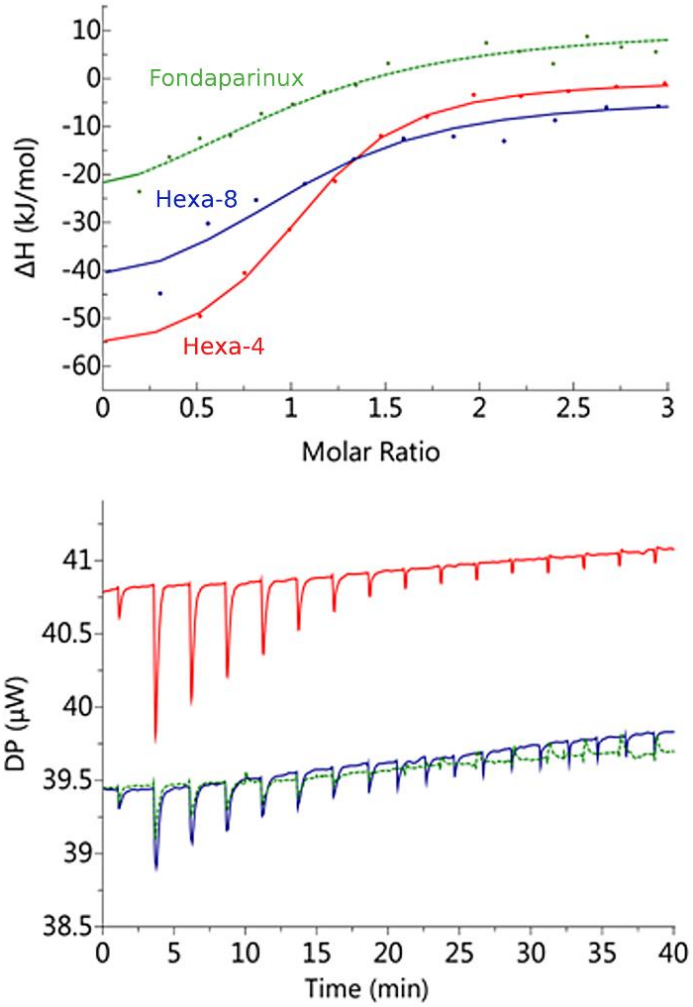


Figure 18. *ITC experimental results. On the bottom panel, the negative peaks resulting from het releases. On the top panel, the binding isotherms created by plotting the integrated heat peaks against the molar ratio of the AT.*

.

Hexa-4 displayed a K_d value about two and six times lower than those of hexa-8 and Fondaparinux, respectively, corresponding to a stronger binding affinity of the former compound. The numerical analysis of the micro-ITC titration curve also provides information about the nature of the interaction of these hexasaccharides with Antithrombin. The thermodynamic parameters in the Table 2 indicate

that the binding reaction is enthalpy driven, while the entropic term is weaker, and usually this correspond to a binding event with a weak conformational change. The structural analysis of the binding pocket describes it as predominantly hydrophilic so the dominant interactions are electrostatic interactions and hydrogen bonds, even if short range Van der Waals forces are still contributing. Hexa-4 shows higher free energy and enthalpy changes upon binding than hexa-8. In fact, the lower enthalpy variation of hexa-8 in comparison with hexa-4 hypothetically correlates both with the missing 2-O-sulfo group, corresponding to an electrostatic interaction lost, and possibly with the higher energy cost that has to be paid to drive a conformational change.

Table 2

Micro ITC thermodynamic parameters

Sample	**Temp. °C**	**Stoichiometry**	**K_d (M)**	**ΔH (KJ/mol)**	**ΔG (KJ/mol)**	**-TΔS (KJ/mol)**
Hexa-4	25.4	1	5.51E-07	-59.5	-35.8	23.7
Hexa-8	25.1	1	1.29E-06	-45	-33.6	11.4
Fondaparinux®	25.1	1	3.52E-06	-45.1	-31.1	13.9

3.3. NMR Characterization

The structure of hexa-4 and hexa-8 was determined in un-bound state, in the same condition used to study the interaction with AT. In this latter case an increased ionic strength of the solvent medium, obtained by addition of salt, modulates the binding affinity of a charged protein due to electrostatic screening. Thus, the off-rate k_{off}, or more specifically the exchange ratio between free and bound, increases passing from nanomolar to micromolar, as required for the analysis of the complex for

reasons relating to NMR timescales. Even if in un-bound state higher ionic strength is not necessary, we preferred to work in the same condition to standardized the characterization for both un-bound and bound oligosacchairdes. Results confirmed the known connectivity previously described in water solution and established the conformational properties of the two ligands in the higher NaCl concentration media used. A full structural and conformational characterization study was set up for these oligosaccharides, following classical NMR approaches and methods. NMR spectra are shown in the experimental section.

3.3.1. ^{1}H STD NMR

STD-NMR is a valid method that can be used as a screening tool for the identification of binding epitopes. Briefly, proton resonances of a protein are selectively saturated at a radiofrequency far from any ligand resonances. During the lifetime of the complex this saturation is transferred to the protons of the bound ligand when the two surfaces ligand-receptor become in contact. If the complex is formed, the resulting proton spectrum, obtained upon subtraction of the spectrum obtained with off-resonance irradiation, will include the resonances of the protons in contact with the protein. The intensity of the STD signal correlates the distance between the ligand proton and the protein surface. As suggested by MicroITC experiments, in which the K_d of the tested glycans were found very similar, suggesting a comparable interaction; the STD experiments showed that both hexasaccharides binds Antithrombin with a similar binding epitope. IdoA/IdoA2S residues showed the most intense STD peaks and was used as references to obtain relative intensities of the other peaks. Outer AGA*IA GlcN residues had comparable intensities to IdoA/IdoA2S central peaks suggesting very close contacts between the AGA*IA sequences and the protein [Figure 19]. We were not able to further confirm this hypothesis with GlcA residues because the anomeric signals are affected by water pre-

saturation, and for this reason H2 and H4 signals are overlapped with other resonances. The GlcA″-(4-nitrophenyl) moiety at the reducing-end is less involved than the AGA*IA sequence in interaction with AT, as shown by the corresponding weaker STD signals, indicating a greater distance between these residues and the HBS. Surprisingly, even if hexa-4 and hexa-8 differ for 2-O sulfation of IdoA, STD signals were comparable for both ligands [Table 3].

Table 3

Selected ^{1}H-STD intensities as percentage relative to the H1 of IdoA2S or IdoA in Hexa-4 or Hexa-8 respectively.

	^{1}H-STD relative intensities (%)							
Ligand	**H1 GlcNS6S**	**H1 GlcNS3S6S**	**H1 IdoA**	**H1 GlcNS6S"**	**H2 GlcNS6S''**	**H1 GlcA"**	**H2 GlcA"**	**H4 GlcA''**
Hexa-4	98	81	100	97	98	43	64	66
Hexa-8	92	92	100	92	89	38	63	56

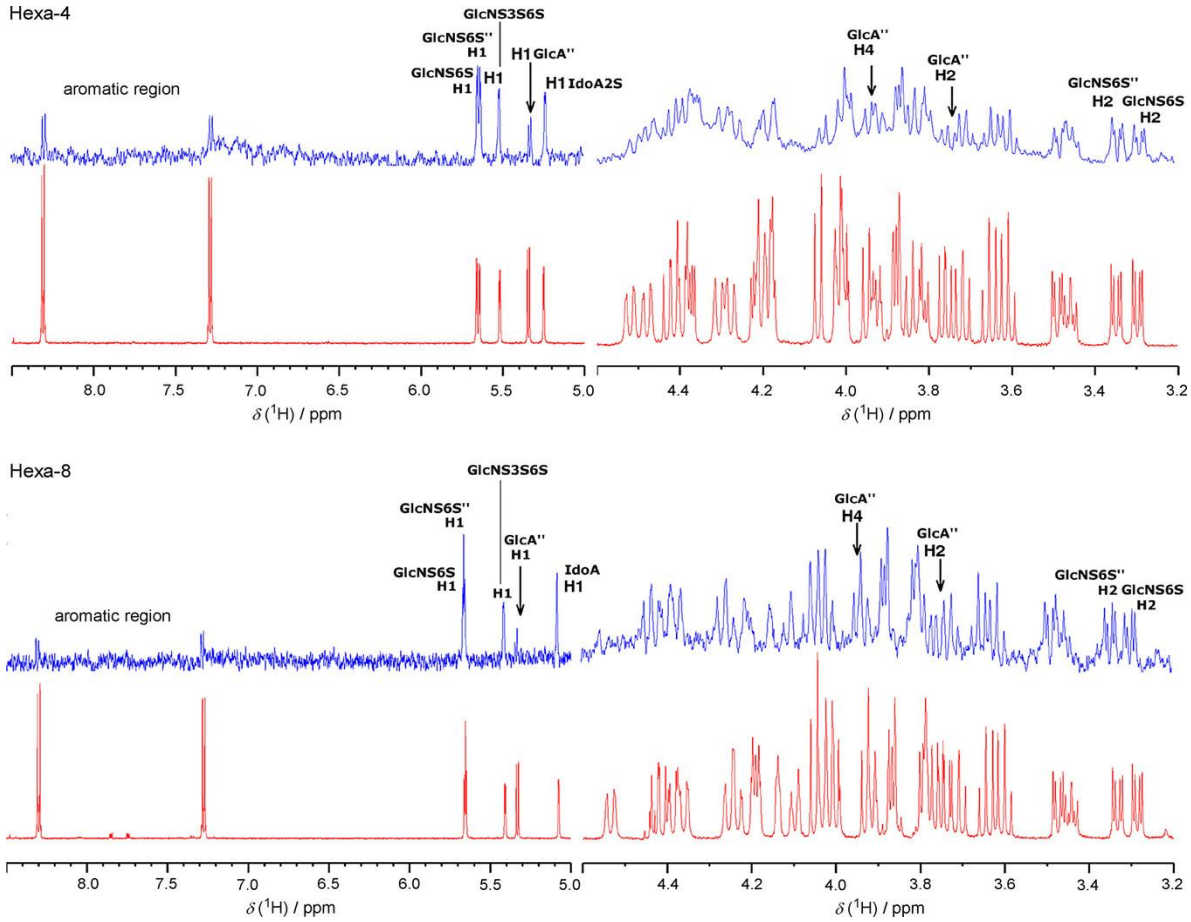


Figure 19. *Partial 600 MHz ¹H-STDNMR spectra of the hexa-4·AT and hexa-8·AT complexes. The experiment was conducted with a glycan/AT molar ratio of 100:1; the temperature was set to 35°C. STD and reference 1H spectra are depicted in blue and red, respectively.*

3.3.2. ^{1}H NMR Chemical Shift Perturbation

High-resolution NMR structure determination of protein/ligand complexes is a time demanding process due to the large amount of data that must be analyzed therefore binding epitope mapping is routinely obtained by 1D-NMR from a comparison of the chemical shifts and the shape of signals of the free versus ligand-protein complex. As a consequence of binding with protein, the correlation time of oligosaccharide increases, and this induces perturbation of chemical shifts, and increased

linewidths. ^{1}H Chemical shift perturbation and signal shape widening induced by the interaction with AT, were observed in the spectra of both hexasaccharides, when compared with ^{1}H spectra of the free glycans. In particular, anomeric signals of GlcNS3S6S and IdoA2S or IdoA presented the highest chemical shift variation, valued as 10 and 5 Hz for hexa-4·AT, and 9 and 7 Hz for hexa-8·AT. On the other hand, disulfated glucosamine residues (GlcNS6S and GlcNS6S″) seem to be slightly less involved in the binding exhibiting weaker chemical shift perturbation than the central GlcNS3S6S-IdoA2S/IdoA moieties. Finally, as expected, the reducing-end glucuronic acid residue GlcA″, as well as the para-nitrophenyl group directly connected, appear not to be involved in the binding, as it was suggested by the unchanged proton chemical shift and signal shape upon binding with AT. This evidence indicates that the closer contacts between hexasaccharides with the HBS involve the pentasaccharide sequence, in particular the GlcNS3S6S-IdoA2S/IdoA moiety showed the strongest interaction, as inferred previously from ^{1}H-STD interaction studies [Figure 20].

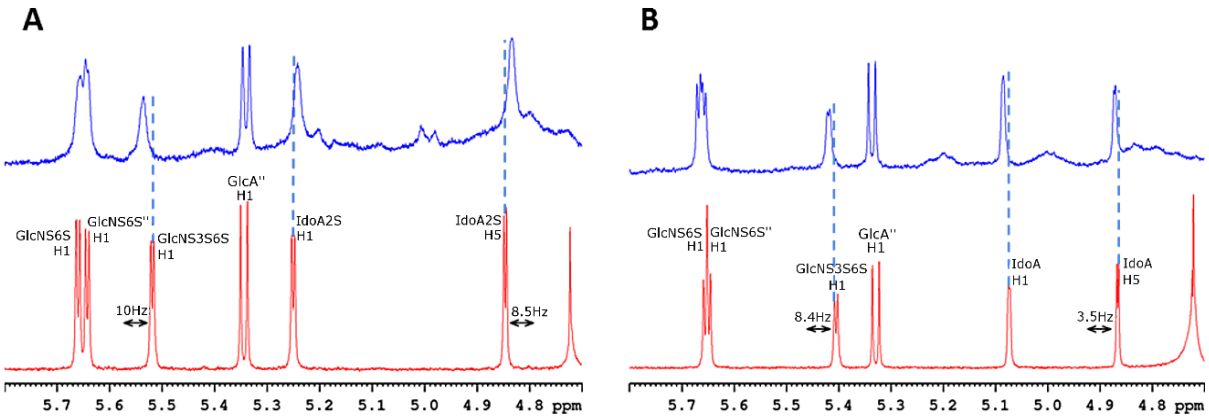


Figure 20. *Anomeric region of the proton spectra of A) hexa-4, and B) hexa-8 in bound (blue line) and unbound (red line) forms at a ligand/protein ratio of 5:1. The anomeric signal of the GlcA residue is hidden by the residual water signal in the spectra of both the hexa-4·AT and the hexa-8·AT complexes.*

3.3.3. Molecular Dynamics

The iduronic acid residue in hexa-4·AT and hexa-8·AT remained in the $^2S_0$ conformation for the whole length of the MD simulation (260 ns approximately), independently of the 2-O-sulfation (see NOESY and tr-NOESY section). MD simulation also allowed sampling of the dihedral angle conformations φ/ψ of the glycosidic linkages associated with hexa-4 and hexa-8 in their free and AT-bound states; these results are reported as Ramachandran plots in Figure 21. In the free state, the glycan backbone was characterized by the average values of φ/ψ calculated from the MD simulation after a relaxation period of 40 ns; in the bound state, the corresponding properties were calculated based on the sets of tr-NOE-selected glycan·AT complex geometries (see tr-NOESY paragraph). The backbone conformations of hexa-4 and hexa-8 in the free state and in the AT-bound state agreed with the AGA*IA as determined by NMR. Similarly to what was observed for a library of heparin-like trisaccharides, both hexasaccharides in the free state showed greater flexibility of the IdoA-GlcNS6S glycosidic linkage than GlcNS3S6S-IdoA (Figure 21 panels A to E)[70]. The conformational freedom of the glycosidic linkages in hexa-4 and hexa-8 showed a consistent reduction from the free to the bound state, particularly evident for the GlcNS3S6S-IdoA2S/IdoA and IdoA2S/IdoA-GlcNS6S″ glycosidic linkages, corresponding to the part of the glycans that better ”adapt” to the AT-HBS. Moreover, significant changes in the glycosidic linkage geometries were observed between GlcA-GlcNS3S6S and GlcNS6S″-GlcA″ in hexa-4 and IdoA-GlcNS6S″ and GlcNS6S″-GlcA″ in hexa-8 (Figure 21, comparison panels B and G, D and I, and E and L). In hexa-4, R13 of AT interacts with sulfate of GlcNS3S6S, 2-*O*-$SO_3^{(-)}$ of IdoA2S and 6-*O*-$SO_3^{(-)}$ of GlcNS6S″. In hexa-8, in which the IdoA sulfate group is missing, the previously cited triple interactions cannot be present, thus strengthening the repulsion between the 6-*O*-sulfate moiety of GlcNS6S″ and the carboxylate group of GlcA″ [Figure 21] and forcing

the last glycosidic bond to assume a torsional state different from that of hexa-4. These results are better represented by the subsets of the glycan·AT structures selected by tr-NOESY analysis, underlined by the yellow symbols in the Ramachandran plots in panels F) to L) of Figure 21. These results suggested that both glycans bind AT from the non-reducing to the reducing end: that is, with involvement of all their pentasaccharide residues constrained in an approximately cylindrical cavity. Only the reducing-end GlcA″-(4-nitrophenyl) moiety is not involved in binding, as shown by the ^{1}H and STD experiments, although it appears to slightly affect the conformation and the position of the pentasaccharide sequence [Figure 22].

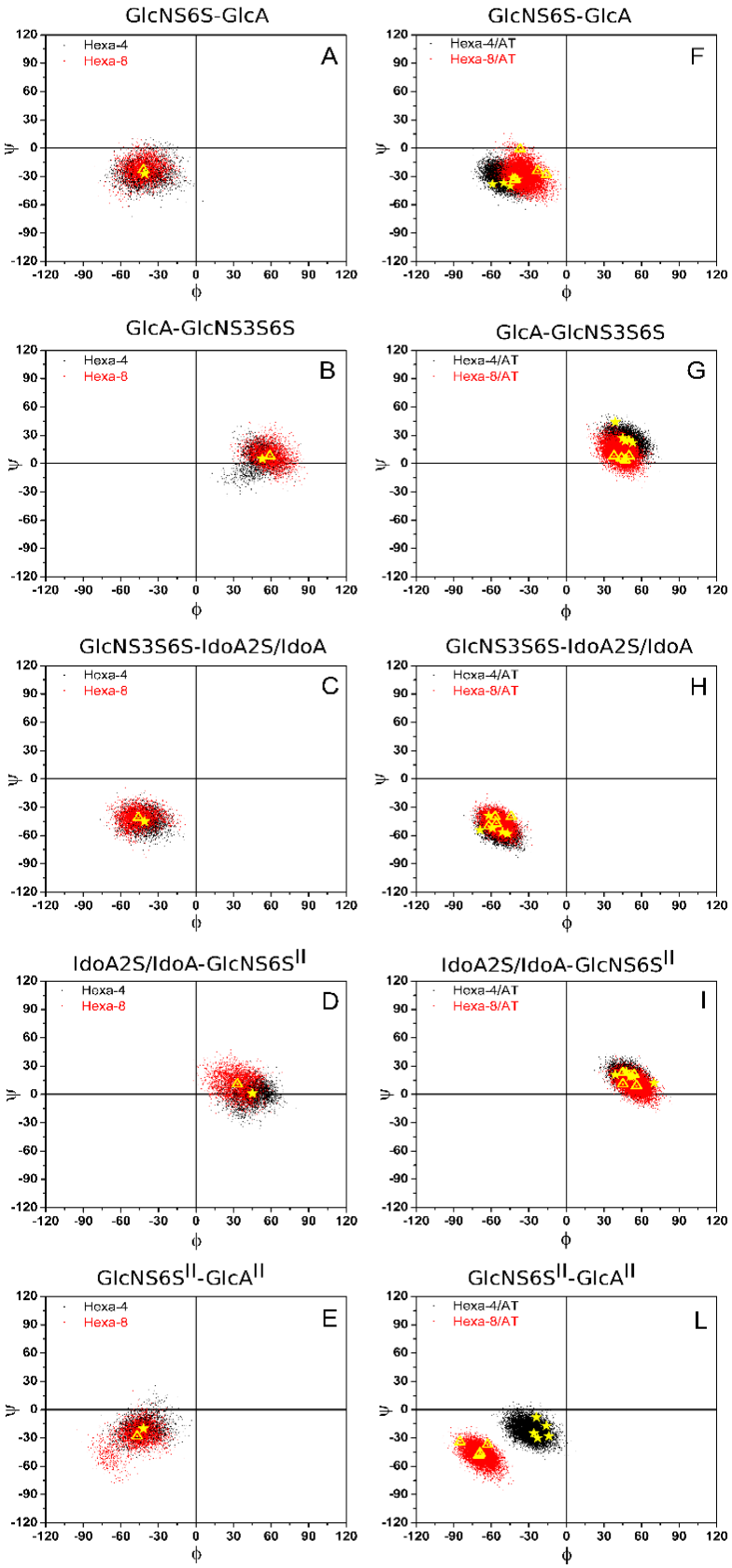


***Figure 21**. Ramachandran plots of the glycosidic dihedral angles for hexa-4 and hexa- 8.*

A)–E) in the free state and F)–L) in the bound state as sampled during MD simulations after a relaxation period of 40 and 80 ns for glycans in the free and the bound state, respectively. In panels A)–E) the yellow stars and triangles correspond to the average values φ/ψ of hexa-4 and hexa-8, respectively, calculated in the free state from the MD simulation trajectories, skipping the first 40 ns. In panels F)–L) the yellow stars and triangles indicate φ/ψ values of the MD-simulated structures validated by tr-NOEs and corresponding to simulation times of 209, 227, 240, 257, 270 ns for hexa-4·AT and 214, 235, 243, 258 ns for hexa-8·AT.

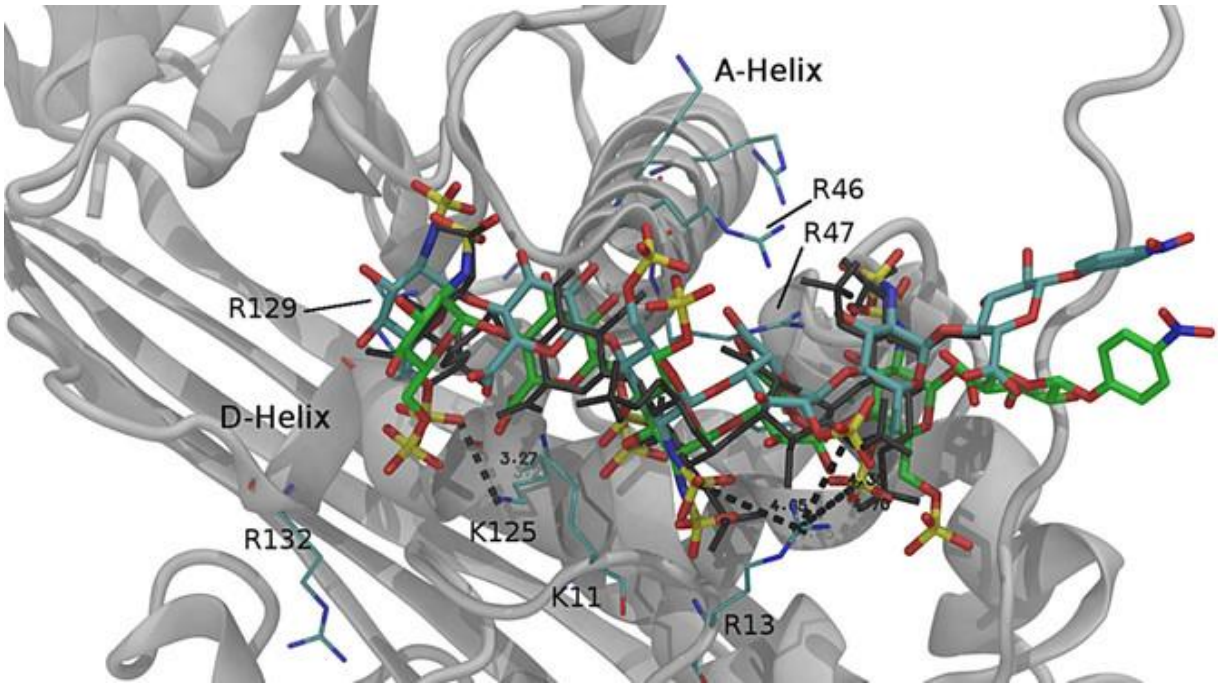


***Figura 22**. Hexa-4·AT and hexa-8·AT complexes selected from the MD simulation trajectories at 227 and 258 ns, respectively, and validated by the interglycosidic tr-NOEs (see text and Table 4). Hexa-4·AT and hexa-8·AT are represented by cyan and green tubes, respectively, whereas the black wireframe represents the reference pentasaccharide as in 1AZX. AT is represented by the grey ribbon of the hexa-4·AT complex only, whereas selected Arg and Lys residues are indicated by thin cyan tubes. The two complexes are superposed to the reference structure 1AZX through the corresponding Ca backbone of the A helix (Ca backbone, between amino acids 45 and 70).*

3.4. NOESY and tr-NOESY

Because the cross relaxation rates in the free and the bound states differ considerably as a consequence of the binding with AT, a variation of cross peak intensities in the two dimensional NOESY (free state) and tr-NOESY (bound state) spectra was observed (Table 4 and Table 5). The conformation of the iduronic acid residue in hexa-4 and in hexa-8 was investigated by analysis of the tr-NOESY spectra [Figure 23]. Similarly, to what was observed for AGA*IA[71], the interaction between hexa-4 and AT changes the intra-residue NOESY ratio H5-H2/H4-H5 of IdoA2S from 0.5 to 1.3 approximately, thus confirming that the conformation of this residue approaches the pure $^{2}S_{0}$ upon binding (Table 6). Surprisingly, the conformation of IdoA of hexa-

8, which adopts an almost pure $^{1}C_4$ conformation in free solution, is also shifted toward $^{2}S_0$ in the bound state to AT, as shown by the change in the H5-H2/H4-H5 NOESY ratio from 0.3 to 0.9 (Figure 23 and Table 6). Comparison between inter-glycosidic NOEs and tr-NOEs, acquired for the glycans in their free and bound states, respectively, allowed the detection of glycan conformational changes upon binding. Pairs of H1-H4 and H1-H3 or H1-H4 and H1-H6 inter-glycosidic NOE and tr-NOE intensities and their ratios are reported for both glycans in Table 5.

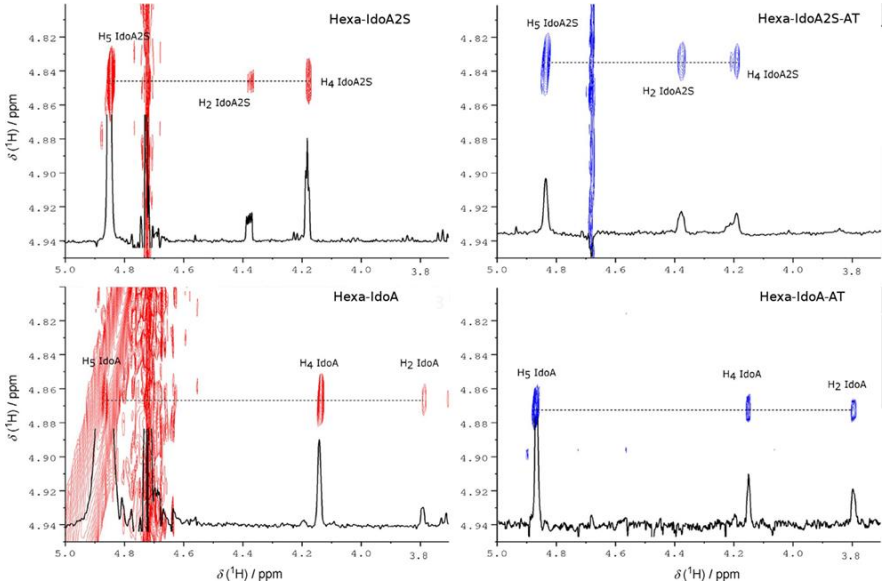


Figure 23. *NOEs (left) and tr-NOEs (right) H5–H2 and H5–H4 Cross-peaks correlations in free and bound state to observe conformational changes upon binding*

Table 4

Percentage of IdoA conformers in free and bound state. Conformation in free state was previously calculated minimizing the residual sum of squares (RSS) between the observed NMR J-values and the total contribution back-calculated from the molecular dynamics (MD).

Sample	H2-H5/H4-H5 ratio	$^{2}S_0$ %	$^{1}C_4$ %
Hexa-4	0.4	75%	25%
Hexa-4+AT	~1	100%	ND
Hexa-8	0.2	19%	80%
Hexa-8+AT	~0.9	100%	ND

Table 5

Selected inter-glycosidic NOEs and tr-NOEs of the glycans in the free and bound states. The ratios between the tr-NOEs and NOEs intensities of the pairs H1-H4 and H1-H3 or H1-H4 and H1-H6 are reported in brackets. Significant differences between ratios of the tr-NOEs and NOEs intensities, indicating glycosidic dihedral angle variation upon binding, are underlined in bold face.

Hexa-4/AT, trNOE/NOE selected intensities						
	GlcNS3S6S-IdoA2S		IdoA2S-GlcNS6S''		GlcNS6S''-GlcA''	
Tmix	H1-H4/H1-H3		H1-H4/H1-H6		H1-H4/H1-H3	
(ms)	Tr-NOE	NOE	Tr-NOE	NOE	Tr-NOE	NOE
200	15/13.9 (1.1)	4.3/4.0 (1.1)	9.8/10.4 (0.9)	3.2/3.4 (0.9)	**16.2/5.7 (2.8)**	**2.3/0.7 (3.3)**
300	25.5/23.4 (1.1)	6.2/6.1 (1.0)	13.3/15.1 (0.9)	4.9/5.2 (0.9)	**25.1/10.1 (2.5)**	**3.7/1.1 (3.4)**
500	34.9/32.3 (1.1)	10.5/10.2 (1.0)	24.1/29.8 (0.8)	8.0/7.8 (1.0)	**43.3/24.4 (1.8)**	**6.0/1.9 (3.2)**
Hexa-8/AT, trNOE/NOE selected intensities						
	GlcNS3S6S-IdoA		IdoA-GlcNS6S''		GlcNS6S''-GlcA''	
Tmix	H1-H4/H1-H3		H1-H4/H1-H6		H1-H4/H1-H3	
(ms)	Tr-NOE	NOE	Tr-NOE	NOE	Tr-NOE	NOE
200	8.6/7.7 (1.1)	5.4/6.7 (0.8)	**9.8/4.8 (2.0)**	**4.0/1.6 (2.5)**	**6.1/1.3 (4.7)**	**1.4/0.1 (14)**
300	17.3/14.1 (1.2)	8.0/10.2 (0.8)	**15.3/6.6 (2.3)**	**6.4/2.2 (2.9)**	**11.0/2.1 (5.2)**	**2.2/0.4 (5.5)**
500	35.7/28.9 (1.2)	13.2/16.7 (0.8)	**32/13.5 (2.4)**	**10.5/3.6 (2.9)**	**24.7/7.7 (3.2)**	**4.0/0.7 (5.7)**

Table 6

Selected H2-H5 and H4-H5 intra-residue NOEs and Tr-NOEs for IdoA in Hexa-4 and Hexa-8 in free and in bound state with AT, respectively. The ratios between the two signals are reported in brackets

	Hexa-4 and Hexa-4•AT		**Hexa-8 and Hexa-8•AT**	
Mixing Time	**NOE**	**Tr-NOE**	**NOE**	**Tr-NOE**
200	2.2/4.9 (=0.45)	19.9/15.6 (=1.28)	0.9/3.4 (=0.26)	5.1/5.7 (=0.89)
300	2.8/7.8 (=0.36)	29.8/24.2 (=1.23)	1.7/5.6 (=0.30)	7.4/9.4 (=0.79)
500	5.2/12.4 (=0.42)	51.8/44.4 (=1.17)	2.2/10.0 (=0.22)	15.5/18.5 (=0.84)

As shown by the corresponding Ramachandran plots discussed in the MD simulation paragraph [Figure 21], significant variation in the glycosidic linkage geometry of reducing-end residues upon binding was observed. Notably, hexa-4 showed changes localized at the GlcNS6S″-GlcA″ glycosidic linkage, whereas hexa-8 showed variation of both IdoA-GlcNS6S″ and GlcNS6S″-GlcA″glycosidic linkages (Table 5). The NOE signals acquired for hexa-4 and hexa-8 glycans allowed us to validate the conformation observed during the MD simulations in the free state. Selected inter-glycosidic NOEs, calculated from optimized glycan conformations, showed partial agreement with the measured NOEs (Table 7).

Table 7

Experimental NOEs of Hexa-4 and Hexa-8 in solution for selected Inter-glycosidic proton-proton correlation. The corresponding simulated NOEs are reported in brackets and are generated using NOEPROM, with an isotropic model of motion (Tc = 700 ps). The Hexa-4 and Hexa-8 structure and conformations are obtained from MD simulation, adjusting each ϕi/ψi glycosidic dihedral angle pair with the average values calculated as reported in materials and methods.

	Hexa-4 Experimental NOE (Simulated NOE)			
Mixing Time (ms)	GlcNS6S-GlcA	GlcNS3S6S-IdoA2S	IdoA2S-GlcNS6S''	GlcNS6S''-GlcA''
200	H1-H4	H1-H4/H1-H3	H1-H4/H1-H6	H1-H4/H1-H3
300	2.9 (2.8)	(4.1) 4.3/4.0 (3.4)	(8.8) 3.2/3.4 (2.5)	(6.1) 2.3/0.7 (0.50)
500	4.5 (5.1)	(6.2) 6.2/6.1 (5.2)	(13.2) 4.9/5.2 (4.0)	(9.2) 3.7/1.1 (0.81)

	Hexa-8 Experimental NOE (Simulated NOE)			
Mixing Time (ms)	GlcNS6S-GlcA	GlcNS3S6S-IdoA	IdoA-GlcNS6S''	GlcNS6S''-GlcA''
200	H1-H4	H1-H4/H1-H3	H1-H4/H1-H6	H1-H4/H1-H3
300	2.1 (6.3)	(3.9) 5.4/6.7 (9.5)	(11.4) 4.0/1.6 (2.1)	(4.3) 1.4/0.1 (0.9)
500	3.5 (9.4)	(5.9) 8.0/10.2 (14.3)	(17.1) 6.4/2.2 (3.5)	(6.4) 2.2/0.4 (1.3)

The glycan conformations in the bound state were determined by selecting structures from MD simulation trajectories after a relaxation period of 80–120 ns and simulating the critical tr-NOEs to be compared with the experimental values. The tr-NOEs were simulated by the CORCEMA program[58] with a "two-state" model for the interaction (see the Experimental Section). The thermodynamic dissociation constants for both glycans with AT were estimated by use of micro ITC. To obtain the best agreement with between the MD simulated and the experimental tr-NOEs, the kinetic k_{off} constants (k_{off} hexa-4: 18 s^{-1}; k_{off} hexa-8: 15 s^{-1}) were estimated, reproducing the relatively stable H1-H2 tr-NOEs of the three glucosamines residues GlcNS6S, GlcNS3S6S and GlcNS6S″ by use of selected glycan·AT structures extracted from the MD simulation (data not reported). Five selected structures of the hexa-4·AT complex, at simulation times of 209, 227, 240, 257 and 270 ns, as well as four selected structures of hexa-8·AT, corresponding to times of 214, 235, 243 and 258 ns, showed good agreement with the experimentally measured tr-NOEs. The *R* factors of these geometries averaged on the selected inter-glycosidic H–H correlations, spanned from 0.10 to 0.22 for hexa-4·AT and between 0.22 and 0.27 for hexa-8·AT. The best agreement was found for the structures at time 227 ns for hexa-4·AT and at 258 ns for hexa-8·AT, as reported in Table 8.

Table 8

Selected inter-glycosidic tr-NOEs measured and calculated from the structures at simulation times 227 ns and 258 ns for hexa-4 and hexa-8, respectively. Bold numbers in brackets correspond to the tr-NOEs calculated by use of the CORCEMA program. The R factor is underlined in bold italic.

Tmix ms	GlcNS6S-GlcA	GlcNS3S6S-IdoA2S		IdoA2S- GlcNS6S‘‘		GlcNS6S‘‘-GlcA‘‘	
	H1-H4	H1-H4	H1-H3	H1-H4	H1-H6	H1-H4	H1-H3
			Hexa-4·AT $_t$=227 ns				
200	14.9 **(10.5)**	15 **(5.3)**	13.9 **(17.3)**	9.8 **(10.0)**	10.4 **(9.2)**	16.2 **(20.3)**	5.7 **(5.1)**
300	22 **(15.5)**	25.5 **(8.6)**	23.4 **(25.1)**	13.3 **(14.5)**	15.1 **(13.5)**	25.1 **(28.8)**	10.1 **(8.2)**
500	41.5 **(26.2)**	34.9 **(16.9)**	32.3 **(41.0)**	24.1 **(23.6)**	29.8 **(22.0)**	43.3 **(44.7)**	24.4 **(16.0)**
R	***0.12***	***0.34***	***0.05***	***0.00***	***0.05***	***0.01***	***0.10***
			Hexa-8·AT $_t$=258 ns				
	GlcNS6S-GlcA	GlcNS3S6S-IdoA		IdoA- GlcNS6S‘‘		GlcNS6S‘‘-GlcA‘‘	
	H1-H4	H1-H4	H1-H3	H1-H4	H1-H4	H1-H4	H1-H3
200	5.9 **(11.3)**	8.6 **(5.3)**	7.7 **(15.9)**	9.8 **(7.3)**	4.8 **(8.1)**	6.1 **(2.9)**	1.3 **(2.7)**
300	9.8 **(16.0)**	17.3 **(8.1)**	14.1 **(22.4)**	15.3 **(10.5)**	6.6 **(11.4)**	11.0 **(4.6)**	2.1 **(4.3)**
500	21.1 **(25.5)**	35.7 **(14.9)**	28.9 **(35.1)**	32 **(17.1)**	13.5 **(18.1)**	24.7 **(9.2)**	7.7 (**8.5**)
R	***0.15***	***0.32***	***0.16***	***0.19***	***0.22***	***0.38***	***0.11***

To compare the relative positions of hexa-4 and hexa-8 in the AT HBS graphically, the corresponding helix A of the tr-NOE-validated complexes and that of the reference complex (AGA*IA·AT; PDB ID: 1AZX) were superposed (Cα backbone, between amino acids 45 and 70 in 1AZX numbering) and are reported in Figure 22. These operations were repeated for the whole sets of hexasaccharide·AT validated structures, and the average root mean square distance (RMSD) of each glycan residue and the corresponding residue of the reference complex was calculated (Table 9). The AGA*IA sequence in both hexa-4·AT and hexa-8·AT, qualitatively matches the corresponding glycan in AGA*IA·AT, even if this similarity improves toward the reducing end (IdoA2S, GlcNS6S″) for hexa-4 and in a more extended region including GlcA and GlcNS3S6S for hexa-8. This result was unexpected because hexa-4 has the AGA*IA sequence closer to that of the active pentasaccharide

than hexa-8; this possibly indicates that the reducing GlcA″-(4-nitrophenyl) terminal end of these hexasaccharides might affect the binding of the AGA*IA sequence to AT.

Table 9

*Average values of the RMSDs calculated between residue pairs of the common sequence (AGA*IA) with regard to hexa-4·AT, hexa-8·AT and AGA*IA·AT after superposition of the A helix (see the Experimental Section.)*

	Average RMSD (C1, C2, C3, C4, C5, O5) [Å]					
	A helix	**GlcNS6S**	**GlcA**	**GlcNS3S6S**	**IdoA2S/IdoA**	**GlcNS6S″**
hexa-4·AT	0.46	1.84	2.15	1.98	1.23	0.66
hexa-8·AT	0.67	1.81	1.30	1.08	1.40	1.37

Distances between selected interacting groups of the hexa-4·AT and hexa-8·AT models (lowest *R* score) and the corresponding AGA*IA·AT, determined by NMR methods[71] and by X-ray diffraction (″1AZX“, ”3KCG“ and ”2GD4“), are reported in Table 10, In hexa-4 and hexa-8, the GlcNS3S6S, IdoA2S/IdoA and GlcNS6S″ residues are in closer proximity to AT than GlcNS6S and GlcA″, as shown in Figure 24 and Table 10. Independently of the degree of sulfation of the iduronic acid moiety, the interatomic distances in the regions surrounding GlcNS3S6S-Ido2S/IdoA of hexa-4 and hexa-8 in bound state with AT are similar to those found in NMR and crystal-state models (Table 10). K114 was confirmed as one of the key residues, being in contact with GlcNS3S6S (NS, 3S), IdoA2S/IdoA (COO−) and, to a lesser extent, GlcNS6S″(6S), as observed for the AGA*IA specific pentasaccharide·AT complexes.

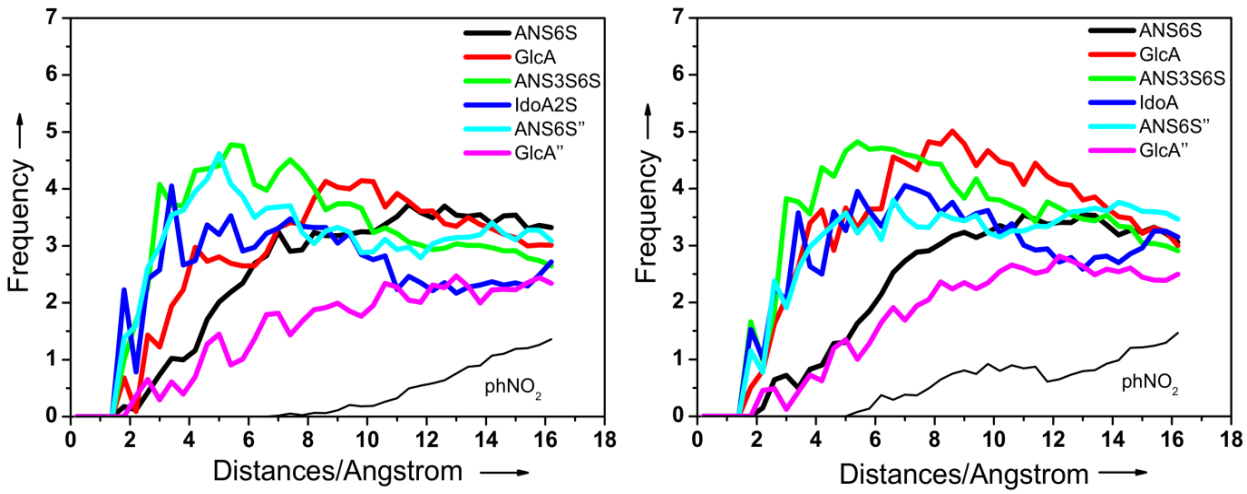


Figura 24. *Diagram of the contacts (histograms of the distances) between each residue and the hydrogen atoms belonging to AT. The diagrams were calculated by using a subset of the tr-NOE-validated structures for A) hexa-4·AT, and B) hexa-8·AT complexes, that correspond to MD simulation times of 209, 227, 257 and 270 ns and of 214, 235, 243 and 258 ns, respectively.*

Additionally, R13 in hexa-4·AT was near to GlcNS3S6S (NS) and IdoA2S (2S; 4.7 and 4.4 Å, respectively). This finding does not exclude partial involvement of the 2-*O*-sulfo group of IdoA2S in interaction with AT, as supported by AT X-ray structures in ”Latent“, ”Intermediate“, and ”Activated“ conformations (PDB IDs: 1E03, 1NQ9, and 3KCG), but in partial disagreement with what was previously discussed and with the X-ray-resolved structure of AGA*IA·AT in PDB ID 2GD4.

Table 10

*Selected distances between interacting groups in Hexa-4·AT and Hexa-8·AT complexes at time 227 and 258 ns validated by tr-NOEs signals (see tr-NOEs/NOEs paragraph) are reported in 2nd and 4th column. The atoms used for distance definitions are underlined in bold face in the 1st column. For comparison the corresponding distances measured in AGA*IA·AT complex in solution by NMR techniques (Hricovini et al. 2001) [1], by x-ray diffraction '1AZX' [2] (chain I Jin et al. 1997), '3kcg' [3] (Johnson, D.J. et al. 2010), and '2gd4' (Johnson, D.J. et al. 2006) are reported in 3rd, 5th , 6th and 7th column respectively. Distances greater than 6 Å and smaller than 4 Å are underlined in blue and red respectively. The symbol '-' indicate distances not reported. In 3kcg structure R46 residue were not accurately resolved.*

Interacting groups	**Distances (Å)**					
	Hexa-4·AT	AGA*IA·AT [1]	Hexa-8·AT	AGA*IA·AT [2]	AGA*IA·AT [3]	AGA*IA·AT [4]
GlcNS6S(6S)---R129(H2N-C-NH_2^+)	6.4	4.3	8.3	4.9	5.2	5.0
GlcNS6S(6S)---K125(H3N$^+$)	4.2	3.8	5.7	4.4	4.8	6.2
GlcA(COO)---K125	7.2	3.5	4.4	4.1	4.1	3.7
GlcNS3S6S(NS)---K114	4.3	-	4.6	4.3	4.1	4.5
GlcNS3S6S(3S)---K114	4.4	-	4.2	4.1	4.1	4.4
GlcNS3S6S(NS)---R13	4.7	-	6.7	5.4	5.6	8.2
IdoA2S(COO)---R47	4.4	4.5	4.1	5.6	4.5	5.1
IdoA2S(COO)---R46	4.2	-	4.0	5.7	-	8.1
IdoA2S(COO)---K114	3.5	3.0	3.6	3.6	3.7	3.4
IdoA2S(SO3-)---R13	4.4	-	/	8.7	4.1	8.6
GlcNS6S''(NS)---R47	4.0	3.9	4.1	4.2	4.7	4.9
GlcNS6S''(NS)---R46	5.1	4.2	5.5	5.9	-	5.9
GlcNS6S''(6S)---K114	7.8	4.7	7.6	5.2	5.2	4.9
GlcNS6S''(6S)---R13	4.7	-	4.6	8.4	5.3	5.4

The tr-NOE-validated complex structures allowed to determine refined details of the glycan·AT complexes, such as the histograms of the distances between the atoms of each of the glycan residues and the hydrogen atoms belonging to AT [Figure 24]. The histograms show the frequencies of population for the distances between the indicated atom pairs, as determined in the tr-NOEs of selected complexes. As previously observed in terms of chemical shift perturbation (Figure 20) and of the glycosidic linkage φ/ψ distribution analysis [Figure 21], in both hexasaccharides the four central residues GlcA, GlcNS3S6S, IdoA2S/IdoA and GlcNS6S″ have closer contacts with the AT HBS than the reducing- and non-reducing-end GlcA″ and GlcNS6S units. Notably, the histogram clearly shows that the 4-nitrophenyl group is further away from the AT HBS than the GlcA″ residue.

The hexa-4·AT complex shows slightly closer contacts between the GlcNS3S6S and IdoA2S residues and AT than in the case of hexa-8·AT, as underlined by the slightly greater population at distances smaller than 4 Å observed for hexa-4 (green and blue lines in Figure 24).

3.5. Conclusions

Chemoenzymatic synthesis has allowed the preparation of a wide range of high-purity artificial GAG oligosaccharides reproducing both natural and other possible sequences occurring in heparan sulfate and heparin. This provides the unique opportunity to obtain variants of AT binding oligosaccharides, the compositions of which can be controlled in terms of position and modification of the AGA*IA sequence[50].Recently, it was demonstrated that two AGA*IA containing hexasaccharides, terminated at the reducing end by the remnant GlcA-R (R=4 nitrophenyl) and differing in the sulfation at the 2-O position of the iduronic acid residue, have comparable abilities to bind and activate AT[69]. Here the interaction of these two hexasaccharides with AT was studied in comparison of results with those previously obtained by X-ray crystallography and NMR techniques, characterized by complexes between AGA*IA containing oligosaccharides and AT. The AGA*IA specific binding is dominated, at least for long range distances, by electrostatic interactions involving negatively charged groups (sulfate and carboxylate) of the glycan and the positively charged residues (mainly Arg and Lys) that characterize the HBS of AT across the helix D subunit[35,37,72]. Convergent studies by several groups[35,36,72,73] based on structural and kinetic investigations, described the AGA*IA·AT binding as a two steps process. In the first step the rigid trisaccharide AGA* approaches the HBS of AT, driven mainly by Lys125 and Lys114, to form a low affinity complex. In the second step, the HBS of AT undergoes a conformational change that allows the positioning of Lys114 and Arg129 in proximity to GlcNS3S6S, thus driving the conformation of

IdoA2S toward the $^{2}S_{0}$ form and allowing the subsequent binding of the reducing-end IdoA2S-GlcNS6S'' disaccharide motif [36].

For both hexasaccharides, STD NMR experiments and MD simulations indicated a glycan binding epitope extending from the non-reducing end GlcNS6S to the reducing end GlcNS6S''-GlcA'' residues. In particular, the two central GlcNS3S6S and IdoA2S/IdoA residues showed the stronger interaction with AT for both glycans, whereas the reducing end GlcA'' appeared less involved in binding with AT. MD simulations with use of explicit solvent were employed to characterize the conformational properties of hexa-4 and hexa-8 in their free and AT-bound states, spanning approximately 100 and 260 ns, respectively. IdoA2S and IdoA remained in the $^{2}S_{0}$ conformation during the MD simulations in the cases both of hexa-4·AT and of hexa-8·AT, independently from the 2-O sulfation state, thus supporting the idea that the HBS surface enforces the $^{1}C_{4}$ to $^{2}S_{0}$ transition upon binding[71]. All glycosidic linkages of both glycans experience decreased conformational freedom on going from the free to the bound state, but this was particularly true for the central GlcNS3S6S-IdoA2S/IdoA-GlcNS6S'' residues, indicating their closer contacts with the HBS, as inferred previously from ^{1}H and ^{1}H STD interaction studies. These results are reinforced by considering the glycosidic dihedral angles (yellow symbols in Figure 21) of the subsets of hexa-4·AT and hexa-8·AT geometries in agreement with tr-NOE ratios, and additionally by the histogram contacts between the glycan residues and the HBS of AT. For residues at the reducing end, the glycosidic linkage conformations change upon binding. This was more evident in hexa-8, in which the missing 2-O sulfation of IdoA weakly affects the nearest neighbor IdoA-GlcNS6S'' linkage, but significantly perturbing the next GlcNS6S''-GlcA'' (Figure 21). This could be explained in terms of a possible contribution by the reducing GlcA'' residue to allow better adjustment of the pentasaccharide in the HBS of AT; it also agrees

with the higher affinity of hexa-8 in relation to the pentasaccharide containing the 2-O desulfated IdoA residue[74]. Intra residue tr-NOE ratio analysis clearly illustrates that both IdoA2S and IdoA move towards the pure $^{2}S_{0}$ conformation upon binding. The tr-NOE validated geometries of hexa-4·AT and hexa-8·AT showed comparable positioning of the AGA*IA sequence once superposed through the A-helix on the AGA*IA·AT (1AZX). In this description the analysis of the glycan·AT contacts indicated that GlcNS3S6S, IdoA2S/IdoA and, to a lesser extent, the reducing end GlcNS6S'' preserve the contacts previously observed in the reference NMR and X-ray diffraction AGA*IA·AT complexes (Table 10). These results also confirm the significant role of the contact network between AT HBS and residues surrounding the trisaccharide sequence GlcNS3S6S-IdoA2S/IdoA-GlcNS6S'' in changing the IdoA conformation from $^{1}C_{4}$ to $^{2}S_{0}$ upon binding, independently from the 2-O sulfation state of IdoA.

4. Octasaccharide

4.1. Introduction

Structural variability of the AGA*IA sequence was considered several years ago, after the isolation and characterization of different oligosaccharides obtained by enzymatic depolymerization of heparins from different animal sources. The currently known AT-binding sequence comprises a disaccharide -GlcA-GlcNS3S ± 6S- unit. The substitution of the -GlcA- in the above-mentioned disaccharide unit with the -IdoA2S was perceived to abolish its AT-binding property.

Herein we want to investigate about the conformational properties of a novel chemoenzimatically synthesized octasaccharide recently described by Professor Liu at University of North Carolina. This glycan belongs to a series of oligosaccharides designed to reproduce possible HS structure variability, able to bind and activate AT [Figure 25]. Characteristic of these oligosaccharides is that their biological activity (anti-FXa) depends by both the position of 3-O-sulfated glucosamine along HS chain and the structure of surrounded residues. Particularly, the octasaccharide 5 [Figure 26], even though it contains the -IdoA2S-GlcNS3S6S-IdoA2S- sequence instead of the typical -GlcA-GlcNS3S6S-IdoA2S- sequence, shows affinity to AT and exerts anti-FXa activity similarly to Fondaparinux (Table 10).

We have therefore focused on the 3D-structure characterization of this octasaccharide in solution and in presence of AT, comparing its conformation to that of Fondaparinux and paying particular attention to the conformation of the two key iduronic acid residues positioned before and after the GlcNS,3,6S residue.

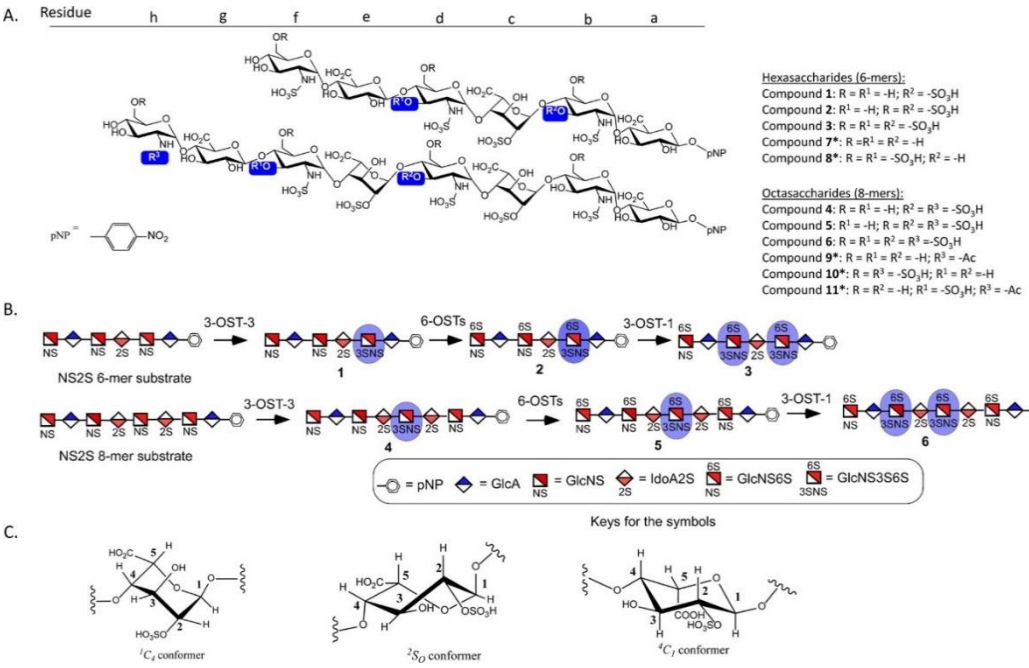


Figure 25. Series of chemoenzimatically synthesized oligosaccharide to investigate the structure-activity relationship. Residues of oligosaccharides are labelled with letter from a to h starting from the reducing-end.

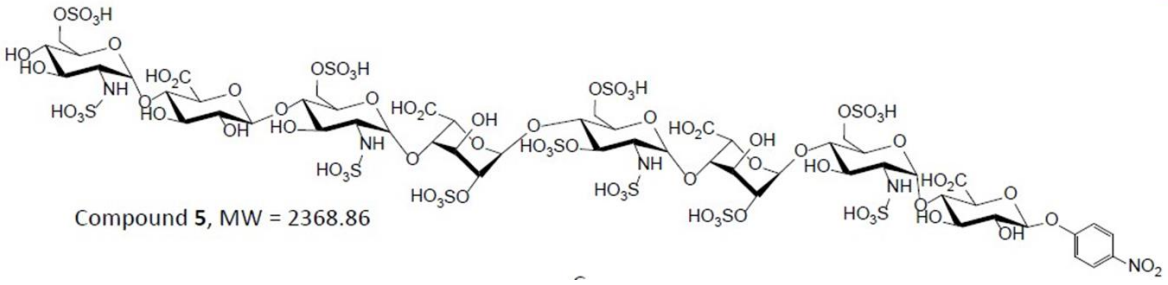


Figure 26. Structure and molecular weight of octasaccharide 5.

Table 10

Correlation between the saccharide sequence and AT-Binding Constant (K_d) as well as anti-FXa Activity (IC_{50})

Compound	Abbreviated Structure	Binding Affinity (K_d) to AT (nM)	Anti FXa activity (IC_{50}, nM)
Octa-5	GlcNS6S-GlcA-GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S-GlcA-pNP	5.1±1.4 nM	7.7 nM
Fondaparinux®	GlcNS6S-GlcA-GlcNS3S6S-IdoA2S-GlcNS6S	14.8±1.4 nM	12.2 nM

4.2. Temperature coefficient and hydrogen bond strength determination

The temperature dependence of sulfamate proton resonance chemical shifts of the central GlcN3S6S, and the line widths provide information regarding the strength of the intra-residue hydrogen bond between SO3(-)-NH--- and ---O-SO2(-)-O3- equatorial substituents, that correlate with the kinetics of the exchange of the sulfamate protons. Important differences in exchange behavior may be related to the involvement of the GlcNS NH in a persistent hydrogen bond, a possibility already observed in oligosaccharides that include GlcNS3S6S residue. The temperature dependence of the Octa-5 GlcNS NH resonances can be ascertained by the visual inspection of the spectra in Figure 27. The frequencies of the sulfamate protons of the two GlcNS residues “b” and “h” at the reducing and non-reducing-ends of the octasaccharide shift toward the water resonance at similar rates (−7.3 ppb/K) and sharpen as the temperature is lowered, while the central GlcNS3S6S residue “d” resonance undergoes only a slight change in chemical shift and remains sharp even at 30°C. The lower temperature coefficient determined for the central GlcNS3S6S residue “d” (−0.7 ppb/K) provides evidence for a stronger hydrogen bond, in comparison to the corresponding observed for GlcN6Ss “h” and “b”; in fact, the analysis of the 3D structure of octa-5, suggest how the former hydrogen bond correspond to an intra-residue, and is typical of the GlcNS3S6S residue.

The hydrogen bonds formed by “h” and “b” residues are formed across the glycosidic bond, involving the groups: SO3(-)-NH---O3-, of GlcNS6S and the nearby GlcA; these reasons suggest how the latter hydrogen bond is weaker compared to the former case.

Table 12.

GlcNS NH temperature coefficients measured in 90% H2O/10% D2O 10 mM phosphate buffer. Temperature range from 273 °K to 303°K

Residue	δ ppm 273°K	δ ppm 280°K	δ ppm 287°K	δ ppm 295°K	δ ppm 303°K	Temperature coefficient (ppb/K)
GlcNS6S “b” and “h”	5.63	5.71	5.77	5.81	5.85	-7.3
GlcNS6S3S “d”	5.38	5.39	5.40	5.40	5.40	-0.7

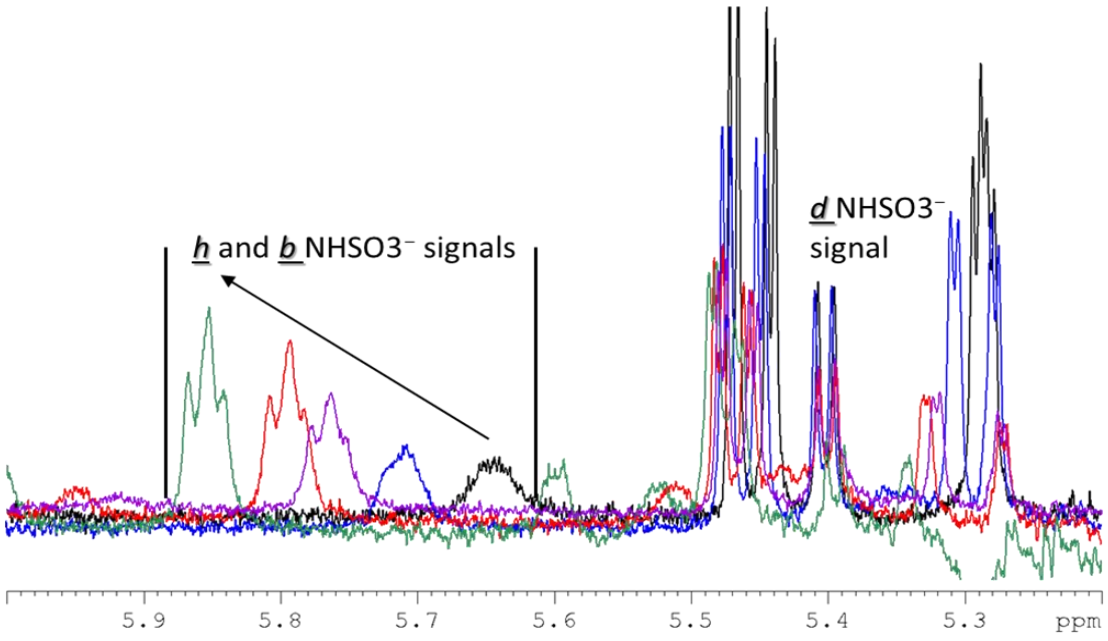


Figure 27. *Selected ^{1}H NMR spectra measured for Octa-5 as a function of temperature. The three NH resonances are labeled according to the positions of the GlcNS residues in the octasaccharide structure.*

In Figure 28 are shown the distances and the orientation of the groups involved in the abovementioned hydrogen bonds. The distance between SO_3^--NH---O3- in the intra-residue hydrogen bond of GlcNS3S6S (residue "d") is constant (lower than 2 Å), because the ring geometry is less affected by rotation of bonds and the conformation does not change dramatically. Moreover, the two interacting groups are oriented frontally increasing the chance to maintain a longer contact time. On the other hand, owing the high flexibility of inter-glycosidic linkages and the greater distance, the possibility to keep in contact the two interacting groups are lower reducing the strength of hydrogen bond.

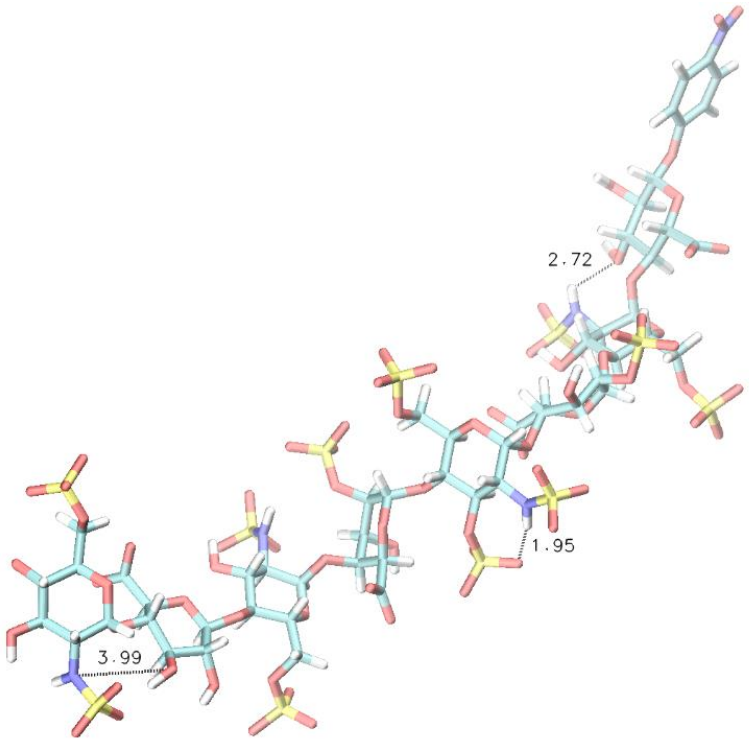


Figure 28. *Selected dynamic simulation structure in which are highlighted the difference in length between the hydrogen bonds.*

To better understand the role of sulfamate group hydrogen bonding in the Octa-5 and reinforce our experimental results, a preliminary molecular dynamics simulation of the full octasaccharide was performed sing NAMD 2.12 with the Glycam-06 force field. The analysis of trajectories involving the distances between GlcNS3S6S NH and the neighbor 3-O-sulfo group [Figure 29] confirm these data suggesting that during the whole simulation time NH proton interacts at least with one of the oxygen atoms belonging to sulfo-group.

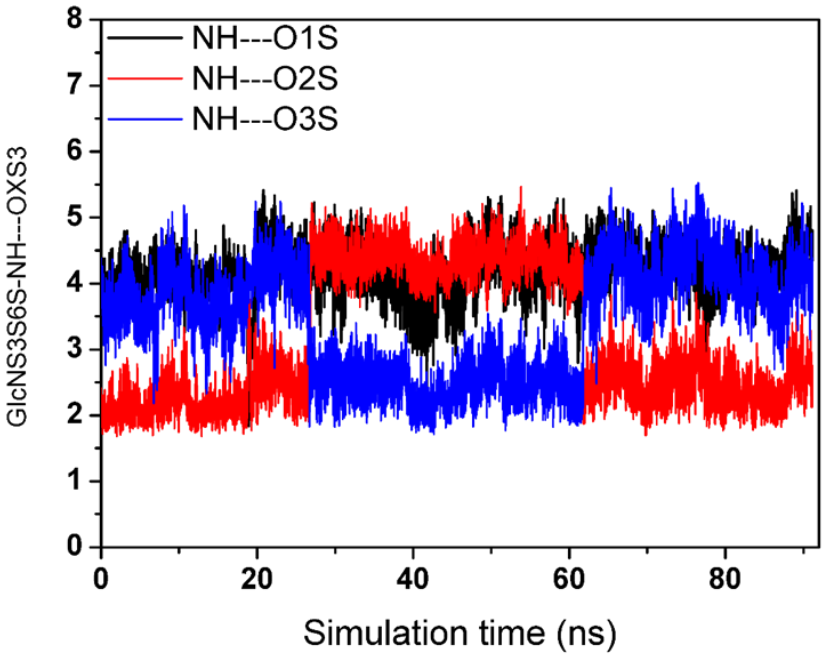


Figure 29. *Trajectories during a 90 ns simulation time of distances between the NH proton of GlcNS3S6S sulfmate group and the oxygen atoms belonging to 3-O-sulfo group of the same residue.*

4.3. Saturation Transfer Difference NMR and water-LOGSY

To identify the binding epitope, ^{1}H STD experiments were performed on octa-5 in the presence of AT (about 100-fold excess of the ligand). The comparison between the ^{1}H STD and the reference proton spectrum is shown in Figure 30. The STD signals of the AIA*IA residues, in particular the trisaccharide moiety AIA*, are the most intense ones, whereas signals of residues belonging to the

reducing and non-reducing end extensions show lower intensity. H1 anomeric signal of GlcNS6S residue f was the most intense and then was chosen as reference peak. The signal superposition only selected H2 resonances to be compared with the reference signal (Table 13). Relatively intense signals were observed also for H1 and H2 of the GlcNS3S6S residue in the STD spectrum, indicating that this residue is close to the binding region. On the contrary, both H1 and H2 of GlcNS6S residue "h" show weak STD signals, indicating a larger distance of this residue from AT.

Table 13.

Selected ^{1}H-STD intensities as percentage relative to the H1 of GlcNS6S residue "f" of Octa-5.

	^{1}H-STD relative intensities (%)							
Ligand	**H1 GlcNS6S "h"**	**H1 GlcNS6S "f"**	**H1 IdoA2S "e"**	**H1 GlcNS3S6S "d"**	**H1 IdoA2S "c"**	**H1 GlcNS6S "b"**	**H1 GlcA "a"**	
Octa-5	36	100	65	82	60	62	25	
	H2 GlcNS6S "h"	**H2 GlcNS6S "f"**		**H2 GlcNS3S6S "d"**		**H2 GlcNS6S "b"**		**H2 GlcA "g"**
Octa-5	50	97		90		74		38

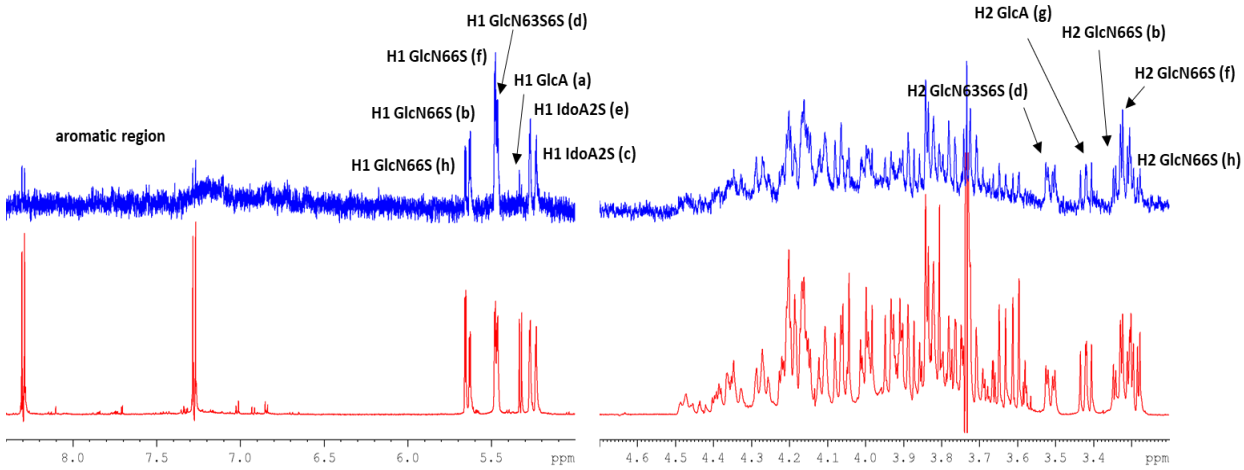


Figure 30. *Partial 600 MHz 1H-STDNMR spectra of the octa-5·AT complexes. The experiment was conducted with a glycan/AT molar ratio of 100:1; the temperature was set to 35°C. STD and reference 1H spectra are depicted in blue and red, respectively.*

The water-LOGSY experiment is an alternative method to STD for the identification of the binding motif of the ligands. The experiment is based on the NOE effect and implies the transfer of magnetization, via intermolecular NOE and spin diffusion, from the water (water interacting with to the protein surface or bulk water) to the ligand protons. In a sample containing a small molecule and a macromolecular receptor, the signals of the small molecule appears with the same phase than the water peak when the ligand is bound to the receptor (slow tumbling motion) and in anti-phase when the ligand interacts only with the bulk water (fast tumbling motion). In our case we applied the water-LOGSY experiment to confirm what observed in the Saturation Transfer Difference spectra. As shown in Figure 31, signals of the reducing-end GlcA (residue A) and para-nitro-phenol moieties appear in opposite phase compared to the whole spectrum. These data strengthen the results of the STD and the ^{1}H chemical shift perturbation spectra, confirming that the inner pentasaccharide moiety is the part of the molecule the strongly interacts with the HBS, while both reducing and non-reducing end show weaker interaction because placed at higher distance from the binding site.

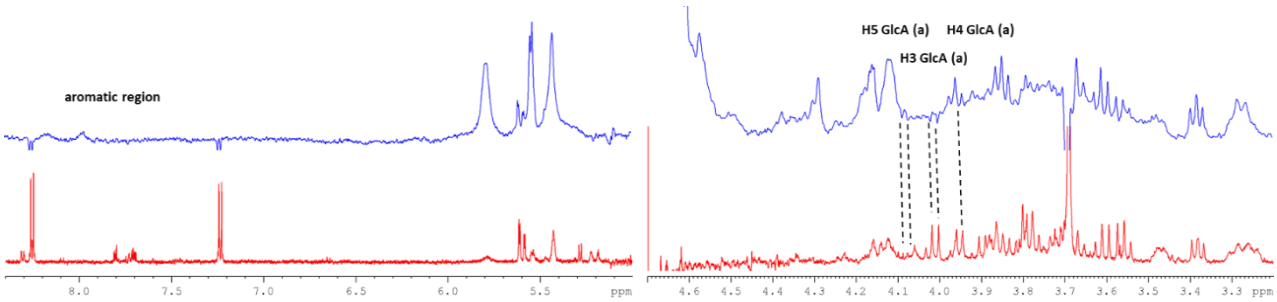


Figure 31. *Partial 600 MHz water-LOGSY spectra of the octa-5·AT complexes. The experiment was conducted with a glycan/AT molar ratio of 100:1; the temperature was set to 30°C. Reference ^{1}H spectra and w-LOGSY are depicted in blue and red, respectively.*

4.4. NOE and tr-NOE

Adding AT to a solution of octa-5 (molar ratio 1:5 oligosaccharide/AT) the line width of signals in the ^{1}H spectrum increased by up to 3 Hz, indicating an interaction between the two molecules [Figure 32]. The small protein-induced shift observed in the spectrum is consistent with an equilibrium regulated by the slow to intermediate dynamic exchange between the free and bound states. ^{1}H Chemical shift perturbation and signal shape widening induced by the interaction with AT, were observed for all the residues of the octa-5. In particular, the inner tetrasaccharide sequence GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S shows the most intense chemical shift perturbation and the highest line broadening. On the other hands the outer residues on both reducing and non-reducing end appear slightly less involved in the binding. These evidences suggest that, as already observed in ^{1}H-STD interaction studies, the strongest contacts with HBS involve the central part of the octa-5, also in according to the hexasaccharide·AT interaction.

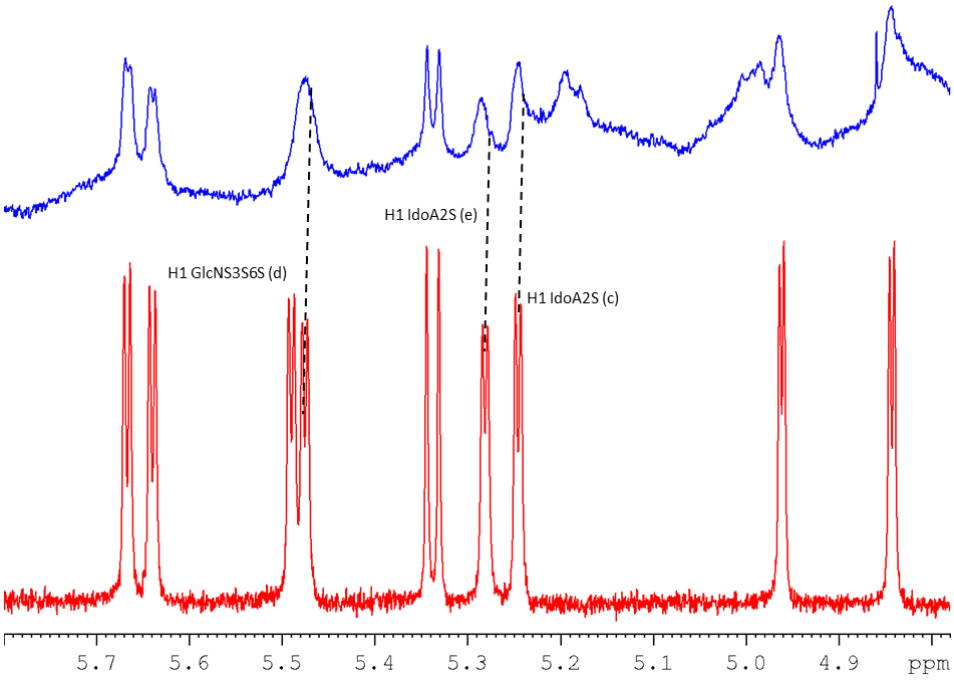


Figure 32. *^{1}H NMR spectrum of free octa-5 in red and ^{1}H NMR spectrum of octa-5·AT complex in blue (molar ratio 5:1)*

Figure 33 compares NOEs and tr-NOEs of H5 of the two iduronic acid residues in octa-5, in both un-bound and bound state with AT, respectively. Due to the difference in cross-relaxation rates between the free and the bound state, significant variation in the cross-peak intensities could be observed comparing NOESY (free state) and tr-NOESY (bound state) spectra. The conformation of both Iduronic acids were investigated by the analysis of the tr-NOESY experiments (mixing time= 200, 300 and 500 ms). A significant change in the ratio between H5-H2 and H5-H4 NOE magnitudes magnitudes for both IdoA2S (residue c) and IdoA2S (residue e) compared with the free state, was observed for the octa-5–AT complex (Table 14).

Table 14

Percentage of IdoA2S conformers in free and bound state.

Sample	H2-H5/H4-H5 ratio
Octa-5 Idoa2S “c”	0.4
Octa-5 Idoa2S “e”	1.4
Octa-5 Idoa2S “c” +AT	0.4
Octa-5 Idoa2S “e” +AT	0.1

As expected for what observed in AGA*IA[71] the interaction with AT drives the IdoA2S residue *c* from a mix of $^{1}C_{4}$ and $^{2}S_{0}$ forms, towards a pure $^{2}S_{0}$ conformation. This behavior is confirmed by the NOESY ratio between H5-H2/H5-H4 that change from 0.4 to 1.4 upon binding with AT, indicating that the distances H5-H2 initially greater than H5-H4, become comparable when the octa-5 is bound to AT. Differently, the IdoA2S residue *e*, from an equilibrium between a $^{1}C_{4}$ and $^{2}S_{0}$ conformation in the un-bound state, becomes a pure $^{1}C_{4}$ form in the complex octa-5 with AT.

Modelling of the octa·5·AT complex suggests that the IdoA2S residue *e* is in $^{1}C_{4}$ chair form mimics the behavior of GlcA in AGA*IA, allowing the same contacts between the carboxylic group and protein, even if additional contacts that involve the 2-O-sulfo group are possible, in particular with R46 of the helix A. The interaction between octa-5 and AT affects the glycosidic linkage between Iduronic residues and adjacent glucosamines. A significant decrease in the ratio between H4 of residue *c* and H6a of residue *b* was observed. The same decrease of intensity was observed, on a smaller scale, between the residues *e* and *d*. In contrast the ratio between the H4 and the H6b of both couple of residues is unaffected suggesting a variation in glycosidic geometry upon the binding. The validation of the model using theoretical and computed tr-NOEs using CORCEMA software is still in progress, in order to provide useful indications of the bound octasaccharide geometry.

4.5. Octa-5 conformational characterization (MD simulation)

The molecular models of octa-5 in un-bound and bound state with AT was built following the same procedure previously reported for the hexa-4 and hexa-8, and described in the material and method section. In this case the pentasaccharide sequence: GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S (AIA*IA) included in octa-5, that correspond to the AGA*IA sequence in heparin chains, was fitted to the corresponding of the 3D structure of the complex AGA*IA·AT (PDBID 1AZX), matching the trisaccharide sequence GlcNS6S-IdoA2S-GlcNS3S6S. This allow the AIA*IA sequence of octa-5 to fit the AGA*IA sequence of the complex. Longer MD simulations were required to equilibrate each system considering the greater size of the ligand and the complex. The MD simulation spanned a period of 350 and 480 ns for octa-5 in unbound and bound state respectively, while the conformational

equilibration required 200 and 250 ns approximately, as determined following the RMSD of the ligand, and the set of glycosidic dihedral angles ϕ_i/ψ_i. The MD simulation describe all the glucosamines in $^4C_1$ conformation for octa-5 in both un-bound and bound state, as was set at the beginnning, and in accord to what previously observed in the shorther heparin like oligosaccharides. Differently, the trisaccharide moiety IdoA2S-GlcNS3S6S-IdoA2S that characterized octa-5, was built setting the non-reducing and reducing end IdoA2S residues in $^1C_4$ and $^2S_0$ respectively, as suggested by preliminary NOEs investigation on octa-5 in solution; the MD simulation of octa-5 in solution (un-bound state) describe these residues preserving their initial conformation. In the complex octa-5·AT, the IdoA2S located respectively at the non-reducing and reducing end of the central tri-sulfated glucosamine (GlcNS3S6S) were set to $^1C_4$ and $^2S_0$ conformations respectively, in accord to the preliminary tr-NOEs study; also the MD simulation preserve the conformation of these two IdoA2S, as detected comparing the average distances H2-H5 and H4-H5, that in the former case H2-H5 is greater than H4-H5, while in the latter case H2-H5 compare better with H4-H5. This description correspond to a non-reducing end IdoA2S in $^1C_4$ chair, while the reducing end IdoA2S sketch the skew boat $^2S_0$. The set of glycosidic dihedral angle distributions ϕ_i/ψ_i describe the main chain conformation of octa-5 in both un-bound and bound state to AT; this property is reported as Ramachandran plots in Figure 34, in the first and second column respectively. The seven glycosidic dihedral angle states are represented by clusters of pairs ϕ_i/ψ_I centred by the corresponding average (red stars), which numerical values are also reported in table 15. Comparing the ϕ_i/ψ_i distribution of octa-5 in un-bound and bound states, the glycosidic linkage show a significant reduction of the conformational flexibility upon binding, in particular for the sequence GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S (AIA*IA sequence) that fit the HBS in AT, mimicking the standard AGA*IA. This result is clearly seen comparing the size of the ϕ_i/ψ_i distributions in panels C-L, D-M,

and E-N, respectively, in Figure 34. Also interestingly, octa-5 bound to AT show a glycosidic degree of freedom that change significantly from the non-reducing to the reducing end, with the "central" AIA*IA sequence presenting the smallest flexibility, and correspond to the part of the glycan with the strongest contacts with AT (compare panels H, I and P with L M and N of figure 34). The glycosidic conformational change that octa-5 undergoes upon binding, as depicted by molecular modelling, is evident at the two ends of the glycan: at the non-reducing end, where two linkages GlcNS6S-GlcA and GlcA-GlcNS6S are involved (Figure 34 panels A-H and B-I, and tab. 15), and also at the opposite side (Figure 34 panels G-P, and table 15). Differently, the glycosidic conformation of the "central" sequence (AIA*IA) in octa-5 (ϕ_i/ψ_I, i = 3, 6), preserve its initial conformation upon binding, as previously observed for the hexasaccharides 4 and 8 of previous chapter; This reinforces the idea that AT binding GAG sequences, require only mild backbone rearrangements to fit AT, while the only significant conformational change appear located at both IdoA2S residues in the "IdoA2S-GlcNS3S6S-IdoA2S" moiety. Interestingly, these results are also supported by the preliminary analysis of the NOEs and tr-NOEs data, in fact by this comparison tiny conformational changes are detected at the glycosidic linkage GlcNS6S-IdoA2S (*f* and *e* residues) and IdoA2S-GlcNS6S (*c* and *b* residues, see oligosaccharide letter code in Figure 25), while the most significant conformational changes involve the two IdoA2S residues that in both cases from a mixed: $^{1}C_{4}$, $^{2}S_{0}$ conformation, move to a pure $^{1}C_{4}$ and $^{2}S_{0}$ for the non-reducing and reducing IdoA2S in that order (still in progress). The glycosidic backbone conformational properties of octa-5 in bound state with AT show similarities with the corresponding of hexa-4 and hexa-8 analysed previously, considering that octa-5 differs from hexa-4 and hexa-8 by a longer non reducing end, and for the additional IdoA2S that replace GlcA (See Figure 17 chapter 3 and Figure 25 Chapter 4). The significant difference between octa-5 and the two hexasaccharides is located to the reducing end GlcNS6S-GlcA,

where the ϕ_7/ψ_7 distribution present two accessible states (Figure 34 panel P and Figure 21 panel L). This suggest that the longer non-reducing end, and the IdoA2S that replace GlcA in octa-5 slightly affect the position of the AIA*IA sequence in AT HBS, in comparison to the standard AGA*IA, allowing the reducing end glycosidic linkage GlcNS6S-GlcA (ϕ_7/ψ_7) to populate two dihedral angle states instead of one. The structural details of the octa-5·AT molecular recognition are summarized in Figure 35 where two geometries of this complex are extracted from the MD simulations at 200 and 400 ns respectively, corresponding to middle and long-time spanned respectively. The two geometries are superimposed to the reference AGA*IA·AT geometry of the PDB ID 1AZX by the A-helix Cα backbone, this allow to visualize the AIA*IA sequence of octa-5 matching the reference pentasaccharide. Interestingly, the IdoA2S that replaces GlcA in the reference pentasaccharide, preserves the carboxyl group orientation of this latter residue, and the corresponding AT contacts K11 and K125, while the 2-O-sulfo-group contribute as an additional contact for the side chain of R46 (results in development). Figure 35 gives also a pictorial representation of the higher conformational freedom (mobility) that both non-reducing and reducing end of octa-5 in bound state with AT, as previously described by the distribution of glycosidic dihedral angles; in this case the two terminals of octa-5 visibly show a greater mobility in comparison to the firmly bound AIA*IA sequence. The difference in binding strength between the "inner" and "terminal" residues of octa-5 complexed with AT, as drawn by the molecular modelling investigation, is supported by the ^{1}H-STD intensities of the anomeric protons reported in Tab 11 and complemented by the water-LOGSY description. (Paragraph 4.3 and Fig. 30, and Fig. 31) In this case the AIA*IA sequence in octa-5 show the strongest ^{1}H-STD intensities, corresponding to the glycan part firmly bound to HBS of ATATBR, while both terminals show weaker interactions, that correspond to greater distances between glycan and residues of AT contacts, but also this correspond to a higher solvation degree.

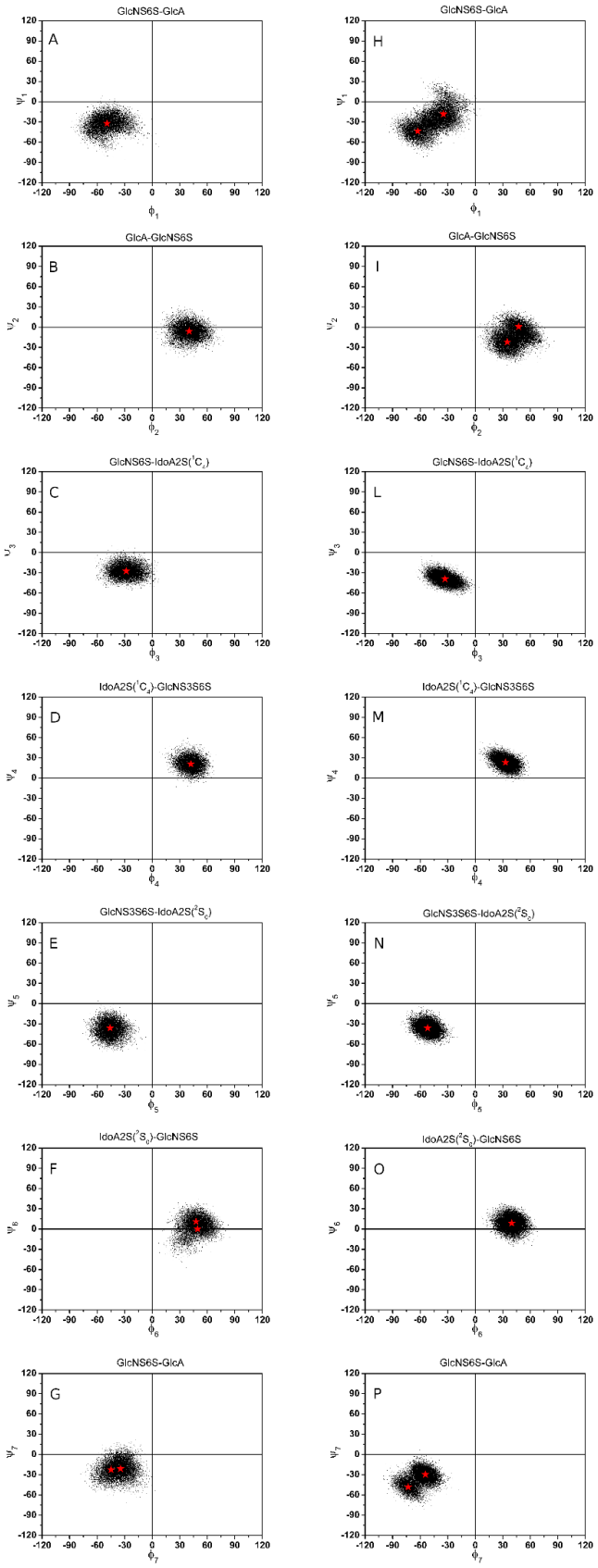


Figure 34. *Glycosidic backbone dihedral angle distribution ϕ_i/ψ_i (Ramachandran plots) that octa-5 sample in un-bound state (panels from A to G) and bound state with AT (panels from H to P) respectively. The red stars indicate the centre of each clusters of dihedral angle states as determined by geometrical average.*

Table 15

*Glycosidic dihedral angle states (degree) predicted by the MD simulation description of octa-5 in un-bound and bound state with AT. The glycosidic dihedrals in the AIA*IA that fit the HBS of AT are underlined in red.*

	Glycosidic linkage type	Octa-5	Octa-5-AT
ϕ_1/ψ_1	GlcNS6S-GlcA	-49/-32	-63/-44, -35/-19
ϕ_2/ψ_2	GlcA-GlcNS6S	40/-6	35/-22, 47/1
ϕ_3/ψ_3	GlcNS6S-IdoA2S	-28/-27	-33/-39
ϕ_4/ψ_4	IdoA2S-GlcNS3S6S	42/21	33/23
ϕ_5/ψ_5	GlcNS3S6S-IdoA2S	-46/-36	-52/-36
ϕ_6/ψ_6	IdoA2S-GlcNS6S	48/11, 49/0	40/9
ϕ_7/ψ_7	GlcNS6S-GlcA	-35/-21, -45/-22	-73/-48, -54/-30

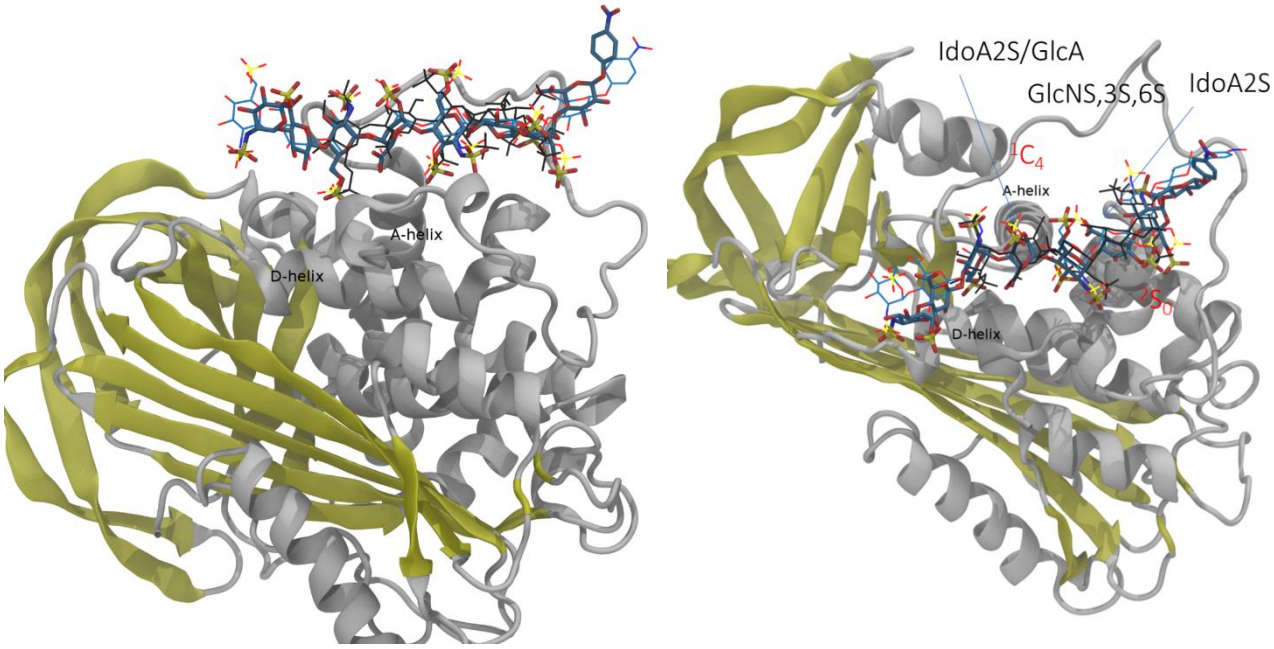


Figure 35. *Octa-5-AT complex snapshots (side and top views) after 200 and 400 ns of MD simulation superposed by the A-helix of AT. The AT is represented by the white and yellow ribbon, while the octa-5 is drawn in licorice and blue wire. The AGA*IA pentasaccharide is drawn by the black wire superposing the complex AGA*IA-AT (PDB ID 1AZX) by the A-helix to the previous.*

Table 16

Selected intra-residues and inter-glycosidic ^1H NOEs and tr-NOEs for octa-5 and octa-5 interacting with AT.

		NOEs %			Tr-NOEs %		
Tmix (ms)		200	300	500	200	300	500
IdoA2S(E)	H2-H5	3.3	5.0	9.0	1.0	2.2	2.9
	H4-H5	7.4	10.7	20.1	10.6	18.9	21.8
IdoA2S(C)	H2-H5	2.7	4.0	7.1	6.9	14.9	20.7
	H4-H5	6.3	9.0	15.9	5.5	9.2	12.7
Tmix (ms)		200	300	500	200	300	500
GlCNS-IdoA2S	F1-E4	4.9	7.0	11.3	12.3	20.4	24.7
	F1-E3	6.7	10.2	16.5	7.8	11.3	12.5
IdoA2S-GlcNS3S6S	E1-D4	3.3	5.2	9.4	2.9	3.5	3.6
	E1-D6a	9.5	12.7	20.2	5.6	6.4	11.5
	E1-D6b	3.9	5.2	10.9	1.6	4.4	9.5
IdoA2S-GlcNS6S	C1-B4	5.7	7.6	12.7	6.3	9.3	14.3
	C1-B6a	1.5	2.1	4.5	-	-	-
	C1-B6b	4.8	5.8	10.5	5.2	5.6	10.6

4.6. Conclusions

The chemoenzymatic procedure allowed the preparation of an artificial heparin-like octasaccharide (octa-5) characterized by a central pentasaccharide sequence variant: GlcNS6S-GlcA-**GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S**-GlcA-$PhNO_2$, in which -IdoA2S-GlcNS3S6S- replaces the typical -GlcA-GlcNS3S6S- sequence of AGA*IA, known to be the shortest glycan that specifically bind and activate AT. The research group of Prof. J. Liu demonstrated for the first time the biological activity of octa-5 on the coagulation system, showing that octa-5 binds and activates AT with greater potency than fondaparinux pentasaccharide[75]. To rationalise these results, we started a structural and conformational characterization of the octa-5 in un-bound and bound state with AT. The basic structural characterization confirms the chemical structure/connectivity of this glycan, as predicted by the chemoenzymatic synthesis. The preliminary conformational characterization of the octa-5 in un-bound state, indicated that the GlcN residues and the glycosidic backbone conformations does not change significantly from that of the heparin-like hexasaccharides previously investigated (hexa-4,

hexa-8). Molecular dynamic supported ^{1}H-STD and the complementary water-LOGSY experiments, show that the AIA*IA central moiety presents the strongest contacts with the HBS of AT, while both non-reducing and reducing ends are positioned further from the residues of AT preferring contacts with the solvent. The most relevant conformational change of octa-5 upon binding involve the two IdoA2S, that change from mixed $^{1}C_{4}$ and $^{2}S_{0}$ forms, to a defined $^{1}C_{4}$ and $^{2}S_{0}$ conformation for the IdoA2S located in "*e*" and "*c*" positions, respectively. Comparing the few inter-glycosidic geometries determined by NOEs and tr-NOEs, weak changes of the glycosidic backbone of octa-5 were detected. The modelling description of octa-5 supports the conformational change of the two IdoA2S residues in going from the un-bound to the bound state, preliminary observations suggest that the contact network between the AIA*IA moiety and the HBS of AT, forces IdoA2S in "e" and "c" positions to assume the $^{1}C_{4}$ and $^{2}S_{0}$ conformation, respectively. This evidence was the first example in solution, in which two IdoA2S residues belonging to a single oligosaccharide, adopted two different conformations when bound to protein, highlighting the importance of the conformational flexibility of IdoA residues in intermolecular interactions. A refined analysis of inter-glycosidic tr-NOEs signals, together the MD simulation description of the octa-5·AT complex through a semi-quantitative analysis based on the CORCEMA approach, is still in progress. These results are important because they contribute and refine the Structure Activity Relationship description of the still growing set of heparin-like oligosaccharides able to bind and activate AT.

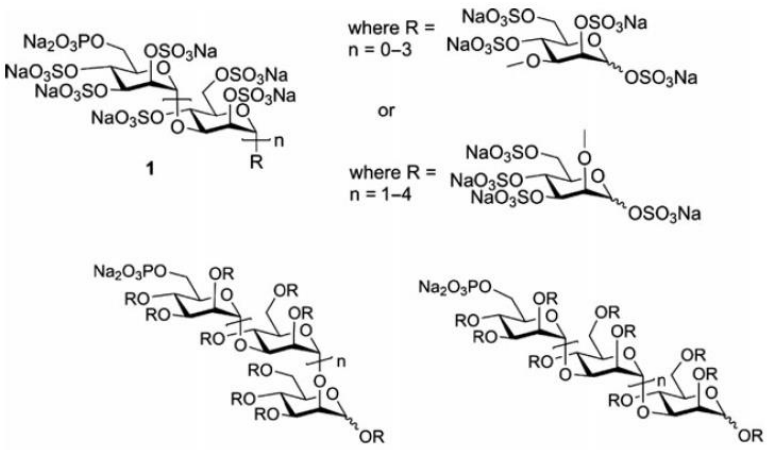
5. Structural and Conformational Studies of the Heparan Sulfate Mimetic PI-88

5.1. Background

The heparan sulfate (HS) mimetic PI-88 (also known as muparfostat or "phosphomannopentaose sulfate")[76] is an inhibitor of angiogenesis, tumor growth and metastasis that has been in clinical development for the past two decades for various cancer indications[77,78]. Most recently, PI-88 was evaluated in an international, multi-center Phase III clinical trial as an adjuvant therapy for post-resection hepatocellular carcinoma[79]. PI-88 is a potent inhibitor of angiogenesis[76] . This activity is mediated through inhibition of heparanase[76], an endo-β-glucuronidase that plays a key role in both metastasis and angiogenesis[80], and more directly by blocking the interactions of angiogenic growth factors, such as FGF-1, FGF-2 and VEGF, and their receptors with HS[76].

PI-88 (1, Figure 36) is a complex mixture of monophosphorylated, polysulfated mannose oligosaccharides. It is prepared by the exhaustive sulfonation of the oligosaccharide phosphate fraction (OPF) obtained following mild acid-catalysed hydrolysis of the extracellular phosphomannan produced by the yeast *Pichia holstii* NRRL Y-2448[81,82]. This mixture is primarily composed of phosphorylated α(1→3)/α(1→2)-linked penta- and tetrasaccharides, which together account for approximately 90% of the total oligosaccharide content, with the remaining 10% comprised of phosphorylated di-, tri- and hexasaccharides[83].Recently, the OPF was successfully fractionated for the first time by preparative ion exchange chromatography and the major oligosaccharides were isolated and characterized by NMR spectroscopy and MS[84]. The presence of the previously reported oligosaccharide phosphates 2, 4, 6 and 8, which possess α(1→3)- linked mannoses with a terminal

α(1→2)-linked residue, was confirmed with 4 and 6 as the major components [Figure 36]. The presence of the all-α(1→3)-linked isomers 12, 14, and 16 resulting from hydrolytic cleavage of higher oligosaccharides, and present in amounts that vary inversely with oligosaccharide chain length, was also confirmed. It was also shown that the only disaccharide phosphate present is α(1→3)-linked[84]. Given that the OPF is exhaustively sulfonated to produce PI-88, it was proposed that the structure of PI-88 is best represented by structure 1 [Figure 36] as this more accurately reflects its composition.



	n	R		n	R
2	1	H	**10**	0	H
3	1	SO_3Na	**11**	0	SO_3Na
4	2	H	**12**	1	H
5	2	SO_3Na	**13**	1	SO_3Na
6	3	H	**14**	2	H
7	3	SO_3Na	**15**	2	SO_3Na
8	4	H	**16**	3	H
9	4	SO_3Na	**17**	3	SO_3Na

Figure 36. *Revised structure for PI-88 (1), and structures of the oligosaccharides from PI-88 and its nonsulfated precursor (2-17).*

Herein we describe the preparation of the individual components of PI-88 from the successfully separated OPF oligosaccharides and their detailed structural characterization by NMR spectroscopy, which, when combined with detailed NMR spectroscopic analysis of PI-88 itself, allowed for the assignment of many minor resonances in the ^{1}H NMR spectrum of PI-88 and provided new insights into its composition and level of sulfation. In addition, a detailed conformational analysis of the phosphomannopentaose sulfate, the major component of PI-88, was carried out by combining data from NOESY spectra with molecular dynamics simulations.

5.2. Results and Discussion

Samples of the purified OPF oligosaccharides[84] were individually sulfonated by treatment with excess sulfur trioxide pyridine complex to yield the corresponding PI-88 component oligosaccharides following purification by size exclusion chromatography (Bio-Gel P2). The purity was determined to be ≥95% by capillary electrophoresis. After sulfonation, the corresponding sulfated pentasaccharide fraction was examined in D2O at both 500 and 600 MHz using one- and two-dimensional NMR techniques and the ^{1}H and ^{13}C NMR chemical shifts were fully assigned using the HSQC, HMBC, COSY and TOCSY pulse sequences (data not presented). Ambiguous assignments, such as the interring connectivity were resolved by NOESY [Figure 38] and HMBC experiments [Figure 39]. These data clearly established the fully sulfated α-anomer of pentasaccharide **7** as the major component, with ~10% of the β-anomer also present. A significant downfield shift was generally observed in the ^{1}H NMR spectrum for the resonances of protons geminal to hydroxy groups upon sulfation, with no signals between 3.6 and 4 ppm, indicative of full *O*-sulfation [Figure 39].

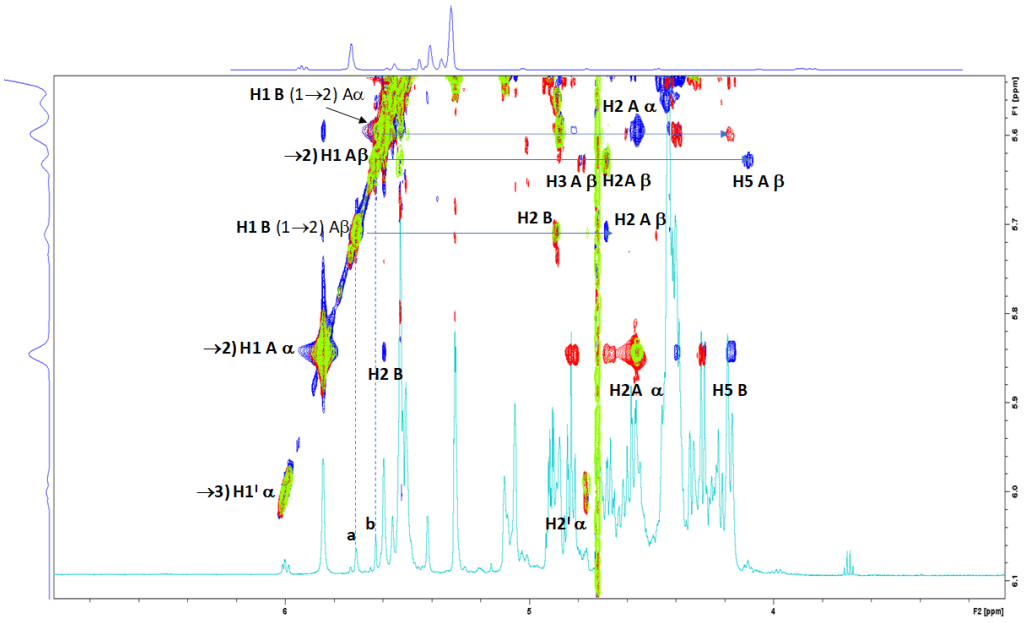


Figure 37. *Partial 600MHz NOESY (blue), COSY (green) and TOCSY (blue) spectra of PI-88 (**1**).*

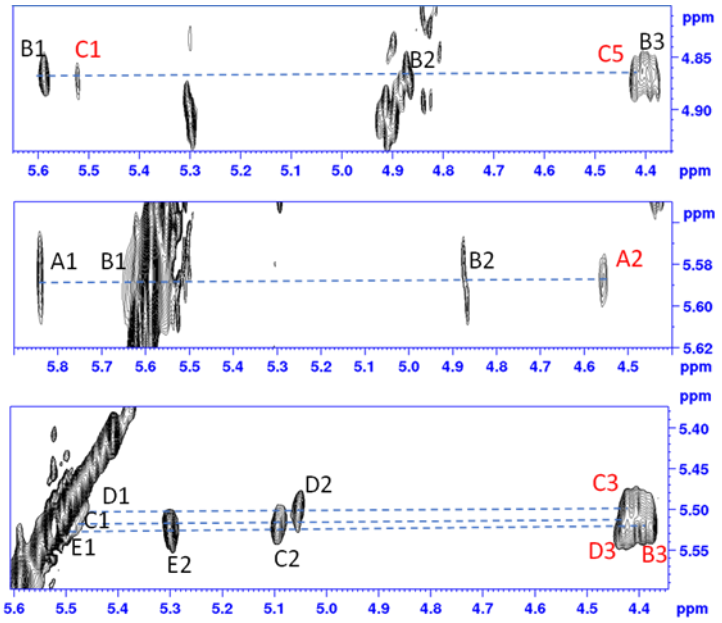


Figure 38. *Portions of 600MHz NOESY spectrum of PI-88 (1). Intra and inter-residue noe's are indicated in black and red, respectively.*

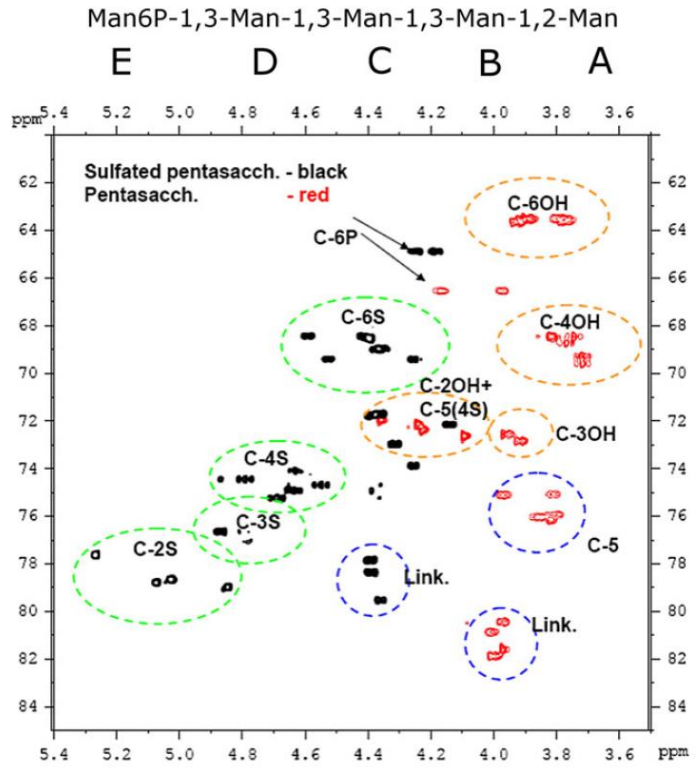


Figure 39. *Superimposed HSQC spectra of sulfated (7) and nonsulfated (6) pentasaccharide fraction, showing significant chemical shift differences.*

Similarly, the resonances of carbon atoms attached to sulfo groups in **7** were shifted downfield relative to the corresponding resonances in the non-sulfated precursor **6** [Figure 39]. Carbon atoms involved in glycosidic linkages were observed at slightly higher field upon sulfation elsewhere in the molecule (<3 ppm for ^{13}C). The average resonances of the hydroxylated carbon atoms shifted 6.0 ppm downfield upon sulfation, and the corresponding shifts in the proton spectrum were 0.73 ppm. The small shift for C6 of the non-reducing end mannose residue is consistent with the presence of a phosphate group in this position. The shapes of the signals for the anomeric protons of the reducing end residue (A) in the NMR spectra of **7** are noteworthy. The broader line widths (i.e., A1-α in Figure 40) are probably due to shorter relaxation times (T_2) with respect to the other anomeric protons. The presence of a very high degree of sulfation influences the internal mobility of the molecule with the introduction of local rigidity due to steric hindrance and/or inter-residue interactions, as discussed in the MD simulations section. The major components of the sulfated tetra-, tri-, and disaccharide fractions were similarly assigned as described above. The anomeric ratios observed in the non-sulfated oligosaccharide precursors were roughly preserved upon sulfation. Although the minor β-anomers (Aβ ring) were fully assigned in the case of all-(1→3)-linked structures 11 and 13, in the case of oligosaccharides 3, 5 and 7 with a terminal (1→2)-linked residue, the minor β-anomers were much less abundant.

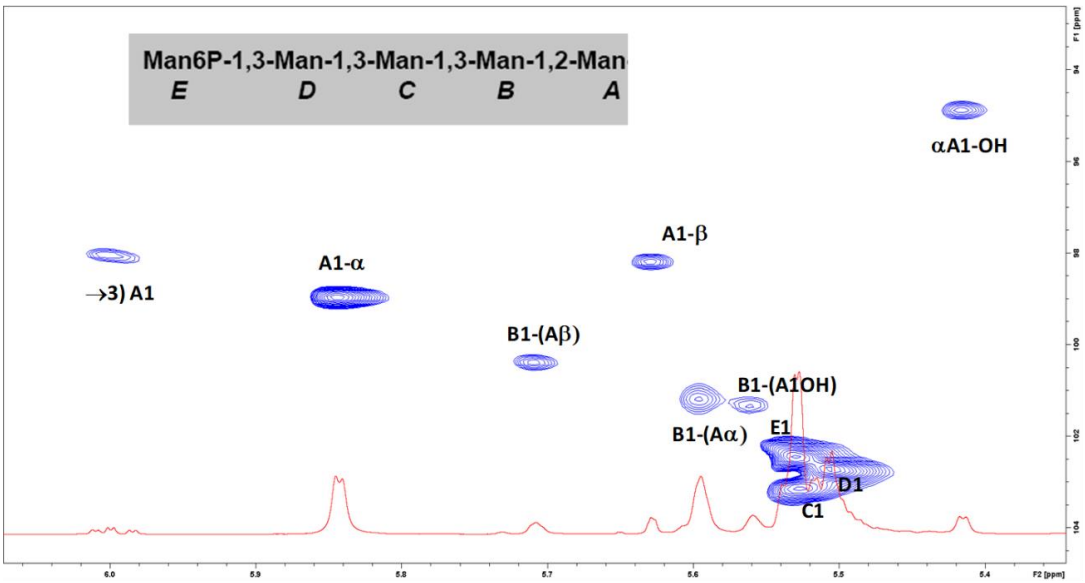


Figure 40. *Partial HSQC spectrum of PI-88 (1) in which is possible to observe differences in linewidth correlate to different T2.*

The NMR spectral resonances of the latter compounds were insufficiently resolved for full assignment; partial assignments were made as illustrated for the pentasaccharide 7 where the protons nearest the reducing-end residue could be determined. Confirmation of the β-D-configuration at the anomeric centre was available from the unusual downfield shift of the anomeric proton of residue (B) (labelled peak H1 B-(Aβ) in Figure 41) upon NMR analysis.

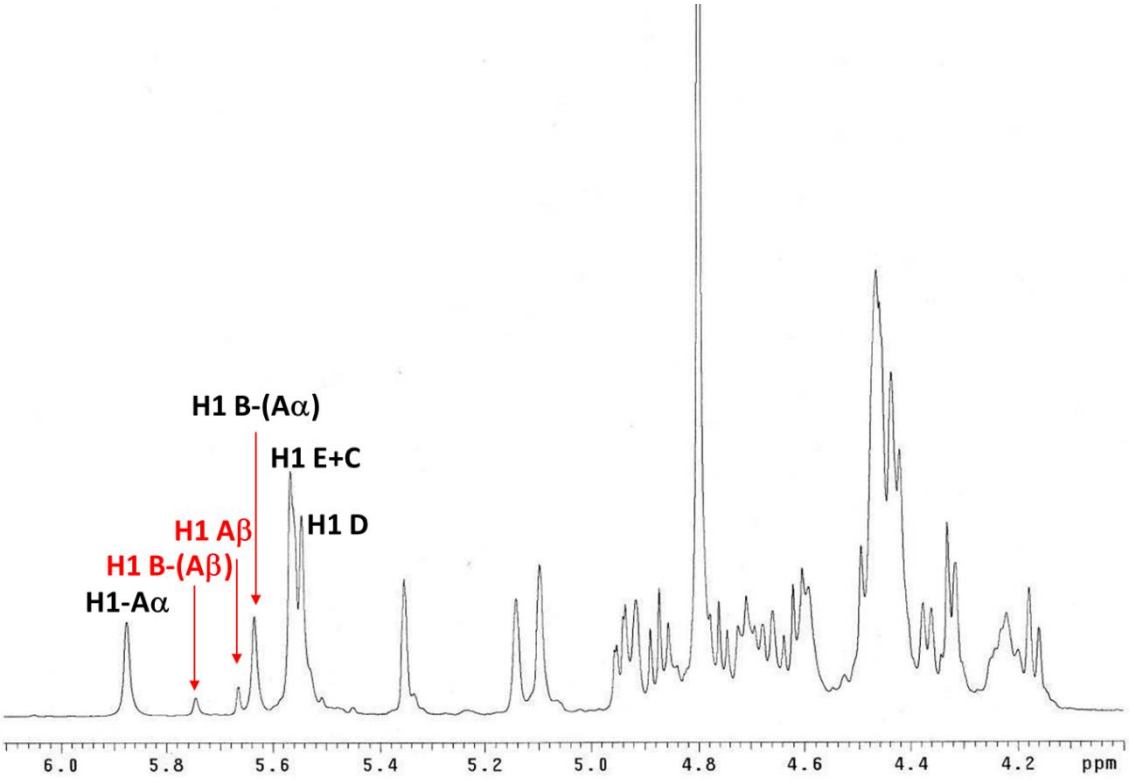


Figure 41. *^{1}H NMR (600 MHz) spectrum of pentasaccharide 7 in D_2O showing peaks associated to the minor b-anomer and major a-anomer of residues A and B, highlighted in red and black, respectively.*

This proton resonates at a chemical shift of δH 5.76, which is at unusually low field for a non-sulfated anomeric center. Figure 41 also shows the peak assigned to proton H-1 Aβ at the sulfated anomeric center. The chemical shift of peak H1 B-(Aβ) can be rationalised as the result of deshielding by a through-space interaction with at least three sulfate groups that are linked to C2 of residue B, and also to C1 and C3 of residue A[85]. Further confirmation of the assignment of peaks H1 B-(Aβ) and H1 Aβ was available from the NOESY spectrum and, as the expected, cross-peaks were observed between both protons and H2 Aβ [Figure 33]. The NMR spectra of PI-88 and of the sulfated tetrasaccharide 5 exhibited a very similar pattern of resonances assigned to α- and β-anomers [Figure 40]. NMR spectroscopic characterization of the sulfated trisaccharide fraction confirmed that the major components were the α- and β-anomers of compound 13, and the α- and β-anomers of compound 3,

in the ratio of 5:2:3:0.2. The ^{1}H NMR spectra of the first three compounds were fully assigned. The reducing end anomeric proton of the α-anomer of compound 13 is very distinctive, resonating at δH 6.02 ppm, the lowest field of any anomeric proton observed in this study. Thus, the small group of signals around 6 ppm in the ^{1}H NMR spectrum of PI-88 are characteristic of the minor all-(1→3)-linked isomers. The reducing end anomeric proton of the β-anomer of compound 13 is hidden by the strong signals at approximately δH 5.5 ppm. As discussed above, the NMR chemical shifts of hydroxylated carbon atoms and their attached protons were significantly increased upon sulfation of each OPF fraction. As this shift was observed for all 16 free hydroxyl groups in pentasaccharide 6, it was concluded that the major component in the pentasaccharide fraction was the completely per-O-sulfated pentasaccharide 7, with the α-configuration at the anomeric center. The presence of a small amount (~10%) of the corresponding β-anomer was also noted. Similarly, the major components in the sulfated tetra-, tri-, and disaccharide fractions were shown to be fully sulfated, with anomeric ratios consistent with those observed for the unsulfated precursors. This is well illustrated by superimposition of the HSQC spectra of compound 7 and its precursor 6 [Figure 39]. As the NMR spectra of PI-88 were very similar to those of pentasaccharide 7, it was concluded that the major components in PI-88 were also fully per-O-sulfated. This is consistent with the C:S ratio data from elemental analysis[86] and LC-ESI-FTMS. The LC-MS profile shows a main peak corresponding to the mass of compound 7 determined as Mw = 2189.6 Dalton, and minor peaks attributed to the mono-desulfated pentasaccharide and shorter oligomers from di- to tetrasaccharides. The original PI-88 sample in this study was split into two portions, with one sample examined in Australia at 600 MHz and the other sent to Italy for independent analysis initially at 500 MHz. Although the PI-88 samples were identical, the delay in sample measurement due to the long trip to Italy presumably resulted in more opportunity for decomposition as the 500 MHz spectrum exhibited significant differences

compared with that at 600 MHz; the anomeric proton signal of the reducing-end mannose residue was slightly reduced in intensity, and new resonances appeared around 5.42 ppm and 5.56 ppm. The intensities of these signals further increased after two days from NMR tube preparation [Figure 42].

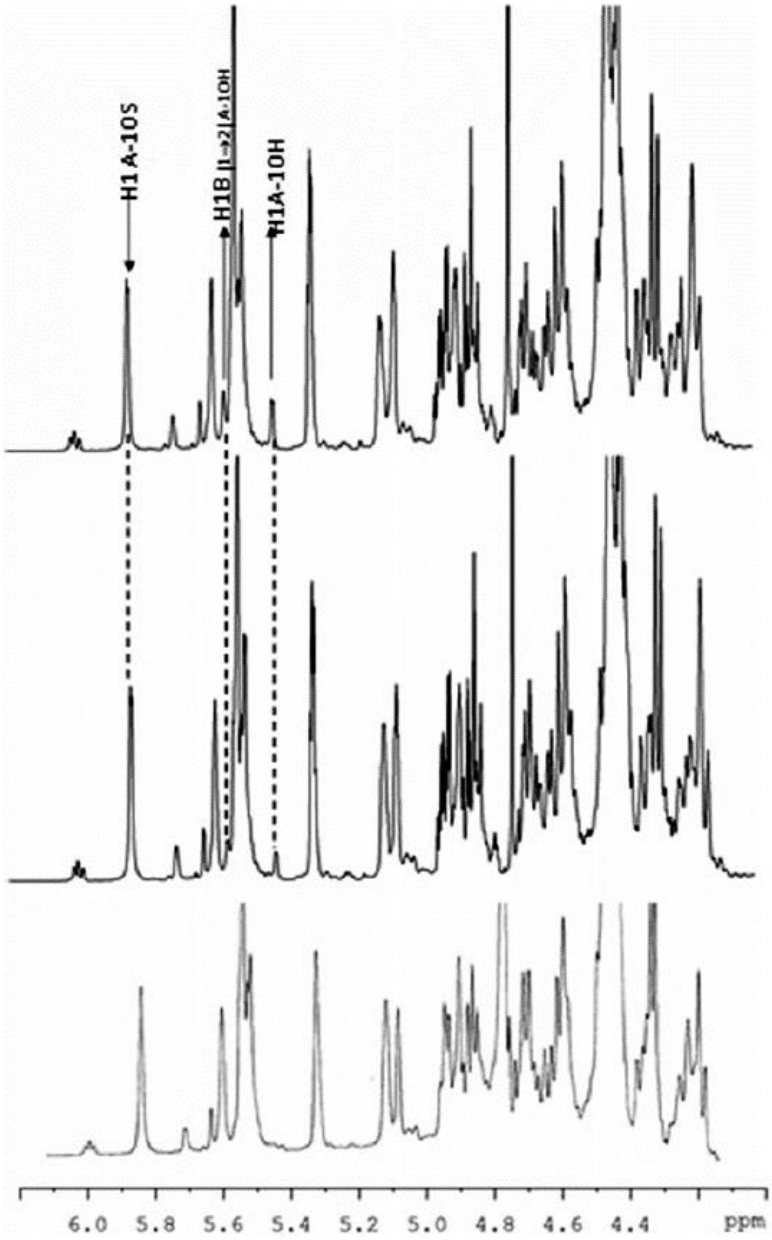


Figure 42. *Decomposition of PI-88. The peaks indicated by the dotted lines differ significantly between the two samples. The bottom trace corresponds to the sample measured in Australia while the intermediate and upper traces correspond to the sample measured in Italy immediately after NMR tube preparation and after two days, respectively.*

These NMR spectral features are consistent with slight decomposition by anomeric desulfation. This was confirmed by superimposition of the HSQC spectra of PI-88 and unsulfated pentasaccharide 6 [Figure 43]. The nascent anomeric resonance corresponding to the reducing end of compound 7, correlates with that of the reducing-end anomeric resonance of pentasaccharide 6, indicating that they are in very similar chemical environments, i.e., both are unsulfated. The extent of anomeric desulfation in the sample was estimated to be 14%, calculated by comparison of the cross-peak volume of its signal with that of its sulfated counterpart. The second peak at δH 5.56 was assigned to mannose residue B linked to the reducing mannose desulfated at the anomeric position. Again, this is best illustrated by the HSQC spectrum [Figure 39] where small resonances at ~ δH/ δC 5.4/95 and ~ δH/ δC 5.55/101 corresponding to the anomeric proton of residues A and B of the desulfated product are indicated by the arrow.

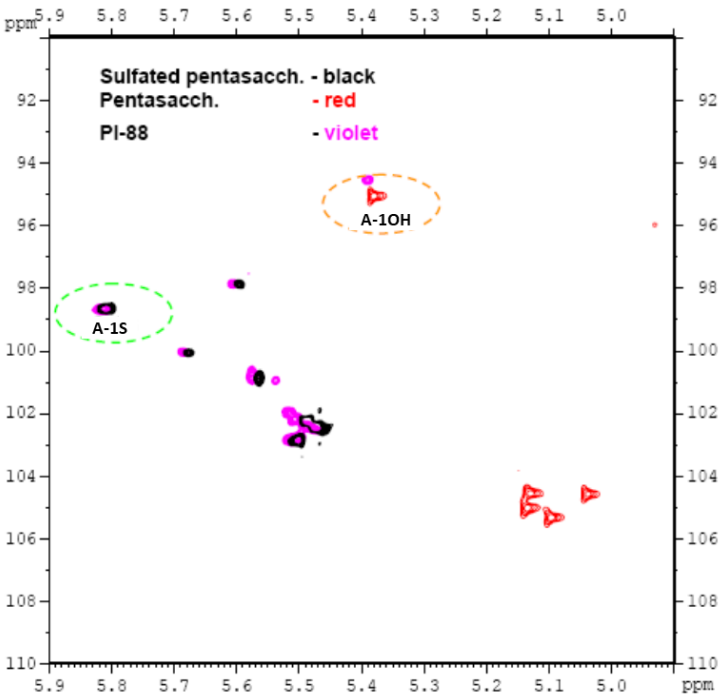


Figure 43. *Identification of the NMR resonances due to slight anomeric desulfation, by comparison of the HSQC spectra of PI-88 (violet) with that of pentasaccharides 6 (red) and 7 (black).*

A detailed conformational analysis of PI-88 in solution was also accomplished. A fresh sample of PI-88 was analysed by NMR spectroscopy at 600 MHz and a complete assignment of the chemical shifts was carried out via COSY, TOCSY, HSQC, HMBC and NOESY experiments. The $^{1}H/^{13}C$ chemical shifts, summarized in Table 17, are largely in agreement with those obtained for the pentasaccharide 7, despite the systematic shift of the carbon resonances (minor for proton), due to the use of a different reference standard.

Table 17

^{1}H and ^{13}C NMR chemical shift data for PI-88 (1) at 600 MHz in D2O.

Residue		H-1/C-1	H-2/C-2	H-3/C-3	H-4/C-4	H-5/C-5	H-6a,b/C-6
α-Man-6-PO4	E	5.53	5.3	4.91	4.84	4.33	4.22/4.18
		102.4	77.9	77	74.9	73.5	64.8
→3)-α-Man	D	5.51	5.06	4.43	4.72	4.39	4.61/4.44
		102.7	78.9	78.1	75.5	72	68.8
→3)-α-Man	C	5.52	5.1	4.42	4.67	4.43	4.43
		103.1	79.1	78.5	75.2	72.1	68.9
→3)-α-Man	B	5.59	4.87	4.39	4.58	4.17	4.4
		101.2	79.2	79.8	75	72.5	69.3
→2)-α-Man-OSO3Na	A	5.845	4.56	4.83	4.67	4.3	4.55/4.28
		98.9	75.2/77.7	77.5	75.2/74.4	74.2	69.7
→2)-β-Man-OSO3Na	A	5.63	4.69	4.78	?	4.11	4.60/4.23
		98.2	75.1	79.2	?	76.9	70.3
→2)-α-Man-OH	A	5.42	4.44	4.83			
		94.9	75.5				
→3)-α-Man-(1→2)-α-Man-OH	B	5.56					
		101.3					
→2)-β-Man-OSO3Na	A	5.63	4.68	4.78	4.1		nd
		98.2	75.3-75.7	79.2	nd	76.9	
→3)-α-Man-(1→2)-β-Man-OSO3Na	B	5.71	4.89	4.91	4.84	4.33	4.22/4.18
		100.4	79.3	77	74.9	73.5	64.8

The inter-residue NOEs not only confirmed inter-residue connectivities, but also provided information on the geometry of the glycosidic linkages. To reduce spin-diffusion effects, the inter-glycosidic proton-proton distances were estimated by two-spin approximation at short mixing time (200 ms). The H1-H2 distance of internal mannose residues (2.53Å) was used as a reference distance. Cross-peaks between H1 and H3 and between H2 and H5 were observed across each glycosidic linkage in going from residue E to B; while the glycosidic linkage involving residues B and A showed H1B-H2A, H1B-H1A and H1A-H5B NOE correlations (See Figures 37 and 38). The corresponding estimated distances are reported in the second column of Table 18. This set of NOE cross-peaks observed between E to A residues of PI-88, agrees with a $^{4}C_{1}$ chair conformation of all mannose residues, similarly to that previously described for unsulfated Man-α(1→3)-Man oligosaccharides[87]. Unfortunately, the heterogeneity of the mixture and the consequent signal broadening does not allow the determination of selected three bond $^{3}J_{HH}$ coupling constants, to directly support the conformation of these residues.

Table 18

Intra- and inter-residue H–H distances of compound 7, estimated by the spin pair approximation using NOE% at the lowest available mixing time (second column).

Atom Pair	Experimental Distance (Å)	MD simulation 2D histogram bins (Å)
E1-D3	2.2	2.3
E2-E1	2.6	2.5
E2-E3	2.8	2.5
D1-C3	2.3	2.2
D2-E5	2.5	2.9
D2-D1	2.4	2.5
D2-D3	2.5	2.4
C1-B3	2.4	2.2
C2-D5	2.5	3.1
C2-C1	2.4	2.5
C2-C3	2.5	2.4
B1-A1	3.4	3
B2-C5	2.5	2.8
B1-A2	2.4	2.2
B2-B1	2.7	2.5
A1-B5	3	3
A1-B1	2.8	3
E1-D3	2.2	2.3
E2-E1	2.6	2.5
E2-E3	2.8	2.5
D1-C3	2.3	2.2
D2-E5	2.5	2.9
D2-D1	2.4	2.5
D2-D3	2.5	2.4
C1-B3	2.4	2.2
C2-D5	2.5	3.1
C2-C1	2.4	2.5
C2-C3	2.5	2.4
B1-A1	3.4	3
B2-C5	2.5	2.8
B1-A2	2.4	2.2
B2-B1	2.7	2.5
A1-B5	3	3
A1-B1	2.8	3

5.3. MD Simulation

The conformational characterization of the polysulfated pentasaccharide and its nonsulfated precursor (compounds 7 and 6, respectively), was carried out by a succession of eleven MD simulation steps with temperature increasing from 300 K (1st step) to a maximum value of 400 K (6th step) and then decreasing back to a final value of 300 K in the final (11th) step, with this last step further considered for glycosidic backbone optimization. The MD simulation procedure was then terminated by a simulated annealing process, in which the temperature was progressively reduced from 300 K to 20 K, until a final energy minimization was applied to generate the final structures to be optimized in their glycosidic backbone, before their comparison with the PI-88 measured NOE restraints. The length of each MD simulation step was set to 20 ns, while the duration of the whole MD simulation sequence was 220 ns (see Material and Methods). This variable temperature approach was chosen considering the high conformational stiffness of the glycosidic linkages of these poly-sulfated oligosaccharides and should allow a more efficient exploration of their conformational space compared with a constant temperature MD simulation of similar length. Conformational transitions of each Man residue, set initially in the $^4C_1$ chair for both sulfated and non-sulfated pentasaccharides, was monitored by the intra-residue distance H3-H5. Even if each residue maintains its initial $^4C_1$ chair for the whole MD simulation, the H3-H5 distance calculated in the final MD simulation sampling (step 11th), suggests a significant distortion of the $^4C_1$ chair in the polysulfated pentasaccharide (compound 7), in comparison with the less crowded nonsulfated precursor (compound 6). In fact, the average H3-H5 distance is significantly greater in compound 7 (> 2.6 Å) than in compound 6 in all the rings except in residue B. (data not presented). The ideal case of a cyclohexane chair conformation shows equal inter-proton distances: H1-H3, H1-H5 and H3-H5 approaching the 2.6 Å value. The optimization of the glycosidic backbone conformations of compound 7 and 6, obtained at the end of

the simulated annealing, involves the determination of the most populated states on the MD simulation trajectories of the last step (11^{th}) in the previously cited multi-step procedure. Ramachandran plots of φ/ψ dihedrals were drawn combining a 2D histogram binning and colored density maps [Figure 44].

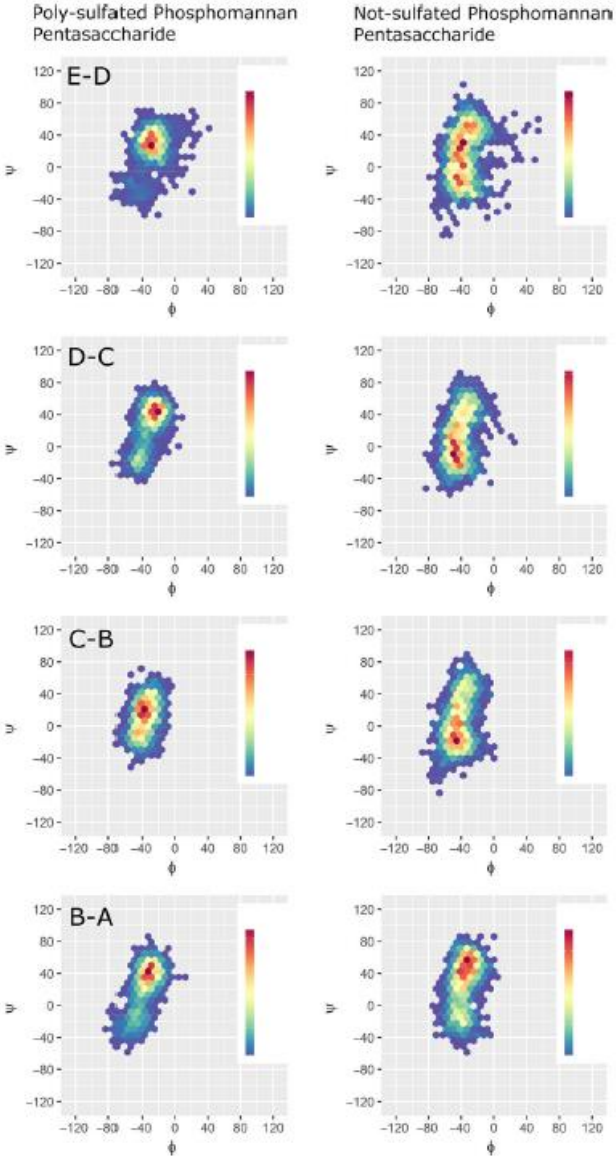


Figure 44. *Glycosidic linkage ϕ/ψ Ramachandran plots of polysulfated pentasaccharide 7 (left column), and nonsulfated pentasaccharide 6 (right column), sampled in the last step of the MD simulation procedure. The glycosidic linkages are indicated by the pairs of residues connected in going from the nonreducing (E) to the reducing end (A). The density color maps qualitatively indicate the population of states ϕ/ψ that increase in going from blue to red color.*

Table 19

Glycosidic dihedral angle pairs ϕ/ψ as determined from the density color map reported in Figure 44.

	φ/ψ Dihedrial angles (°)	
Glycosidic linkage	Sulfated Pentasaccharide 7	Pentasaccharide 6
E-D	−30/30	−30/50,−40/20,−40/−20a
D-C	−30/40	−50/−10
C-B	−40/30	−40/−20
B-A	−30/40	−40/50

The estimated glycosidic dihedral angles with an uncertainty of approximately ±10° are reported in Table 19. The efficiency in the glycosidic backbone conformational sampling was verified on the stiffest molecule (compound 7) comparing the 2D histogram binning analysis on the 1st, the 6th, and the 11th MD simulation steps; the obtained Ramachandran plots are reported in the first, second and third column respectively of Figure 45. Figure 45 shows significant glycosidic linkage conformational change in going from the early to the later stages of the multi-step MD simulation, showing how the stiffest polysulfated phosphomannopentose appears not to reside in its initial conformation, but to visit a growing number of states as the temperature increases, improving the possibility of populating the most favored conformers. Figure 44 shows for both pentasaccharides a comparable φ/ψ distribution at central linkages (D-C and C-B), while the non-reducing and reducing (E-D, B-A) glycosidic linkages significantly differ from the previous. This result is expected considering that the last two glycosidic links include the two terminal residues, and BA has (1→2) instead of (1→3) connectivity. Figure 44 clearly shows the decrease in dihedral angle conformational freedom in going from a non-sulfated pentasaccharide to a polysulfated analogue, the latter being characterized by a

stiffer glycan backbone (compare the right and left columns in Figure 44). In fact, all glycosidic linkages show a significantly narrower φ/ψ distribution and higher gradient of colors (from the red to the blue shadows) in the polysulfated pentasaccharide (compound 7). The E-D glycosidic junction of compound 6 shows a wider φ/ψ conformational space, possibly split in three different regions approximately centered at: -30/50, -40/20, -40/-20, considering the continuity of the yellow/red shadows of the colored bins (Figure 44, E-D panel, right and left columns). A deeper conformational analysis of this glycosidic bond is required to confirm this potential energy minimum. The predicted linear conformation and selected inter-glycosidic distances of compound 7 and 6 are shown in Figure 46 right and left respectively, both molecules appear linear, even if compound 7 shows the stiffest conformation for the highest crowding. The selected inter-glycosidic distances for the polysulfated pentasaccharide are then compared in Table 18 and found qualitatively to be in accord with the corresponding distances estimated by NOEs.

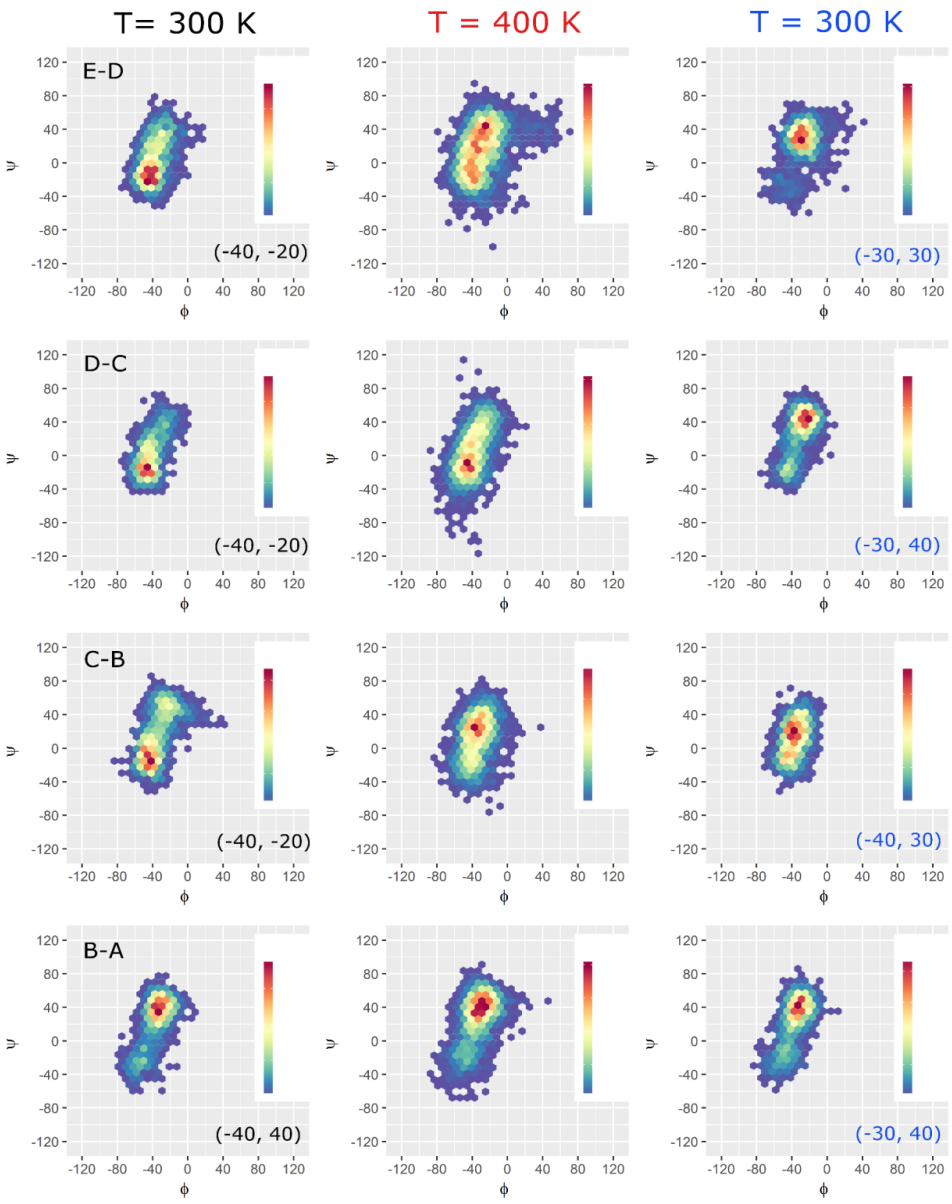


Figure 45. *Glycosidic backbone conformational changes of poly-sulfated phosphomannopentaose (compound 7) in three different steps of the MD simulation procedure in that order: 1st step at T = 300 K (1st column), 6th step at the highest temperature T = 400 K (2nd column), the 11th (final) step at T = 300 K (3th column). On each Ramachandran plots ϕ/ψ a 2D histogram binning with a density colour map is reported to underline glycosidic bonds conformational states in different thermal conditions.*

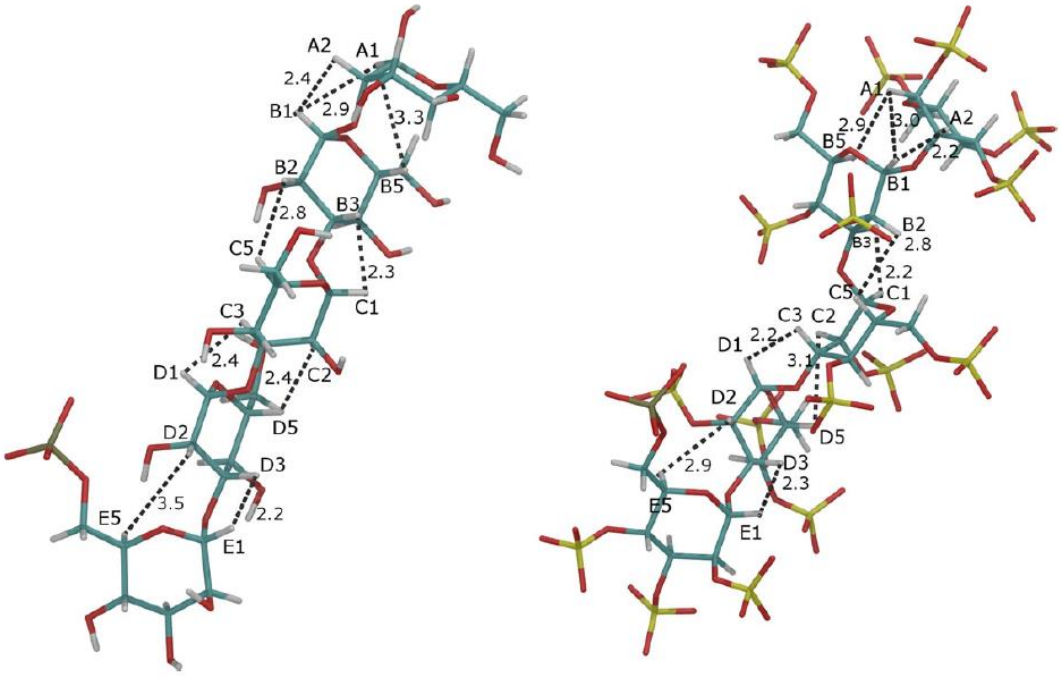


Figure 46. *Polysulfated phosphomannopentaose 7 (right) and the corresponding nonsulfated precursor 6 (left), structures and conformations. The interglycosidic distances matching the corresponding restraints are reported in Table 7. The Man-α residue labels are reported from the nonreducing to the reducing end (E to A).*

5.4. Conclusions

The major oligosaccharide constituents of PI-88 were prepared by sulfonation of individual purified phosphorylated oligosaccharides isolated from the PI-88 precursor and subjected to detailed 1D and 2D NMR spectroscopic analysis. The spectra of the individual components greatly assisted the assignment of minor resonances in the ^{1}H NMR spectrum of PI-88. The data also showed that most of the oligosaccharides in PI-88 are fully sulfated and that any under-sulfated species present are largely due to anomeric desulfation. These studies then allowed the full assignment of the ^{1}H and ^{13}C NMR spectra of PI-88. For the first time, the conformational characterization of the phosphomannopentaose sulfate with Man-α(1→3)-Man connectivities and terminated at its reducing

end by the Man-α(1→2)-Man glycosidic linkage is reported and compared with the corresponding unsulfated precursor. The NOE signals determined from the complex mixture, in which pentasaccharide 7 is prevalent, show correlations across each glycosidic linkage, involving different pairs of protons between internal and reducing end residues. The $^4C_1$ chair conformation of each Man residue, already described for unsulfated Man-α(1→3)-Man oligosaccharides[87], appears to be unaffected by sulfation as supported by the series of inter-glycosidic NOEs proton pair correlations observed, even if a weak distortion of the $^4C_1$ conformation appears correlated with the high degree of sulfation of compound 7, in comparison with the corresponding unsulfated precursor 6. MD simulations in implicit solvent and a 2D histogram binning procedure allowed the sampling and qualitative prediction of the conformation of Man-α(1 →3)-Man and the reducing end Man-α(1→2)-Man glycosidic junctions (Ramachandran plots) in both polysulfated and unsulfated pentasaccharides. In particular, the glycosidic conformation predicted by the ϕ/ψ 2D density maps of the former compound qualitatively matched the interglycosidic NOE restraints. Comparison of the Ramachandran plots of polysulfated and unsulfated pentasaccharides suggests a decrease in conformational freedom (increase in backbone stiffness) upon sulfation at each glycosidic junction. These results predict a linear and rigid backbone conformation for the longest, and most abundant polysulfated oligosaccharides in the PI-88 mixture, and a similar, but less stiff conformation for their corresponding unsulfated precursors, providing useful insights to understand the binding interactions of PI-88 with target proteins.

6. Practical Application of Chemoenzymatic Synthesis

6.1. Background

As discussed in the introduction section the high diversity and complexity of HS represent an obstacle for structural characterization and biological activity studies. The possibility of obtaining structurally defined oligosaccharides is a crucial target for the successful development of heparan sulfate structure –activity relationship. The synthesis series of structurally defined HS oligosaccharides with high purity is generally required for high-affinity interactions with heparin-binding proteins. There are two main approaches for the target driven synthesis of a library of structurally defined HS and heparin oligosaccharides. The first approach is based on a purely chemical synthesis method, involving repetitive steps of protection, activation, coupling and de-protection, and is very challenging for the preparation of oligosaccharides larger than pentasaccharides. For example, Arixtra®, an FDA-approved pentasaccharide anticoagulant drug, requires as many as 60 chemical steps, and is produced in a reported overall yield of only 0.1%[34]. In addition, during the preparation, there are many separation steps needed to remove undesirable isomers (especially the incorrect anomers) as well as the unavoidable side products that increase the cost and decrease the overall yield. The second approach, the chemoenzymatic synthesis, relies on combined chemical and enzymatic methods, mimicking the biosynthetic pathway of heparin, and represents a promising strategy to address many of these synthetic challenges. The enzymatic component of synthesis is usually performed under mild conditions (20–37 °C) using aqueous "green" solvents catalyzing glycosylation reactions with high stereoselectivity (α- or β-glycosidic linkages) and sulfonation reactions with excellent

regioselectivity, without the need for repetitive protection and deprotection steps. Advantages in recombinant enzymes production and the possibility of starting from an inexpensive commercially available acceptor, easily ultraviolet detectable, allow chemoenzymatic synthesis to become a flagship method of obtaining HS oligosaccharides, quickly and in high overall yield. Furthermore, synthetic HS could be engineered to introduce additional beneficial pharmacological properties that do not exist in HS from natural sources. The oligosaccharides analyzed in the previous chapters were fully obtained by chemoenzymatic synthesis. Here, we report this technique to obtain new endo-β-D-glucuronidase-binding compounds that will contribute to a structure activity relationship of a set of HS like oligosaccharides, able to bind and inhibit heparanase. Heparanase is an enzyme capable of cleaving heparan sulfate (HS) at the cell-surface and within the extracellular matrix (ECM), through β-(1-4) glycosidic bond hydrolysis in –GlcUA- β-(1-4)-GlcNX linkages, and thereby contribute to remodeling of the ECM, and in the regulation of manifold cellular processes. The over-expression of heparanase has been linked to numerous malignancies, and is associated with the release and activation of growth factors, facilitating cell migration through intact capillary walls into the tissues and promoting angiogenesis. Inhibition of this enzyme has generated significant interest to design potential anti-metastatic drugs. During the last decades several classes of heparanase inhibitors have been proposed, spanning from nucleic acids, proteins, monoclonal antibodies, polysulfated saccharides to small-molecule; all of these compounds act as competitive inhibitors of this enzyme, mimicking the substrate molecule (HS) in the molecular recognition event. However, only a few of them reached clinical trials and the majority of them belong to the class of polysaccharides, obtained through semisynthetic methods or total synthesis[88].

6.2. Enzymes involved in chemoenzymatic synthesis

The chemoenzymatic synthesis of HS requires several classes of biosynthetic enzymes, some comprised of multiple isoforms. Glycosyltransferases are used to build the HS backbone structure consisting of a disaccharide repeating unit of GlcA and N-acetylglucosamine (GlcNAc). An epimerase (C5-epi) is responsible for converting a GlcA residue to an IdoA residue. Sulfo groups are transferred from 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to oligosaccharide by specialized HS sulfotransferases. For example, 2-O-sulfotransferase (2-OST) specifically transfers a sulfo group to the 2-position of an IdoA residue and, to lesser extent, to the 2-position of a GlcA residue. The 6-O-sulfotransferase (6-OST) and the 3-O-sulfotransferase (3-OST) transfer sulfo groups to the 6-position and 3-position of glucosamine residues. N-deacetylase/N-sulfotransferase (NDST) converts a GlcNAc residue to a N-sulfoglucosamine (GlcNS) residue. NDST is unique displaying two activities, N-deacetylase activity removing N-acetyl group from a GlcNAc residue, and N-sulfotrasnferase activity transferring a sulfo group to a GlcN residue, forming a GlcNS residue.

Currently all the enzymes, that are predominantly mammalian proteins, are expressed with high yield in *E. coli*, with the exception of NDST, EXT-1 and EXT-2. Only the *N*-sulfotransferase (NST) domain of the NDST can be effectively expressed in *E. coli* and, thus, this protein lacks *N*-deacetylase activity. To build the oligosaccharide backbone EXT-1 and EXT-2 are substituted by two bacterial glycosyltransferases known as KfiA (an *N*-acetyl glucosaminyl transferase from *E. coli* K5 strain) and pmHS2 (heparosan synthase 2 from *Pasteurella multicida*).

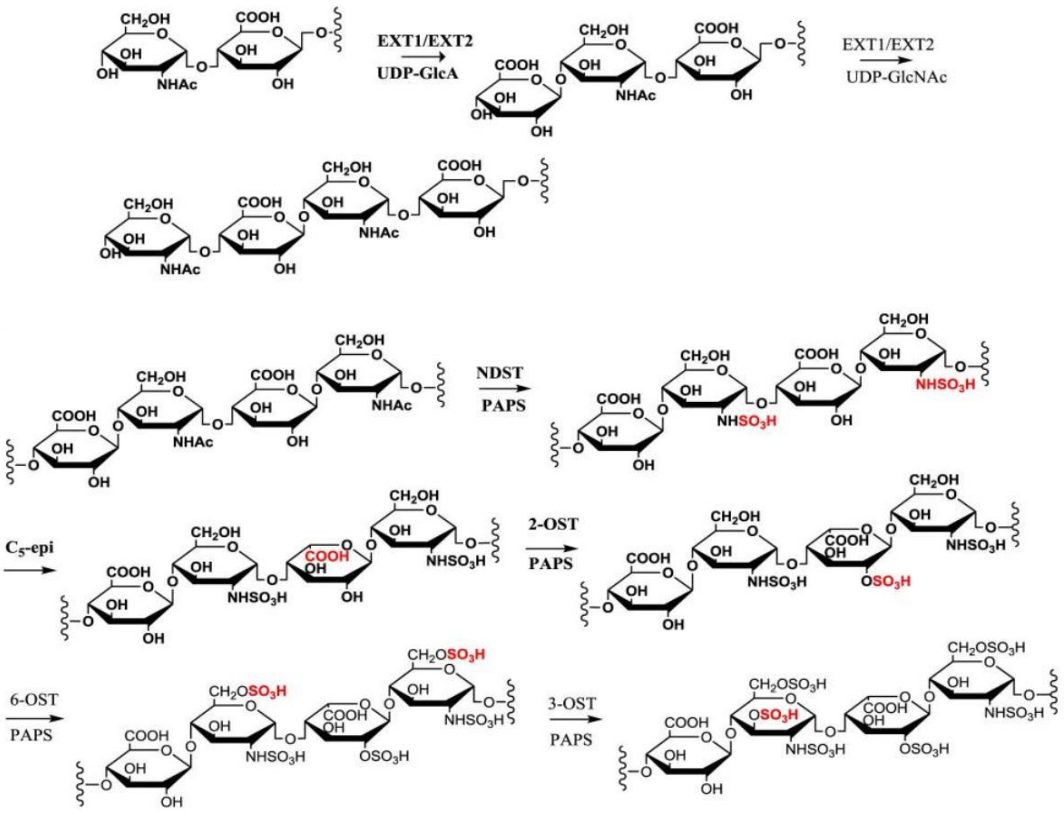


Figura 47. *Enzymes involved in the biosynthesis of HS. Abbreviations: EXT1; exostosin glycosyltransferase 1; EXT1; exostosin glycosyltransferase 2; NDST, N-deacetylase/N-sulfotransferase; C_5-epi, C_5-epimerase; 2-OST, 2-O-sulfotransferase; 6-OST, 6-O-sulfotransferase; 3-OST, 3-O-sulfotransferase; PAPS, 3′-phosphoadenosine 5′-phosphosulfate.*

6.3. Design of cascade enzyme modifications

To synthesize our endo-β-D-glucuronidase inhibitor we followed the synthetic route used by Liu and coworkers to synthesize structurally homogeneous ultra-low molecular weight heparins, low-molecular weight heparins and HS oligosaccharides[50].

Chemoenzymatic synthesis of HS requires a precise cascade of enzymatic modifications. One the most important steps for the success of all enzymatic modifications is the *N*-sulfation of glucosamine. The synthesis of the HS backbone containing GlcNS residues represented a major roadblock due to the lack of efficient recombinant NDST. To overcome this problem an un-

natural sugar nucleotide, UDP-GlcNTFA (uridine diphosphate *N*-trifluoroacetylated glucosamine), was introduced. For HS backbone synthesis, UDP-GlcNTFA appears to be an excellent substrate for KfiA with nearly quantitative conversion in the preparation of heparin oligosaccharide backbone. After the elongation, GlcNTFA residues are then converted to a GlcNS residues by removing the trifluoroacetyl group under mild alkaline conditions followed by *N*-sulfotransferase modification [Figure 48]. The use of UDP-GlcNTFA provides a means to introduce a GlcNS at precisely the desired position within an oligosaccharide.

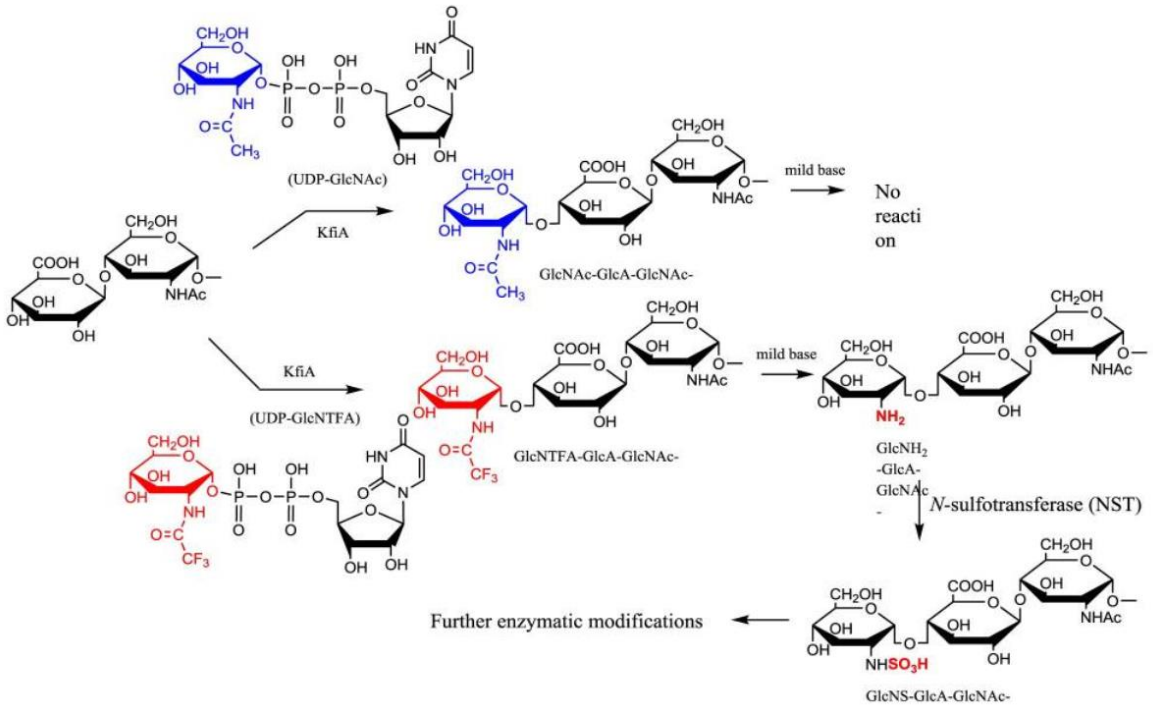


Figura 48. *HS backbone is elongated using a non-natural sugar nucleotide. (UDP-GlcNTFA). The trifluoroacetyl group of sugar nucleotide donor can be removed under mild base condition to form an N-unsubstituted glucosamine (GlcN) residue.*

In vivo, a HS biosynthetic pathway is responsible for attaching the multiple layers of sulfation on the polysaccharide backbone by following a specific order. The cells probably utilize the modification sequence to achieve the control of HS polysaccharide structure. HS biosynthetic enzymes were evolved to only react with selected substrates having a unique sulfation patterns,

while other saccharide substrates lacking these sulfation patterns remain unmodified. Such a feature assures that the biosynthesis of HS follows a preprogrammed modification sequence (For example, 2-*O*-sulfation step catalyzed by 2-OST occurs only after *N*-sulfation step, but prior to the 6-*O*-sulfation step).

6.4. IdoA2S units as enzyme-cleavage blocker

In heparin, more than 80% of the disaccharides are composed of -IdoA2S-GlcNS6S. The synthesis of a domain with the repeating –IdoA2S-GlcNS- disaccharide units is critically important in our work for preparing high sulfation HS-like non-cleavable motifs (i.e. -IdoA2S-GlcNS6S- linkage cannot be cleaved by heparanase).

The synthesis of repeating –IdoA2S-GlcNS6S units is harder due to a unique property of C_5-epi. Unlike other HS biosynthetic enzymes, C_5-epi is a two-way catalyst: it transforms a GlcA to an IdoA residue, but also converts IdoA to a GlcA residue, so the final product will be a mixture of GlcA and IdoA after C_5-epi action. Recent studies suggested that C_5-epi acts as a one-way catalyst within a unique distribution of GlcNAc and GlcNS residues[89]. These results show that C_5-epi can display both two-way and one-way catalytic modes by recognizing a sequence motif designated as the C_5-epi recognition pentasaccharide domain[89]. This pentasaccharide domain consists of an epimerization site (EPS) and a mode of reaction recognition site (MRRS). The MRRS is located three residues away from the non-reducing end side of the EPS. The *N*-substitution status of the MRRS residue, controlled in biosynthesis by NDST, dictates the mode of action of C_5-epi. A GlcNAc residue directs C_5-epi to display an irreversible reaction mode, while a GlcNS or GlcN residue (or the MRRS site being unoccupied) drives C_5-epi to display a reversible reaction mode.

The presence of GlcNS at the -1 residue is essential for rendering the susceptibility to C_5-epi modification. The 2-*O*-sulfation reaction, acts by 2-OST, allows to form an IdoA2S residue. This step is essential because it prevents further reversing action of C_5-epi on this residue and accelerates the C_5-epimerization reaction.

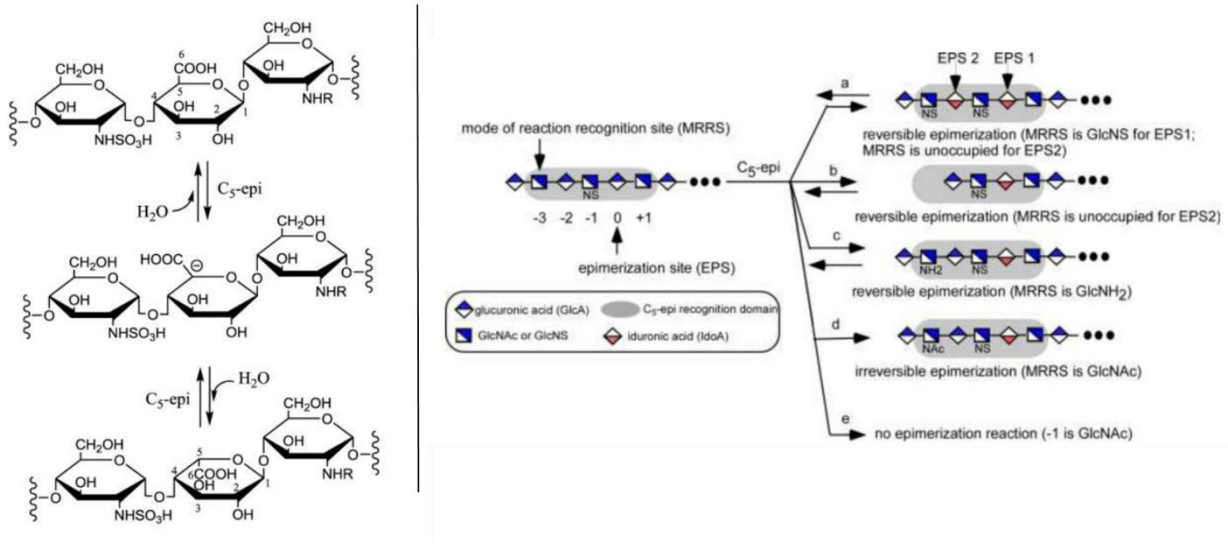


Figure 49. C_5-epi catalyzes the reversible conversion of epimers GlcA and IdoA. The designed epimerization site (EPS) is at the residue 0. The residue -1 must be a GlcNS residue to serve as an ESP, if not the site is not reactive. If the MRRS site (-3) is a GlcNS or GlcN residues or unoccupied, C_5-epi displays a reversible action while if a GlcNAc residue is at the MRRS site the reaction is irreversible.

6.5. Synthesis of hexasaccharide (Compound 1)

The elongation of compound 1 started from a previously chemically synthetized trisaccharide – GlcA-GlcNTFA-X- in which X is the residue directly involved in the binding.

In the current work, we applied a chemical-chemoenzymatic mixed approach. The first step is to assemble the oligosaccharide backbone by enzyme-catalyzed glycosylation and then we used a used a pure chemical approach to overcome all the issues coming from a non-precise recognition by some enzymes of the residue X.

The trisaccharide was synthetized using a pure chemical approach because the X group is completely different from the GlcA-pnp usually used as primer for chemoenzymatic synthesis. The chemical synthesis required a large number of steps with a very low yield. We decided to further elongate up to trisaccharide to avoid the possibility that glycosyltransferase PmHS2 does not recognize it as specific substrate. The limited availability of the starting material was the first problem to overcome in our synthetic route. Fortunately, using the reported conditions for heparosan synthase 2 from *Pasteurella multocida* (pmHS2), tris (hydroxymethyl) amino methane (TRIS) buffer pH 7.2 and $MnCl_2$, only trace amounts of trisaccharide starting material were not converted after overnight incubation time at 37°C. The trisaccharide was elongated using pmHS2 with donor UDP-GlcA to form a tetrasaccharide in 85% yield. It is noteworthy that the glycosylation yield increases as the elongation of the carbohydrate chain proceeds. This is probably because the elongated acceptors have a closer structural similarity to natural substrate structures. The elongation proceeded adding additional GlcNTFA and GlcA moieties to reach an hexasaccharides backbone structure.

After single elongation step, a following purification step was required because the saccharide-donors UDP-GlcA and UDP-GlcNTFA act itself as a building block in a chain growth polymerization reaction with PmHS2. The purification approach was based on the use of BioGel P2 resin, extra fine polyacrylamide beads for size exclusion chromatography with 100–1800 MW fractionation range. Usually in the chemoenzymatic method the purifications with BioGel P2 resin were conducted in basic condition using ddH_2O and NH_4HCO_3 0.1 M as solvent. Due to the presence, in the –X- moiety, of a susceptible functional group in basic environment we decided to conduct the purification in acidic condition (ddH_2O and TFA 0.1 M). These purification steps were conducted at very low flow rate (125µl min^{-1}) in order to better separate products nwith

comparable molecular weight. With hexasacchairde in hand, de-*N*-trifuoroacetylation was first attempted using Li(OH) in the water system at 4°C. However, these reaction conditions were too harsh, generating the cleavage of the intraglycosidic bond of –X- unit. A mixture of Methanol, water and Triethylamine (v/v/v 2:2:1) was subsequently used and successfully afforded the hexasaccharide having three free amino group, in >85% yield.

After NTFA group deprotection, enzymatic *N*-sulfonation catalyzed by *N*-sulfotransferase and PAPS was attempted on hexasaccharide. Unfortunately, using the reported conditions the desidered N-sulfated hexasaccharide was not obtained. We speculate that the difference of the reducing-end part of the backbone could affect the molecular recognizing of NST for its substrate cutting out the possibility to introduce new sulfo group on desired positions.

6.6. A HPLC based approach for rapid access to pure HS-like oligosaccharides

Gel filtration chromatography relying in Bio-gel P2 columns represents an excellent way of removing salts but is inefficient for oligosaccharide purification, since reaction byproducts, often introduced through incomplete glycosylation, have molecular weights similar to the target compounds. The introduction of a UV-detectable para-nitrophenol tag directly linked to NH_2 group of non-reducing end Glucosamine allowed access to easy monitor the reaction using simple LC methods. An HPLC based approach, previously developed, for rapid access to pure HS-like oligosaccharides was used. The elongated products of the HS backbone prior to sulfation reactions were purified using a C-18 column, while all the sulfated products of the oligosaccharides were purified using Q-Sepharose fast flow column.

6.7. Synthesis of octasaccharide (Compound 2)

UV-tagged trisaccharide was used as starting material for the backbone elongation. In this case five glycosylation steps were conducted in order to obtain a longer chain. As previously seen in heparin-Antithrombin molecular recognition, we thought that longer chains correspond to greater binding affinity with the enzyme, figuring out stronger inhibition activity. The followed protocol was the same used for compound 1 but the presence of UV-tag allows us to quickly move further. HPLC separation were conducted after any elongation step. In this way we were able to perform glycosylation with an overnight reaction and purify the day after maximizing the efficiency of the process. De-*N*-trifuoroacetylation step was conducted in the same conditions for compound 1 to avoid degradation of the –X- unit. Due to the failure with the N-sulfation reaction conducted with NST enzyme we decided to use a pure chemical approach to introduce a sulfo group on our compound (octasaccharide). Py-SO_3 in mild basic condition ($NaHCO_3$) was used as donor of a sulfo group. The lability of this salt was the road block in this reaction. To overcome this issue, we decided to add Py-SO_3 in portions in order to let salt in condition to transfer all the sulfo groups and avoid its degradation over time. This approach allows us to obtain more than 75% yield. The possibility to modify experimental conditions could let us to significantly increase the yield. Unfortunately, the low amount of starting material was the bottle-neck in the set-up of all the reaction scheme. The last modification step in this synthetic scheme was the conversion and consequentially the 2-O sufation of GlcA residues in order to obtain as more as possible IdoA2S units to avoid the breakdown of oligosaccharides by the action of the heparanase. The treatment of N-sulfated octasaccharide with C_5-epimerase and 2-OST afforded IdoA2Ss-containing product 2. This specific reaction showed the lowest yield (<30%) and did not give us the desired product. The octasaccharide previously synthetized contains three potential sites that could be recognized

by the C_5-epi even if, that at the non-reducing end has very low probability to be converted because does not fit the MRRS of the enzyme. In such condition only two sites are eligible to be epimerized. Due to the reversible action of C_5-epi and the 2-O sulfation could happens, at high concentration of PAPS, also in GlcA residues, the number of by-products increased. For this reason, we obtained a reaction mix with a several number of by-products that actually are not well purified. The strategy that was followed in the further tests to convert GlcA in IdoA, should be a step by step procedure in order to minimize the number of by-products and better purify the reaction mix to obtain highly pure intermediate for further reactions.

6.8. Conclusion

In this study, we described the chemoenzymatic synthesis developed by Liu at UNC to obtain endo-β-D-glucuronidase inhibitor/HS analogues through a chemoenzymatic approach. Enzyme-based polymerization was employed to synthesize oligosaccharide backbones with controlled size and charge distributions. A well-defined sulfation patterns were obtained through a chemical and enzymatic approach. In summary, we present herein a concise chemical-chemoenzymatic mixed approach for the set-up of a forthcoming robust synthesis of HS analogues as potential endo-β-D-glucuronidase inhibitors.

7. General Conclusions

This work provides new insights into the detailed description of the synthesis and the conformation of chemoenzimatically obtained oligosaccharides in free solution and upon binding with Antitrhombin.

The aim of the first part of the project was the characterization of the dynamic and conformation of two hexasaccharides (hexa-4 and hexa-8) in solution. Detail study involving NMR, MD and numerical analysis of the data generated provided information regarding the conformation of the hexasaccharides and also their dynamic properties during the binding with Antithrombin. For both hexasaccharides, experimental and simulated data indicated a similar glycan binding epitope extending from the non-reducing end GlcNS6S to the reducing end GlcNS6S''-GlcA'' residues. In particular, the two central GlcNS3S6S and IdoA2S/IdoA residues showed the stronger interaction with AT for both glycans, whereas the reducing end GlcA'' appeared less involved in binding with AT. IdoA2S and IdoA remained in the $^{2}S_{0}$ conformation during the MD simulations in the cases both of hexa-4·AT and of hexa-8·AT, independently from the 2-O sulfation state, thus supporting the idea that the HBS surface enforces the $^{1}C_{4}$ to $^{2}S_{0}$ transition upon binding. In this work we found that the residues GlcNS3S6S, IdoA2S/IdoA and, to a lesser extent, the reducing end GlcNS6S'' preserve the contacts previously observed in the reference NMR and X-ray diffraction AGA*IA·AT complexes. These results also confirm the significant role of the contact network between AT HBS and residues surrounding the trisaccharide sequence GlcNS3S6S-IdoA2S/IdoA-GlcNS6S'' in changing the IdoA conformation from $^{1}C_{4}$ to $^{2}S_{0}$ upon binding, independently from the 2-O sulfation state of IdoA.

In the second part of the work the investigation was focused on the characterization of the interaction between a novel chemoenzimatically synthesized octasaccharide with a non-natural sequence and Antithrombin. The new artificial heparin-like octasaccharide was characterized by a central pentasaccharide sequence variant: GlcNS6S-GlcA-**GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S**-GlcA-$PhNO_2$, in which -IdoA2S-GlcNS3S6S- replaces the typical -GlcA-GlcNS3S6S- sequence of AGA*IA, known to be the shortest glycan that specifically bind and activate AT. NMR experiments like ^{1}H-STD and the complementary water-LOGSY supported by molecular dynamic, show that the AIA*IA central moiety presents the strongest contacts with the HBS of AT, while the strength decreasing both to non-reducing and reducing ends. The most relevant conformational change of octasaccharide upon binding involve the two IdoA2S, that change from mixed $^1C_4$ and $^2S_0$ forms, to a defined $^1C_4$ and $^2S_0$ conformation for the IdoA2S located in the non-reducing and reducing end parts, respectively. The modelling description supports the conformational change of the two IdoA2S residues in going from the un-bound to the bound state. The contact network between the AIA*IA moiety and the HBS of AT, forces IdoA2S in the non-reducing and reducing end positions to assume the $^1C_4$ and $^2S_0$ conformation, respectively. These results are important because they contribute and refine the Structure Activity Relationship description of the still growing set of heparin-like oligosaccharides able to bind and activate AT.

During the last part of the study, the practical chemoenzymatic synthesis was studied and applied. Using the strains of engineered enzymes involved in HS biosynthesis we tried to obtain a HS-analogue that acts as endo-β-D-glucuronidase inhibitor. Enzyme-based polymerization was employed to synthesize oligosaccharide backbones with controlled size and charge distributions. A well-defined sulfation patterns were obtained through a chemical and enzymatic approach.

The future of developments of this project will regard the chemoenzymatic synthesis of new HS like molecules and characterize the conformation and topology of interaction with Antithrombin and other proteins to go deeper in the Structure Activity Relationship description. These efforts may lead to the development of new well-targeted and safer drugs.

8. Experimental Section

8.1. Hexasaccharides and Octasaccharide

8.1.1. Antithrombin Purification

Pasteurized Antithrombin (Kybernin P) was solubilized in 10 ml of buffer A (10 mM PBS, 0.25 M NaCl pH 7.4) and was loaded onto aXK 16/40 column (GE Healthcare, Italy) previously prepared with50 ml of Heparin Sepharose 6-Fast Flow and equilibrated at 4◦C in the same buffer. A low affinity fraction was eluted with 300 ml of buffer A, while the high affinity fraction was eluted with the same volume of PBS, 2.5 M NaCl pH 7.4 elution buffer. The fractions were loaded onto 3 K_d a MWCO filters (Millipore, Italy) in a stirred ultra-filtration cell (Millipore, Italy) for desalting and buffer exchanged, with 20 mM phosphate buffer pH 7.4. The final concentration of the solutions was determined using the bicinchoninic acid assay following the manufacturer's instruction.

8.1.2. ITC

The equilibrium dissociation constant (K_d) was assessed at 25°C in 10 mM PBS (pH 7.1) and 0.5 M NaCl. The K_d value was obtained by MicroCal ITC, monitoring the heat released during a biomolecular binding event and obtaining a direct measure of binding affinity and thermodynamics parameters. The sample cell was filled with 200 μL 6 10^{-6} M of protein solution while the syringe released 2 μL 1.4 10^{-6} M aliquot of ligand every 150 s to perform the titration. K_d were measured using a MicroCal PEAQ-ITC and ITC200 Malvern. The titration curve of kcal mol^{-1} vs molar ratio (ligand/sample) was obtained sampling the heat pulse, generated after every injection of 2 μL aliquot, integrated with respect to time and normalized for concentration. The K_d values were determined

fitting the thermodynamics data of the binding enthalpy (ΔH) during titration. From the fitting the free energy of binding (ΔG) and the variation of entropy (ΔS) could be estimated. Thermodynamics data were elaborated with MicroCal PEAQ-ITC Analysis Software. In this study the stoichiometry parameter *n* was set to 1 considering the ratio 1:1 glycan/AT.

8.1.3. NMR

NMR spectra were recorded using a Bruker Avance III spectrometer operating at 600.13 MHz equipped with a high-sensitivity 5 mm TCI cryoprobe at 303 K. Samples were lyophilized two times to remove residual solvents then were dissolved in 220 μL 10 mM phosphate buffer, 0.5 M NaCl (pH 7.4) and 3 mM EDTA in $^{2}H_2O$ (99.996%) and placed in 3-mm NMR tubes. For the experiments involving free ligands, the samples were prepared to obtain a final concentration of 1.2 and 1.0 10^{-3} M by dissolving 500 μg of hexasaccharides and octasaccharide respectively in phosphate buffer. Proton spectra were recorded with water presaturation with a recycle delay of 12 s and 16 scans. HSQC (heteronuclear single quantum coherence) experiments were performed in phase sensitivity enhanced pure-absorption mode using 16 dummy scans, 4 scans with decoupling during acquisition period and 2.5 s of relaxation delay. The matrix size of 2048×128 data points was increased to 4096×1024 by linear prediction and zero filling. 2D homonuclear correlation COSY experiment were obtained in phase sensitivity mode using 8 scans with water presaturation during acquisition. The matrix size of 2048x128 data points was zero filled to 2048x1024. Bidimensional TOCSY were acquired using 8 scans per increment of 1024×320 data points with zero filling in F1 (2048×1024). HSQC spectra were obtained in phase sensitivity enhanced pure-absorption mode with decoupling in the acquisition period. The matrix size of 2048×128 data points was zero filled to 4096×1024. For the tr-NOE experiments the samples were prepared by dissolving in phosphate buffer respectively

1.61 mg of AT and 250 μg of the hexasaccharide and 998μg of AT and 200 μg of octasaccharide reaching a molar ratio of 1:5 AT/oligosaccharides. All bidimensional NOESY and tr-NOESY experiments were performed at 303 K and 308 K respectively, in order to observe a suitable ligand exchange upon binding. A total of 16 scans were collected for each free induction decay (matrix 2048×320 points), the data were zero filled to 2048x1024 points before Fourier transformation and mixing time values of 200, 300 and 500 ms were used. ^{1}H-STD experiments involving oligoasaccharide·AT interaction were prepared so that the final protein concentration was 6.3 10^{-6} M with a ligand/AT molar ratio of 100:1. To remove broad resonances of the protein, the pulse sequence used for monodimensional ^{1}H-STD experiments includes a 10 ms spin-lock pulse. The on-resonance irradiation was performed at the high-field protein resonances (548 Hz) whereas the off-resonance control irradiation was performed at 644000 Hz. The ^{1}H-STD spectrum was obtained by phase cycling subtraction of the on-resonance and off-resonance data acquired in interleaved mode. The number of scans and dummy scans were 128 and 32 respectively with a recycle delay of 8 s and a saturation time respectively of 1, 1.5 and 3 s were used.

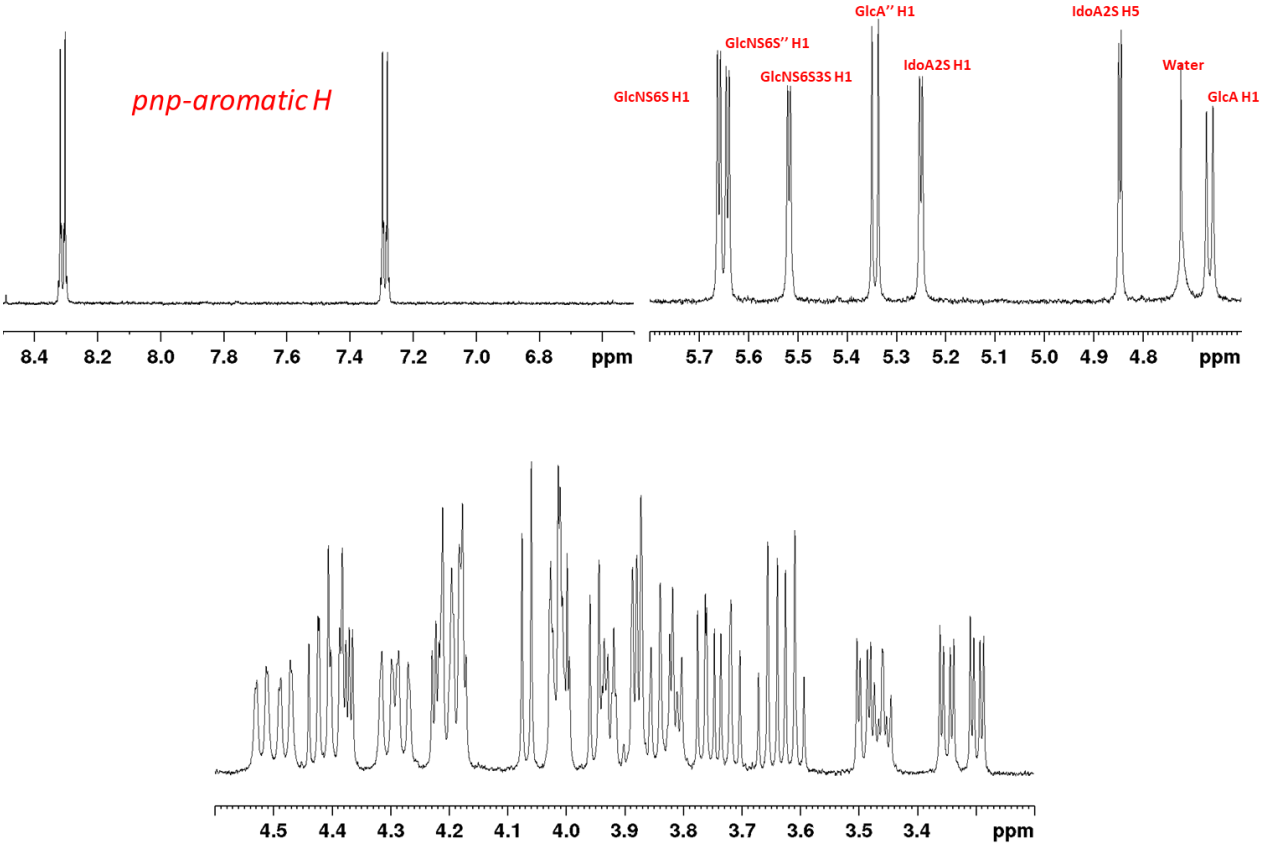


Figure Exp1. *Proton spectrum of the Hexa-8, where two different spectra intervals are highlighted: the aromatic and the anomeric region.*

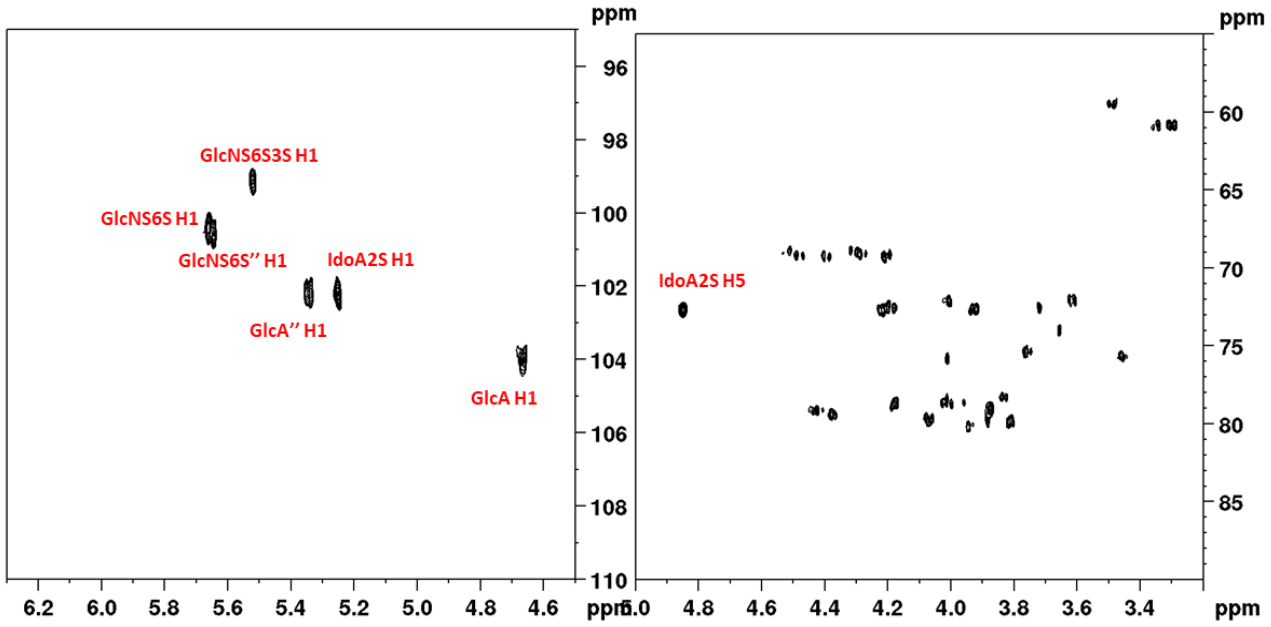


Figure Exp2. *HSQC spectrum of the Hexa-8, where the the anomeric region is highlighted.*

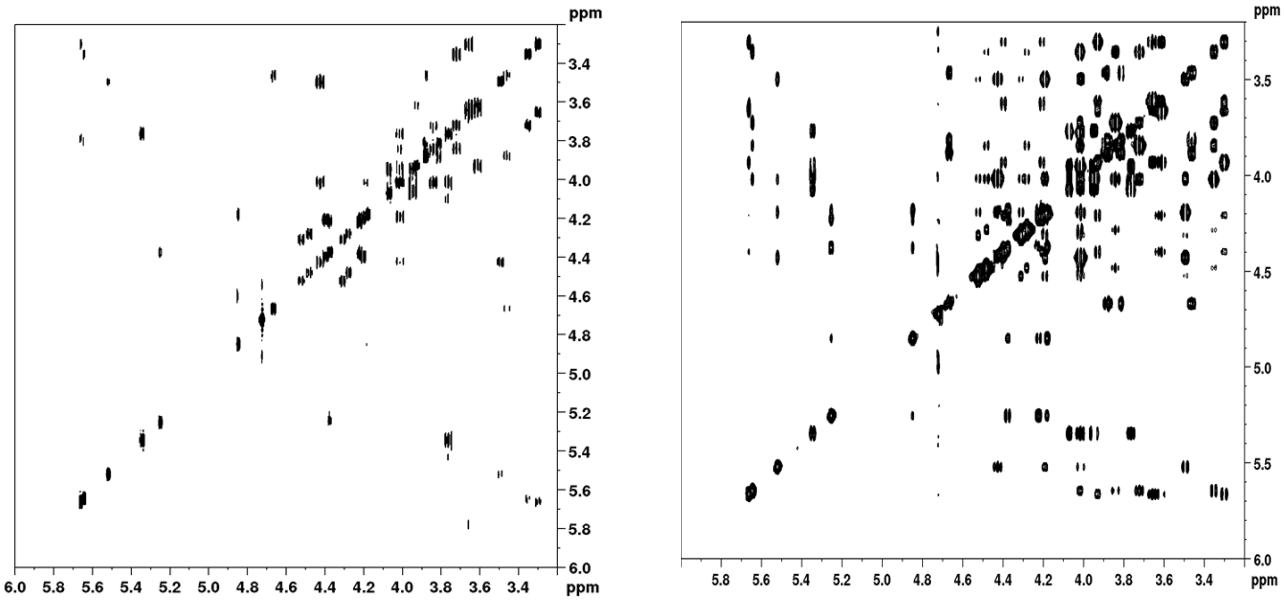


Figure Exp3. *COSY (left) and TOCSY (right) spectra of the Hexa-8.*

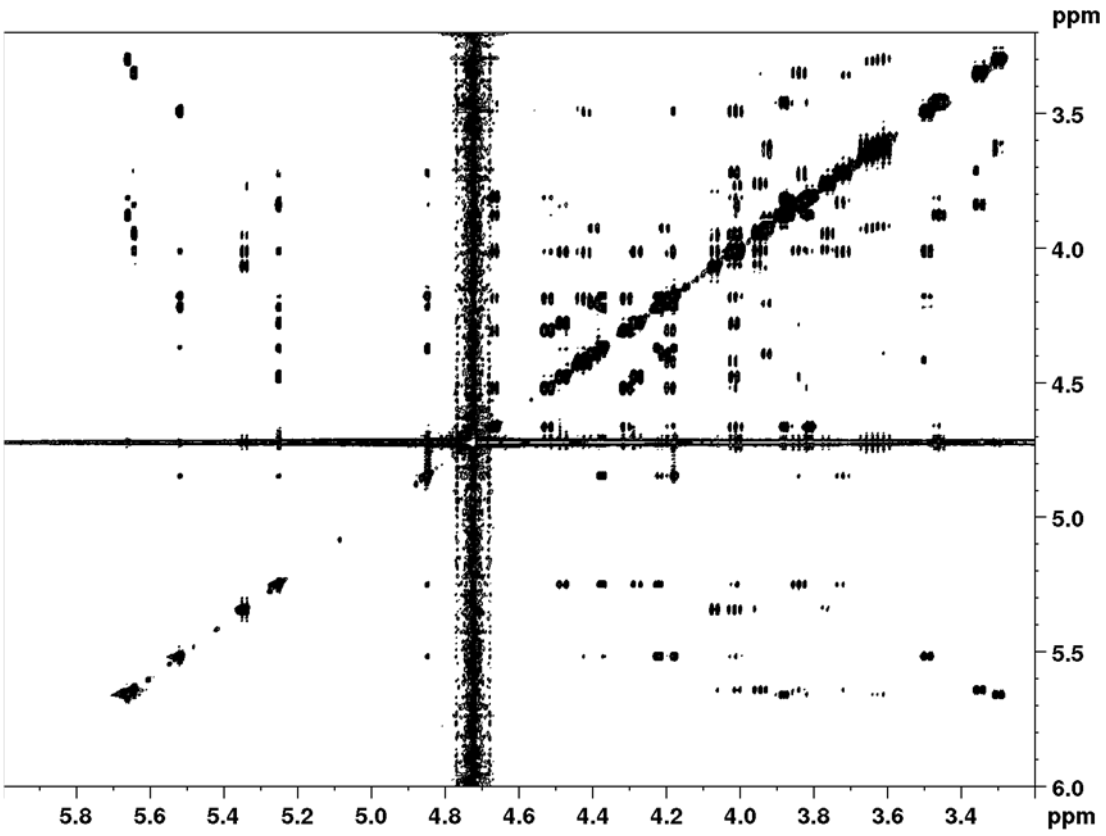


Figure Exp4. *NOE spectrum of the Hexa-8 at 500ms mixing time.*

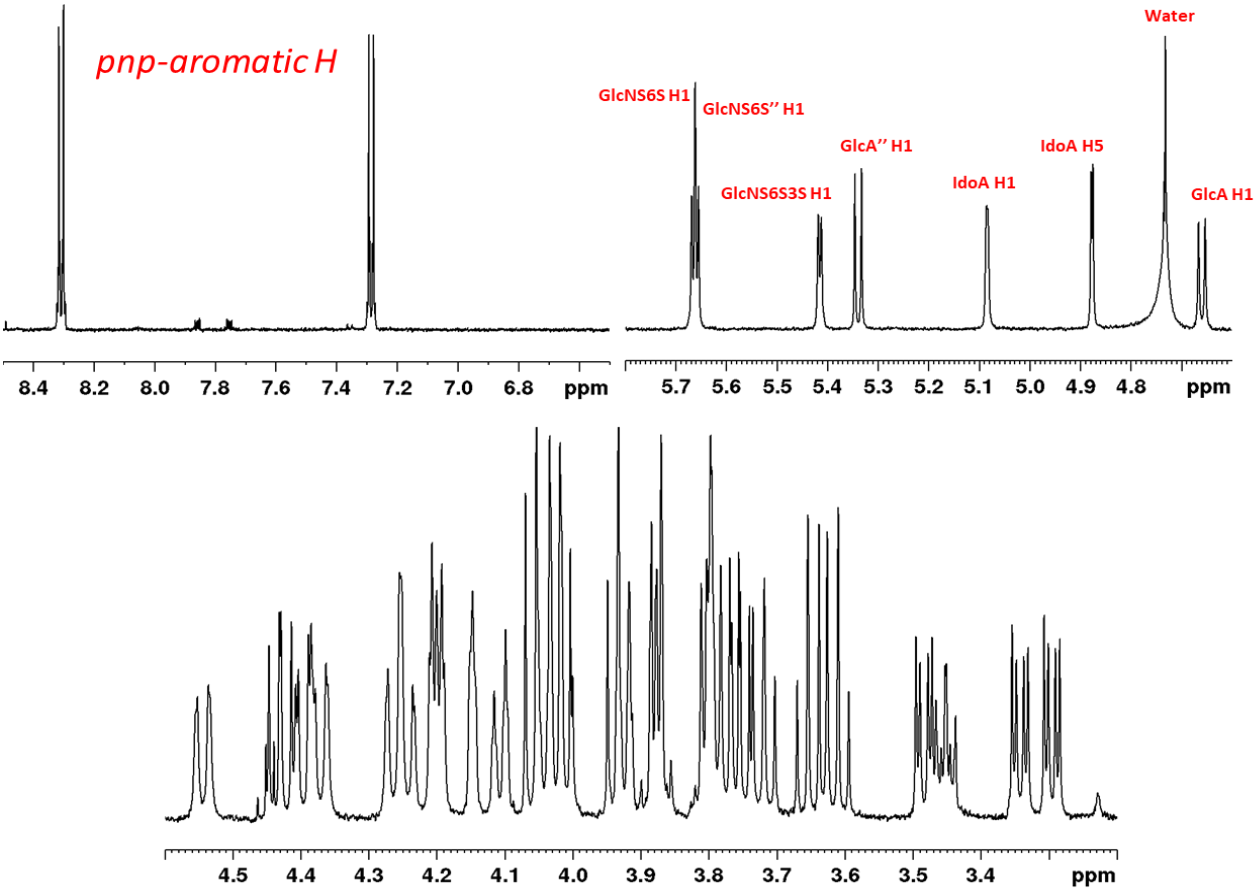


Figure Exp5. *Proton spectrum of the Hexa-4, where two different spectra intervals are highlighted: the aromatic and the anomeric region.*

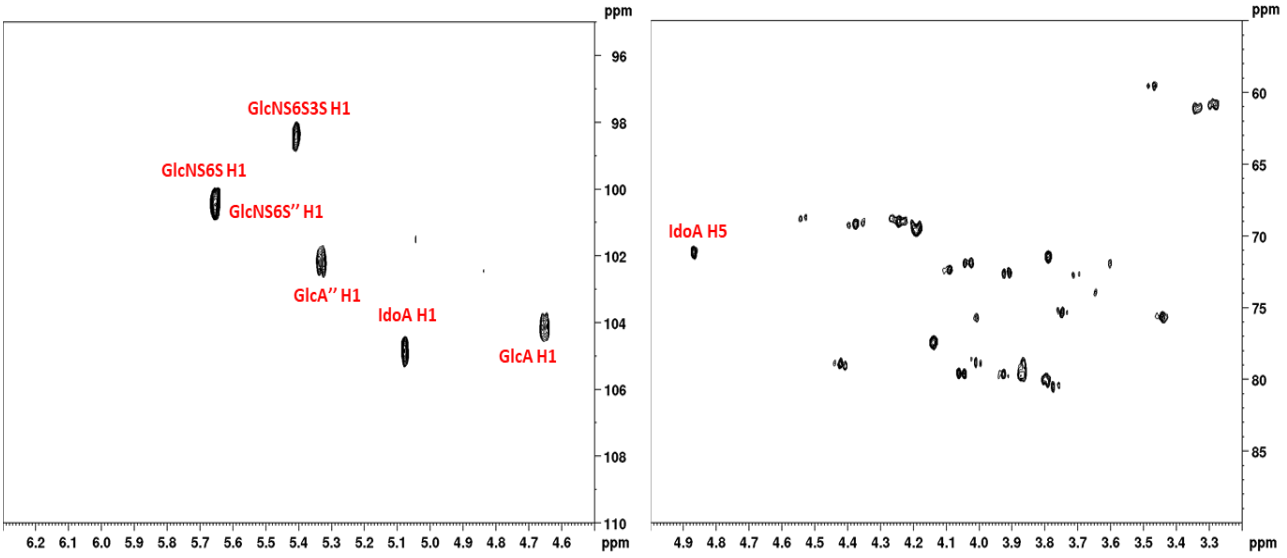


Figure Exp6. *HSQC spectrum of the Hexa-4 where the anomeric region is highlighted.*

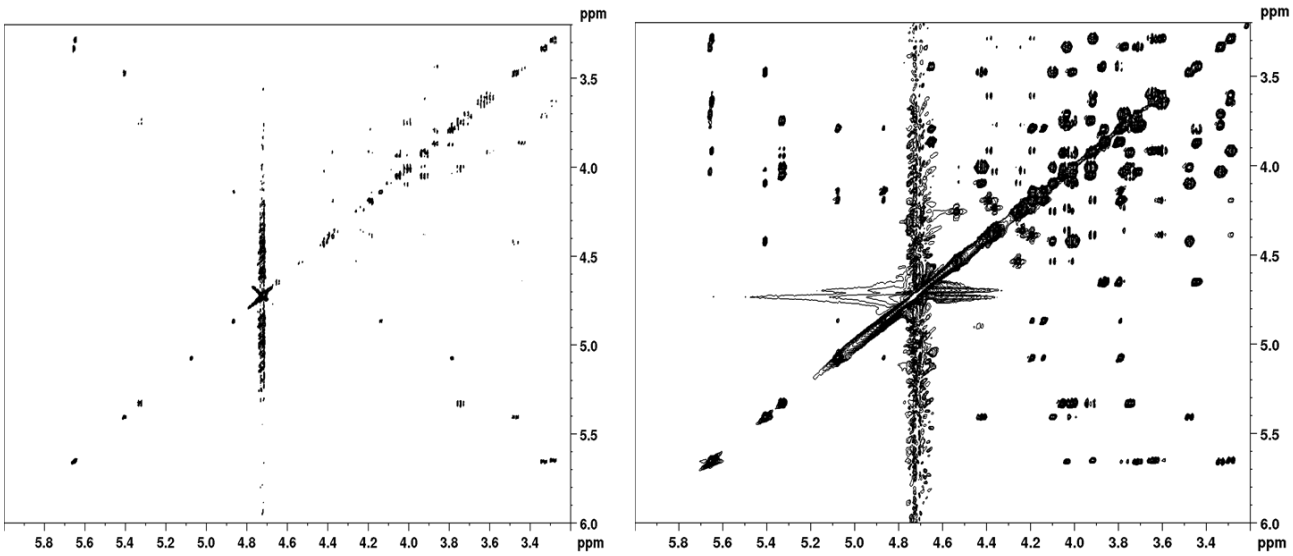


Figure Exp7. *COSY (left) and TOCSY (right) spectra of the Hexa-4.*

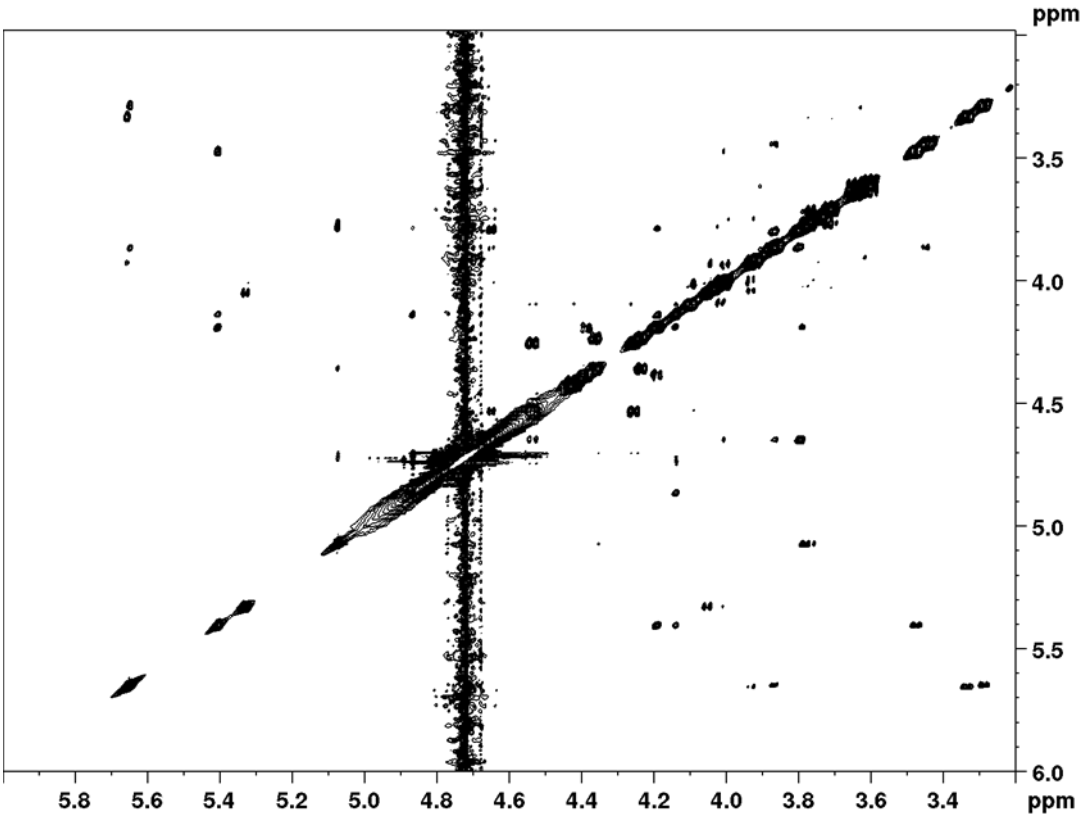


Figure Exp8. *NOE spectrum of the Hexa-4 at 500ms mixing time.*

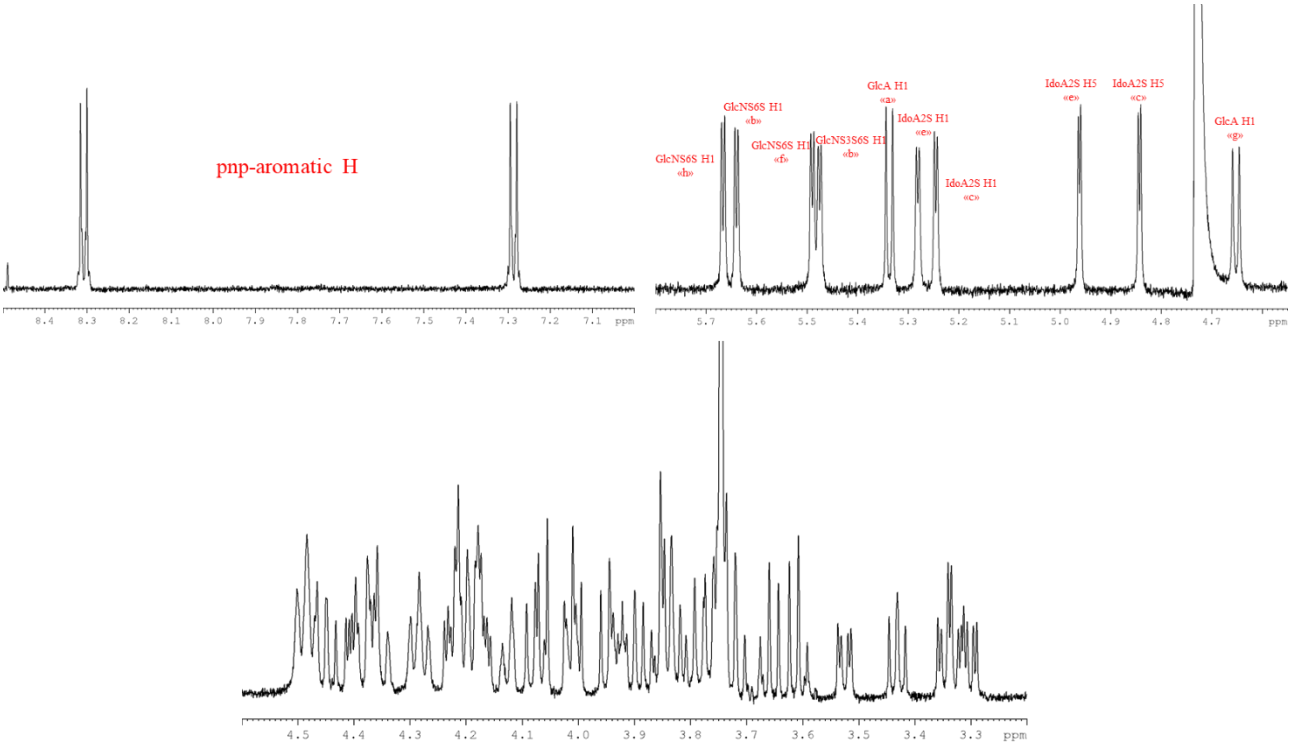


Figure Exp9. *Proton spectrum of the Octa-5, where two different intervals: the aromatic and the anomeric regions are highlighted.*

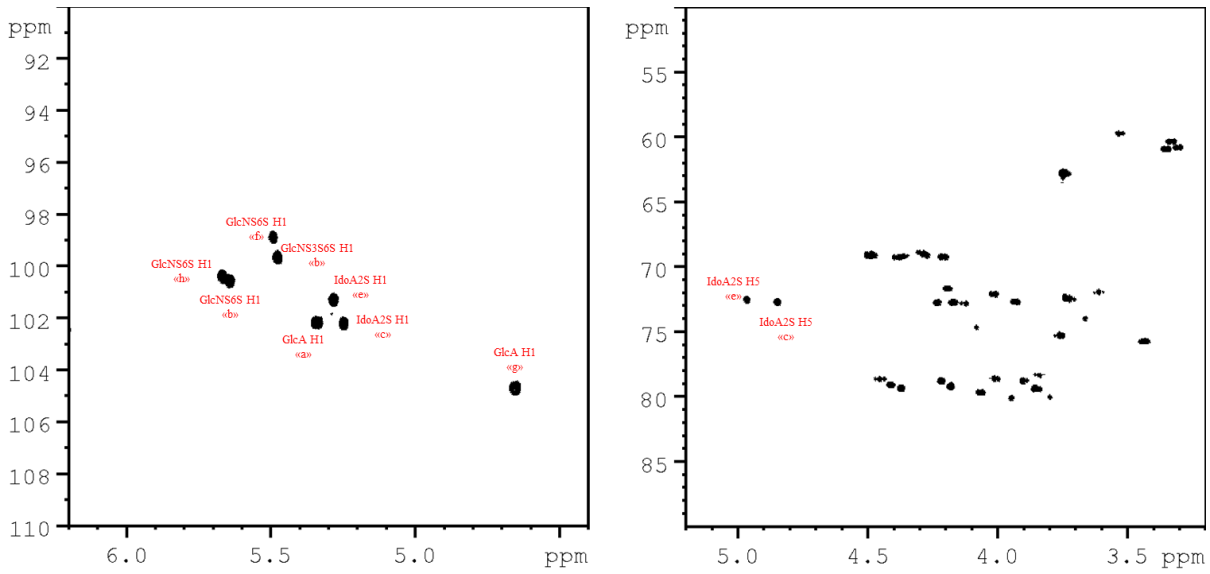


Figure Exp10. *HSQC spectrum of the Octa-5, where the anomeric region is highlighted.*

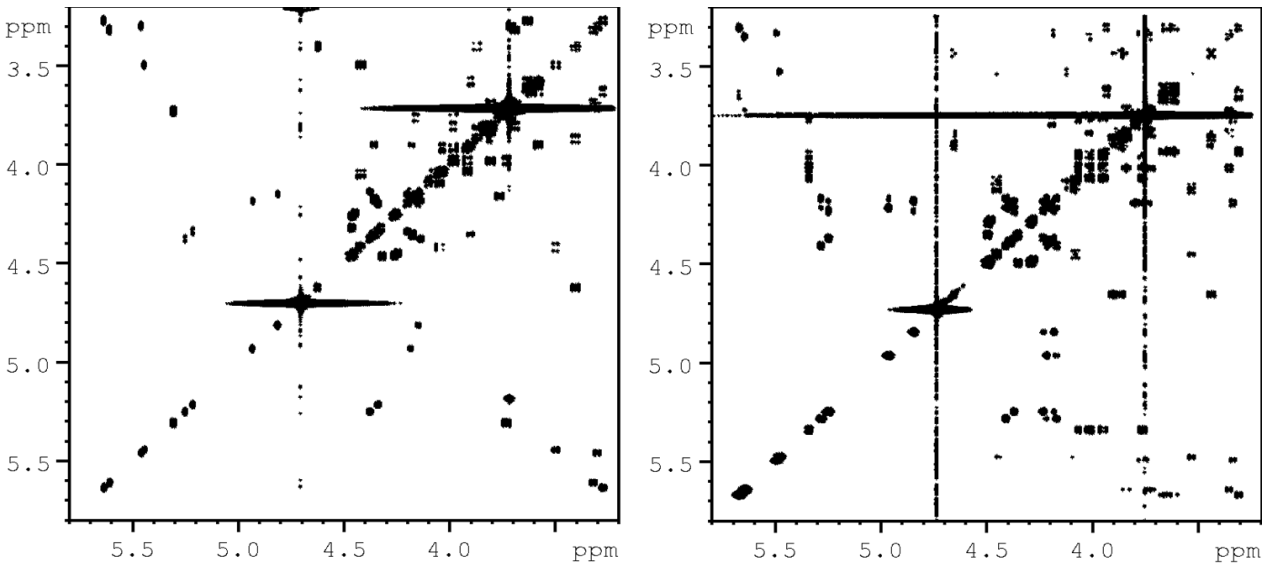


Figure Exp11. *COSY (left) and TOCSY (right) spectra of the Octa-5.*

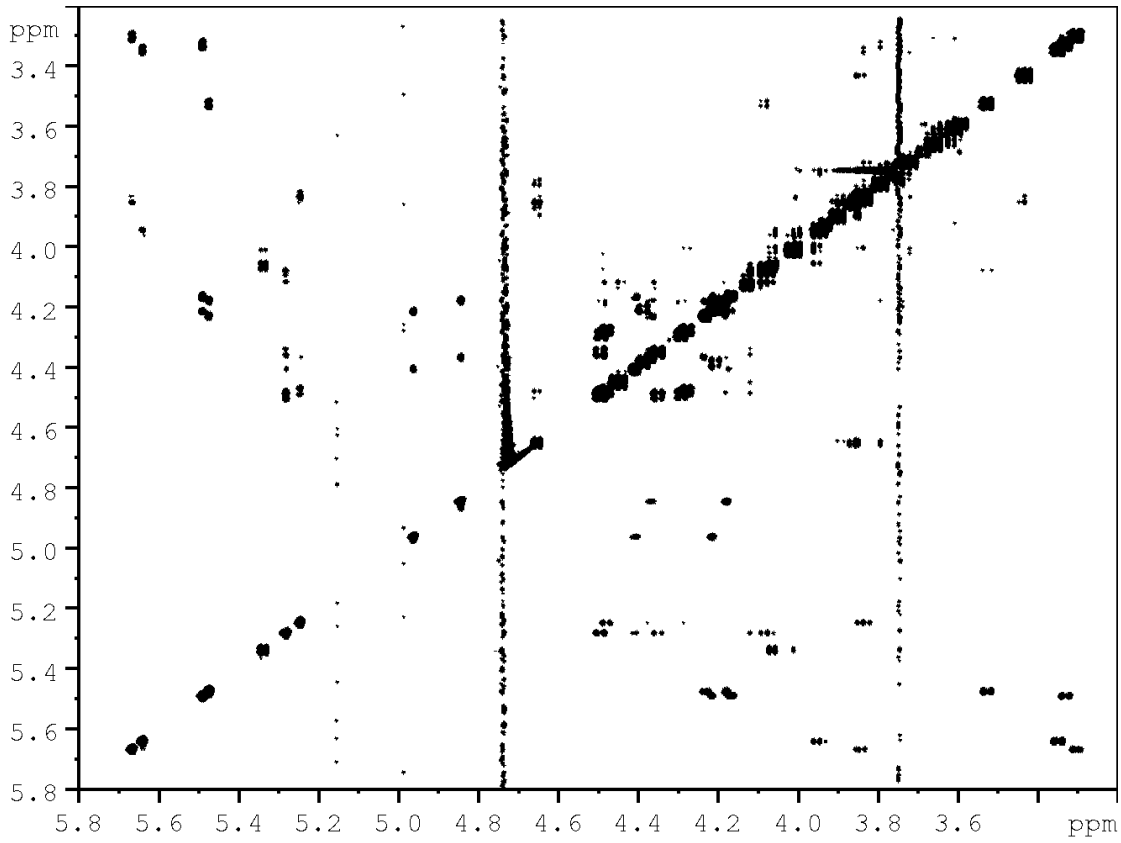


Figure Exp12. *NOE spectrum of the Octa-5 at 500ms mixing time.*

8.1.4. MD simulation

Atomic scale models for the Hexa-4 and Hexa-8 in free and bound state with AT were built, using the "state of the art" GLYCAM06 Force Field, while AT was described using the Amber Force Field. Explicit solvent TIP3P water model were used in all cases. The parameters of the terminal residue $PhNO_2$ at the reducing end of each glycans, were obtained from the Amber parameters and added to the GLYCAM06 Force Field. The partial charges of this residue were calculated after geometry optimization using the quanto-chemical software GAMESS ('General Atomic and Molecular Electronic Structure System'), and the HF/6-31G* level of theory, while the RESP (Restrained Electrostatic Potential) method was applied for charge determination as in GLYCAM06. Ambertools 1.4 was used to build the parameters and topology files for MD simulation. The initial conformation of the free state hexasaccharides was given in accord to previous studies on heparin like oligosaccharides. The initial dihedrals for the glycosidic linkages: GlcNS6S-GlcA, GlcA-GlcNS3S6S, GlcNS3S6S-IdoA2S/IdoA, IdoA2S/IdoA-GlcNS6S'', GlcNS6S''-GlcA'', and GlcA''-$PhNO_2$ were defined as -49/-14, 59/-6, -59/-26, 57/11, -59/-26, 50 in that order. The conformation of each residue was $^4C_1$ chair for GlcNS6S, GlcNS3S6S and GlcA, while the IdoA2S and IdoA were set in $^2S_0$ and $^1C_4$ conformation for Hexa-4 and Hexa-8 in free state. In fact, in accordance with Hsieh *et al.* and Ferro *et al.* Hexa-4 and Hexa-8 showed IdoA2S and IdoA conformation dominated by $^2S_0$ and $^1C_4$ respectively. Differently in bound state with AT: The IdoA2S and IdoA residues of Hexa-4 and Hexa-8 were set in $^2S_0$ conformation in agreement with NOEs tr-NOEs results. The AT model was built using the 3D structure of the L-chain (Latent AT) in PDB ID 1AZX. The specific pentasaccharide sequence AGA*IA defined the initial glycosidic linkage conformations and positions for Hexa-4 and Hexa-8 in bound state with AT. L-chain instead of I-chain (inhibitory AT) was used to model AT, because the initial idea was to reproduce the glycan/AT complex at early time instants

of the interaction. This was justified considering that contact between AGA*IA and AT RBS does not change significantly during AT conformational rearrangement. Before MD simulation the values of glycosidic linkage dihedrals in Hexa-4·AT and Hexa-8·AT: GlcNS6S-GlcA, GlcA-GlcNS3S6S, GlcNS3S6S-IdoA2S/IdoA, IdoA2S/IdoA-GlcNS6S'', GlcNS6S''-GlcA'', and GlcA''-$PhNO_2$ were set up -31/-35, 41/1, -71/-33, 44/16, -59/-26, 50. The missing amino-acids between 26 to 38 (1AZX sequence numbering) were added by molecular editing, using as template the 3D structure of the corresponding sequence in PDB ID 1E03, in which only 27, 28, 29, 30, 31 amino-acids are missing. These last missing residues were then included by molecular editing (MAESTRO 9.8 graphical interface), adjusting the polypeptide dihedral angles ϕ/ψ until the two terminals matched. Hexa-4 and Hexa-8 were superposed to the AGA*IA sequence (1AZX) matching the H1 position of GlcA, GlcNS3S6S, and IdoA2S/IdoA, achieving a RMSD of 1.74 Å. The simulation cell was set by enveloping each macromolecule with a water layer 15 Å wide in the three directions, resulting in an orthogonal cell with edge lengths of approximately 100 Å. The no-bonded potential energy applied the standard cut-off technique (12 Å) for both electrostatic and dispersive interactions. Each cell was minimized using 100K steps of the default minimization algorithm included in the NAMD 2.12 simulation engine. The MD simulations were run sampling the NPT ensemble for the whole length, even if cell density equilibration required approximately 3 ns. The simulation temperature was set at 300 K and maintained by a Lowe-Andersen thermostat as implemented in NAMD 2.12, while the Nosé−Hoover Langevin piston algorithm controlled the pressure (1.01325 bar) applied on the cell walls. During the cell density equilibration steps (3 ns), a harmonic potential energy restraint was applied (harmonic constant of 5 kcal mol^{-1}) to all atoms of the complex, while water molecules were left free to move. The total length of the MD simulations were 100 ns and 270-260 ns for both glycans in free state and in bound state with AT respectively. The glycan·AT complex relaxation could be

visualized plotting the RMSD distance between the two glycans upon superposition of the A-helix of AT (Cα backbone, between amino-acids 45 and 70). The ligands RMSD showed an oscillation behaviour that decrease after 80-120 ns (approximately) indicating a convergence to a relative "more relaxed" distances.The RMSD reported in table 9 correspond to average values of the RMSD (Root Mean Square Distance) calculated between residue pairs of the common sequence (AGA*IA) concerning Hexa-4·AT, Hexa-8·AT and AGA*IA·AT after superposition of their A-helix. This analysis involves the set of hexasaccharides·AT complexes validated by the tr-NOEs, and correspond to geometries at instant time: 209, 227, 240, 257, 270 ns for Hexa-4·AT and 214, 235, 243, 258 ns for Hexa-8/AT. The five Hexa-4·AT and four Hexa-8·AT complexes were superposed by the Cα backbone of the A-helix (45 to 70 amino-acids in 1AZX sequence) and the corresponding average RMSD are reported in the second column of table 9. The residue ring atoms C1, C2, C3, C4, C5, O5 were used for the RMSD calculation.

The octa-5 and octa-5·AT complex were built following the same procedure previously reported for the hexa-4 and hexa-8; in octa-5 and octa-5·AT the initial conformation of GlcNS6S, GlcNS3S6S, and GlcA was initially set to the chair $^{4}C_{1}$, while that of IdoA2S at the non-reducing and reducing end were set at the beginning as $^{1}C_{4}$ and $^{2}S_{0}$ respectively, in accord to early NOEs analysis. The initial dihedral angle conformation (ϕ_i/ψ_i) of octa-5 in both bound and un-bound state was set in accord to the disaccharide linkage: GlcNS6S-GlcA-GlcNS6S-IdoA2S-GlcNS3S6S-IdoA2S-GlcNS6S-GlcA-PhNO2, and correspond to the sequence: -49/-14, 59/-6, -31/-35, 41/1, -71/-33, 44/16, -55/-28. The initial geometry of the octa-5·AT complex was built fitting the AIA*IA sequence on the corresponding AGA*IA pentasaccharide in bound state with AT, as included in the PDB ID 1AZX (L chain); this goal was achieved matching the C4 and C1 carbons of the residues in the moiety GlcNS6S-IdoA2S-GlcNS3S6S of octa-5 with the corresponding GlcNS6S-GlcA-GlcNS3S6S of the

AGA*IA of the 1AZX complex. Octa-5 and octa-5-AT complex were solvated covering each molecule/macromolecule by a water molecule layer (TIP3P) 15 Å wide, the final simulation cell has an orthogonal geometry and periodic boundary condition were applied on the hedges. MD simulations were run using the same conditions and parameters as done before for hexa-4/hexa-8 and the corresponding complexes. The whole simulation time span 350 and 480 ns for octa-5 and octa-5·AT respectively; while the equilibration periods were estimated to be 200 and 250-300 ns for the octa-5 and the complex, following the relaxation of the set of glycosidic dihedral angles $\phi_i(t)/\psi_i(t)$ in the former case, and the RMSD(t) of the ligand in the latter case.

8.1.5. NOEs simulation NOEPROM

Intra-residue and inter-glycosidic NOEs were simulated using NOEPROM software with the isotropic model of motion; the single correlation time (Tc = 700 ps) were estimated matching the simulated H1-H2 intra-residue NOEs of the glucosamines with the corresponding experimental values. The Hexa-4 and Hexa-8 geometries were obtained at the end of the MD simulations adjusting the ϕ/ψ glycosidic dihedrals with the averages calculated on the MD simulation trajectories after a relaxation period of 40 ns.

8.1.6. tr-NOEs simulation CORCEMA

The tr-NOEs simulation applied CORCEMA software on Hexa-4·AT and Hexa-8·AT complex geometries selected from the MD simulation, after a relaxation time between 80-120 ns. A “two state” model for the interaction, in which no significant conformational changes are predicted for the AT Receptor Binding Site (RBS), was used. The kinetics parameters were estimated matching the H1-H2 intra-residue tr-NOEs of the glucosamines residues, using trial values from previous experiences

involving octasaccharides·AT complexes. The kinetic parameters for Hexa-4 were set to k_{off} = 18 s^{-1} k_{on} = 3.27 10^7 M^{-1} s^{-1}, while for Hexa-8 were k_{off} = 15 s^{-1} k_{on} = 2.68 10^6 M^{-1} s^{-1}. Using these kinetic constants, the corresponding thermodynamic dissociation constants (K_d) were 5.5 10^{-7} and 5.6 10^{-6} M for Hexa-4·AT and Hexa-8/AT, in accord to the ITC estimated values. The initial concentration of both Hexa-4 and Hexa-8 was 5.84 10^{-4} M using a glycan/AT (L/E) ratio 5:1, the total AT concentration for both tr-NOE experiments was E_{tot} = 1.12 10^{-4} M. The equilibrium concentrations of the AT in Hexa-4·AT and Hexa-8·AT tr-NOE experiments were 1.10 10^{-7} and 1.11 10^{-6} M respectively, finally the estimated time decay constants for the CORCEMA simulation were τ_E = 5.48 10^{-5} s^{-1}, τ_{L^*} = 0.278 s^{-1}, τ_{EL} = 0.0556 s^{-1} for Hexa-4·AT and τ_E = 6.69 10^{-4} s^{-1}, τ_{L^*} = 0.337 s^{-1}, τ_{EL} = 0.0667 s^{-1} for Hexa-8/AT. The agreement between the theoretical and experimental tr-NOEs was evaluated by calculating the fitting quality parameter R factors as defined in eq 1.

$$R = \sum_{Tmix}\left[NOE_{Tmix}^{EXP} - NOE_{Tmix}^{CALC}\right]^2 \Big/ \sum_{Tmix}\left[NOE_{Tmix}^{EXP}\right]^2 \qquad (1)$$

R Factor values higher than 0.3 are indicative of a pour local structure description.

8.2. **Heparan Sulfate Mimetic**

8.2.1. General Methods

Solvents used were HPLC grade unless stated otherwise. CE analyses were performed in reverse polarity mode on an Agilent CE System using 10mM 5-sulfosalicylic acid, pH 3, as the background electrolyte and detection by indirect UV absorbance at 214 nm, as previously described[82]. Size-exclusion chromatography was performed over Bio-Gel P2 eluting with 0.1M NH4HCO3 supplied by a Pharmacia LKB Pump P-500 at 196 mL/h, and 20 mL fractions were collected using a Pharmacia LKB Frac-200 fraction collector. Fractions were quickly checked

using a combination of the H2SO4 char test and metachromatic staining with 1,9-dimethyl methylene blue, and those containing product were diluted for analysis by CE. Fractions judged to contain a significant amount of salt-free product were combined and lyophilized, then redissolved in HPLC water and lyophilized again.

8.2.2. Sulfation of the pentasaccharide fraction

A mixture of pentasaccharide 6 (980 mg, 1.04 mmol, purity 94.4%) and sulfur trioxide pyridine complex (9.0 g, 56 eq., ~3 eq. per hydroxyl group) in anhydrous DMF (25 mL) was stirred for 45 h. The DMF solvent was decanted, and the gummy precipitate was washed briefly with anhydrous EtOH (2 mL) and then dissolved in water (5 mL). The resulting strongly acidic (pH 1.4) solution was basified to pH 9.3 by the addition of 2M NaOH solution (10.2 mL). Free pyridine was removed by washing with DCM (5 × 5 mL). The pale brown aqueous layer was decolorized by passage through a Waters Sep-Pak Vac 20cc C18-5g solid-phase extraction cartridge, preconditioned by wetting with MeOH, then washing with water: MeOH (1:1→9:1→99:1). The oligosaccharide was loaded as a solution in 99:1 water:MeOH and the cartridge was washed with 99:1 water:MeOH (60 mL) until the eluate tested negative for the presence of carbohydrates (H2SO4 char). The appropriate fractions were combined and evaporated to dryness. The resulting glassy solid was desalted by size exclusion chromatography to give the sulfated pentasaccharide 7 (1.02 g, 37.9%) as a white solid.

8.2.3. Sulfation of the tetrasaccharide fraction

The tetrasaccharide 4 (300 mg, 0.38 mmol) was sulfonated as described above for the pentasaccharide 6 to give the sulfated tetrasaccharide 5 (447 mg, 55%) as an off-white solid.

Sulfonation of the trisaccharide fraction The trisaccharide fraction of the OPF (297 mg, 0.48 mmol) was sulfonated as described above to give the sulfated trisaccharide fraction containing a mixture of compounds 13 and 3 (497 mg, 63%) as an off-white solid.

8.2.4. Sulfatio of the disaccharide fraction

The disaccharide 10 (100 mg, 0.22 mmol) was sulfonated as described above to give the sulfated disaccharide 11 (150 mg, 58%) as an off-white solid.

8.2.5. Mass spectrometry

LC–MS analysis was performed on a HPLC 1100 (Agilent) coupled to a 7.0 T Fourier transform ion cyclotron resonance (FTICR) mass spectrometer equipped with electrospray (ESI) source (Solarix, Bruker). The ion pair reversed phase (IP RP) liquid chromatography separation was conducted on a Kinetex C18 column (2.1 × 100 mm, 2.6 μm particle size, Phenomenex, Aschaffenburg, Germany) using dibutylamine for ion paring. A volume of 3 μL of the sample solution prepared at the concentration of 0.2 mg/mL was introduced via an autosampler; the separation was carried out at 30°C by running the mobile phases A (dibutylamine 10 mM, acetic acid 10mM in water) and B (dibutylamine 10 mM, acetic acid 10mM in methanol) at 0.1 mL/min according to the following gradient scheme: isocratic step at 25% B for 2 min, followed by a fast linear gradient to 60% B in 3 min and a second slower gradient from 60% to 95% B in 30 min; then, the column was held at 95% B for 5 min and reequilibrated to the initial conditions. High resolution full mass scan spectra were recorded in negative polarity in the mass range from 200 to 3000 using the following ESI source parameters: capillary voltage: +3200 V, nitrogen gas used

as nebulizer and heater gases was set at 1.0 bar and 3.7 L/min, respectively. Mass calibration was performed by using sodium trifluoroacetate solution (0.05 mg/ml in water-acetonitrile 1:1 v/v).

8.2.6. NMR spectroscopy

NMR data at 500 and 600 MHz were recorded in D2O; chemical shifts of the isolated oligosaccharides are reported in ppm relative to D2O at 4.80 for ^{1}H and externally referenced to TMS 0.0 for 13C at 30°C. Chemical shifts of PI-88 are reported in ppm externally referenced to TSP 0.0 for 13C and ^{1}H at 30°C. ^{1}H resonances for oligosaccharides were fully assigned using the HSQC and COSY pulse sequences, and the inter-ring connectivity was established using HMBC correlations. Oligosaccharide residues are indicated with consecutive letters from the reducing end. Proton spectra were recorded with water presaturation with a recycle delay of 12 s and 16 scans. HSQC (heteronuclear single quantum coherence) and HMBC (heteronuclear multiple bond correlation) experiments were performed in phase sensitivity enhanced pure-absorption mode using 16 dummy scans, 16 and 32 scans, respectively, with decoupling during acquisition period and 2.5 s of relaxation delay. The matrix size of 2048 × 320 data points was increased to 4096 × 1024 by linear prediction and zero filling. 2D homonuclear correlation COSY and TOCSY experiments were obtained in phase sensitivity mode using 8 scans with water presaturation during acquisition. In TOCSY experiments isotropic mixing was performed applying a MLEV pulse train of 100 ms. NOESY experiments were obtained with a total of 16 scans, a mixing time of 150 ms and a recycle delay of 5.5 s. All homonuclear correlation experiments were acquired with a matrix size of 1048 × 320 data points with zero filling in F1 (2048 × 1024) before Fourier transformation.

8.2.7. NOESY analysis

The estimation of the interglycosidic distances was done by correlating the ratio between the % of NOEs generated by a selected Hx–Hy pair across the glycosidic bond, whose distance is unknown, and the % of NOEs of the reference pair H1–H2, that belong to a Man residue whose distance is 2.53 Å. Only the NOE % measured at the lowest mixing time is considered. Due to the large line widths of the anomeric signals of residue A and B, NOE intensities were referenced to one third of H1 of E + C + D diagonal peak volumes.

8.2.8. MD simulations

Models of polysulfated and nonsulfated pentasaccharides (compounds 7 and 6, respectively) were built using Maestro/Macromodel 9.8 software, where the Man residues were set initially in the $^4C_1$ conformation, and the connectivity between residues is Man-α (1→3)-Man, except for the glycosidic junction at the reducing end, whose connectivity is Man-α(1→2)-Man (see NMR section and Almond et al. 2001). For Molecular Mechanics and MD simulations the Amber* force field as implemented in Maestro/Macromodel 9.8 was used, which include Homan's parameters and atom types for pyranoses. The non-bonded cut-offs were set to 20.0, 8.0 and 4.0 Å for electrostatic, van der Waals and hydrogen bond interactions, respectively. The solvent description involved the Generalized Born Implicit Solvent method as implemented in Macromodel. The two glycan models after building were energy minimized (bmin) setting the Max number of steps = 10 K, and Energy Gradient Threshold = 10^{-3} Kj $mol^{-1}Å^{-1}$. The energy minimization algorithm chosen was the PRCG (Polak-Ribiere Conjugate Gradient). The two glycan models were then submitted to a multistep sequence of eleven MD simulation runs with temperature progressively increasing from 300 K (1st step) to the highest temperature value of 400 K (6th step) using steps

of 20 K, decreasing again to the final value of 300 K, this last (11^{th} step) being the production step, in which the glycosidic ϕ/ψ distribution was determined using a 2D histogram binning approach as further explained. The time length of each fixed temperature MD run was 20 ns for a whole duration of 220 ns. The MD simulations were terminated by a simulated annealing, in which the temperature is progressively reduced (using steps of 100 ps length and $\Delta T = 20$ K) from T = 300 K to 20 K, before applying a final energy minimization; the obtained geometries of compound 7 and 6 were optimized in their backbone using the glycosidic dihedral angles determined by 2D histogram binning applied on the final stage (11^{th} step) of the multistep MD simulation. The conformational sampling of the two pentasaccharides was analyzed using Ramanchandran plots for each glycosidic linkage; each plot was built using a 2D histogram binning approach, where a colored density map allowed the localization of the most populated ϕ/ψ states with an accuracy of ±10°, that corresponds to approximately half of the smallest unit of a 2D histogram binning grid. To evaluate the efficiency of the multistep MD simulation approach on the conformational sampling of compound 7, the 2D histogram binning analysis is applied also on the first and sixth MD step and compared to the production step (11^{th}). The R-statistical software (hexbin package), was used to generate the 2D histograms of the Ramachandran plots.

8.4. **Chemoenzymatic Synthesis**

8.4.1. Expression of HS biosynthetic enzymes

A total of four enzymes were used for the synthesis, including NST, C_5-epi, 2-OST, and PmHS2. All the enzymes were expressed in *E. coli* and purified by appropriate affinity chromatography as described previously [2,3].

8.4.2. Preparation of enzyme cofactors

A sulfate donor, 3'-phosphoadenosine 5'-phosphosulfate (PAPS), was prepared from ATP and sodium sulfate using adenosine phosphokinase and ATP-sulfurylase [4]. The purity of PAPS was more than 95% as measured by PAMN-HPLC and its structure was confirmed by ESI-MS. The preparation of UDP-GlcNTFA was started from glucosamine (Sigma-Aldrich), which was first converted to GlcNTFA by reacting with *S*-ethyl trifluorothioacetate (Sigma-Aldrich) in accord to the protocol described previously [2]. The resultant GlcNTFA was converted to GlcNTFA-1-phosphate using *N*-acetylhexosamine 1-kinase [5].

The UDP-GlcNTFA synthesis was completed by transforming GlcNTFA-1-phosphate using glucosamine-1-phosphate acetyltransferase/*N*-acetylglucosamine-1-phosphate uridyltransferase (GlmU) as described [2]. The resultant UDP-GlcNTFA was ready for the elongation reaction. The successful synthesis of UDP-GlcNTFA was confirmed by PAMN-HPLC analysis as UDP-GlcNTFA was eluted at 17 to 18 min as a distinct UV 260 nm peak.

8.4.3. Synthesis of compound 1

The conversion of a starting material trisaccharide –GlcA-GlcNTFA-X- to compound **1** involves five steps, including three elongation steps, one detrifluoroacetylation and one *N*-sulfation step. Elongation of –GlcA-GlcNTFA-X- to GlcNTFA-GlcA-GlcNTFA–GlcA-GlcNTFA-X- was completed in three steps using bacterial glycosyl transferases heparosan synthase-2 (PmHS2)

from *Pasteurella multocida*. To introduce a GlcNTFA residue, –GlcA-GlcNTFA-X- (1.2 mM) was incubated with PmHS2 (250 µg ml^{-1}) in 4 ml buffer containing Tris (25 mM, pH 7.2), $MnCl_2$ (5 mM) and UDP-GlcNTFA (2 mM) at 37°C overnight. To introduce a GlcA residue, tetrasaccharide substrate, GlcNTFA-GlcA-GlcNTFA-X- (1.8 mM), was incubated with PmHS2 (250 µg ml^{-1}) in a buffer containing Tris (25 mM, pH 7.2), $MnCl_2$ (5 mM) and UDP-GlcA (5.5 mM) at 37°C overnight. The product after each elongation step was purified using a BIOGEL P2 column (1 × 220 cm), which was eluted with H_2O and 0.1% TFA in 18 hours at a flow rate of 125 µl min^{-1}. The identity of the product was confirmed by ESI-MS only because this compound lacks of UV detectable groups. The addition of GlcNTFA residue was repeated one more time to form a hexasaccharide backbone for the subsequent *N*-detrifluoroacetylation and *N*-sulfationreaction. Purified N-trfuoroacetyl hexasaccharides was dried and resuspended in a solution of 1 mL containing Methanol, ddH_2O and Triethylamine (v/v/v=2:2:1). The reaction was incubated at room temperature for 20 hours. The conversion degree of the detrifluoroacetylation reaction was monitored by ESI-MS. The sample was then dried by a vacuum concentrator (Centrivap Concentrator, Labconco) and reconstituted in ddH_2O to recover the N-deacetylated hexsaccharide. Upon the completion of detrifluoroacetylation, the pH of the reaction mixture set to 7.0 and 100 µg of the compound incubated with 50 mM MES (pH 7.0), 10 µg ml^{-1} NST and 1 mM PAPS at 37°C overnight, where the amount of PAPS was about 1.5-times molar amount of NH2 groups in the pentasaccharide. The reaction was monitored by ESI-MS.

8.4.4. Synthesis of compound 2

The conversion of a starting material trisaccharide –GlcA-GlcNpnp-X- to compound **1** involves five steps, including three elongation steps, one detrifluoroacetylation and one *N*-sulfation step. Elongation of –GlcA-GlcNpnp-X- to GlcNTFA-GlcA -GlcNTFA-GlcA-GlcNTFA–GlcA-GlcNpnp-X- was completed in five steps using bacterial glycosyl transferases heparosan synthase-2 (PmHS2) from *Pasteurella multocida.* To introduce a GlcNTFA residue, –GlcA-GlcNTFA-X- (1.2 mM) was incubated with PmHS2 (250 µg ml^{-1}) in 4 ml buffer containing Tris (25 mM, pH 7.2), $MnCl_2$ (5 mM) and UDP-GlcNTFA (2 mM) at 37°C overnight. To introduce a GlcA residue, tetrasaccharide substrate, GlcNTFA-GlcA-GlcNTFA-X- (1.8 mM), was incubated with PmHS2 (250 µg ml^{-1}) in a buffer containing Tris (25 mM, pH 7.2), $MnCl_2$ (5 mM) and UDP-GlcA (5.5 mM) at 37°C overnight. The product after each elongation step was purified using a C_{18} column (0.75 × 20 cm; Biotage), which was eluted with a linear gradient of 0–100% Methanol and 0.1% in H_2O and 0.1% TFA in 60 min at a flow rate of 1 ml min^{-1}. The eluent was monitored by the absorbance at 310 nm and 260, and the identity of the product was confirmed by ESI-MS. The addition of further GlcNTFA, GlcA and GlcNTFA residues was repeated to form an octasaccharide backbone for the subsequent *N*-detrifluoroacetylation/*N*-sulfation. Purified N-trfuoroacetyl octasaccharide was dried and resuspended in a solution of 1 mL containing Methanol, ddH_2O and Triethylamine (v/v/v=2:2:1). The reaction was incubated at room temperature for 20 hours. The conversion ratio of the detrifluoroacetylation reaction was monitored by ESI-MS. The sample was then dried by a vacuum concentrator (Centrivap Concentrator, Labconco) and reconstituted in ddH_2O to recover the N-deacetylated octasaccharide. The deacetylated compound were resuspended in ddH_2O and $NaHCO_3$ (40 eq for free NH_2 group) and then added in four different portions in 2 hours Py-SO_3 (20 eq for free NH_2 group). The reacton mixture was let under stirring a room temperature overnight. The reaction was monitored using a C_{18} column (0.75 × 20 cm; Biotage), which was eluted with a linear gradient

of 0–100% Methanol and 0.1% in H_2O and 0.1% TFA in 60 min at a flow rate of 1 ml min^{-1}. *N*-sulfated product was purified by Q-Sepharose (GE Health Care), and the purified product with the structure of GlcNS-GlcA -GlcNS-GlcA-GlcNS–GlcA-GlcNpnp-X- was dialyzed. The reaction mixture was incubated in 50 mM MOPS (pH 7.0), $MnCl_2$ 2 mM, 10 μgml^{-1} C_5-epi, 2-OST 10 μgml^{-1}, and 0.2 mM PAPS at 37^oC overnight, where the amount of PAPS was about 1.5-times molar amount of the pentasaccharide. The reaction mixture was then purified by Q-Sepharose column to obtain **2.**

9. Abbreviations

AGA*IA: Minimun pentassacharide sequence that binds and activates Antithrombin

AT: Antithrombin

EXT1/EXT2: Exostosin-1/2

GAG: Glycosaminoglycan

GalN: Galactosamine

GlcA: Glucuronic Acid

GlcN: Glucosamine

HBP: Heparan sulfate/Heparin binding proteins

HBS: Heparin Binding Site

Hep: Heparin

HS: Heparan sulfate

HSQC: Heteronuclear Single Quantum Coherence

IdoA: Iduronic Acid

NMR: Nuclear Magnetic Resonance

NOE: Nuclear Overhauser Effect

STD: Saturation Transfer Difference

UFH: Unfractionated Heparin

10. Bibliography

1. Casu, B. & Lindahl, U. Structure and biological interactions of heparin and heparan sulfate. *Adv. Carbohydr. Chem. Biochem.* **57**, 159–206 (2001).

2. Lindahl, U. *et al.* Structure of the antithrombin-binding site in heparin. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3198–202 (1979).

3. Rosenberg, R. D. & Lam, L. Correlation between structure and function of heparin. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 1218–22 (1979).

4. AB 2.1.1.1 (or EC 2.1.1.?). *Eur. J. Biochem.* **247**, 733–736 (1997).

5. NOMENCLATURE ANNOUNCEMENT. *Arch. Biochem. Biophys.* **344**, 242–252 (1997).

6. Casu, B., Naggi, A. & Torri, G. Re-visiting the structure of heparin. *Carbohydr. Res.* **403**, 60–68 (2015).

7. Esko, J. D. & Selleck, S. B. Order Out of Chaos: Assembly of Ligand Binding Sites in Heparan Sulfate. *Annu. Rev. Biochem.* **71**, 435–471 (2002).

8. Casu, B. *et al.* The structure of heparin oligosaccharide fragments with high anti-(factor Xa) activity containing the minimal antithrombin III-binding sequence. Chemical and 13C nuclear-magnetic-resonance studies. *Biochem. J.* **197**, 599–609 (1981).

9. BjÖrk, I. & Lindahl, U. Mechanism of the anticoagulant action of heparin. *Mol. Cell. Biochem.* **48**, 161–182 (1982).

10. Tovar, A. M. F. *et al.* Heparin from bovine intestinal mucosa: Glycans with multiple sulfation patterns and anticoagulant effects. *Thromb. Haemost.* **107**, 903–915 (2012).

11. Casu, B. Structure and Biological Activity of Heparin. *Adv. Carbohydr. Chem. Biochem.* **43**, 51–134 (1985).

12. Comper, W. D. & Laurent, T. C. Physiological function of connective tissue polysaccharides. *Physiol. Rev.* **58**, 255–315 (1978).

13. Gallagher, J. T. & Walker, A. Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides. *Biochem. J.* **230**, 665–74 (1985).

14. Hagner-McWhirter, A. *et al.* Biosynthesis of heparin/heparan sulfate: kinetic studies of the glucuronyl C5-epimerase with N-sulfated derivatives of the Escherichia coli K5 capsular polysaccharide as substrates. *Glycobiology* **10**, 159–171 (2000).

15. Ferro, D. R. *et al.* Evidence for conformational equilibrium of the sulfated L-iduronate residue in heparin and in synthetic heparin mono- and oligo-saccharides: NMR and force-field studies. *J. Am. Chem. Soc.* **108**, 6773–6778 (1986).

16. CASU, B. *et al.* Controversial glycosaminoglycan conformations. *Nature* **322**, 215–216 (1986).

17. Ferro, D. R. *et al.* Conformer populations of L-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **195**, 157–67 (1990).

18. Mulloy, B., Forster, M. J., Jones, C. & Davies, D. B. N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochem. J.* **293**, 849–858 (1993).

19. Ferro, D. R. *et al.* Conformer populations of l-iduronic acid residues in glycosaminoglycan sequences. *Carbohydr. Res.* **195**, 157–167 (1990).

20. Hricovíni, M. Solution Structure of Heparin Pentasaccharide: NMR and DFT Analysis. *J. Phys. Chem. B* **119**, 12397–12409 (2015).

21. Guerrini, M., Mourier, P. A. J., Torri, G. & Viskov, C. Antithrombin-binding oligosaccharides: Structural diversities in a unique function? *Glycoconj. J.* **31**, 409–416 (2014).

22. Das, S. K. *et al.* Synthesis of Conformationally Locked Carbohydrates: A Skew-Boat Conformation ofL-Iduronic Acid Governs the Antithrombotic Activity of Heparin. *Angew. Chemie Int. Ed.* **40**, 1670–1673 (2001).

23. Petitou, M. & van Boeckel, C. A. A. A Synthetic Antithrombin III Binding Pentasaccharide Is Now a Drug! What Comes Next? *Angew. Chemie Int. Ed.* **43**, 3118–3133 (2004).

24. Raman, R., Sasisekharan, V. & Sasisekharan, R. Structural Insights into Biological Roles of Protein-Glycosaminoglycan Interactions. *Chem. Biol.* **12**, 267–277 (2005).

25. Perrimon, N. & Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **404**, 725–728 (2000).

26. Thompson, S. M. *et al.* Novel 'phage display antibodies identify distinct heparan sulfate domains in developing mammalian lung. *Pediatr. Surg. Int.* **23**, 411–417 (2007).

27. Izvolsky, K. I. *et al.* Heparan sulfate–FGF10 interactions during lung morphogenesis. *Dev. Biol.* **258**, 185–200 (2003).

28. Ori, A., Wilkinson, M. C. & Fernig, D. G. A Systems Biology Approach for the Investigation of the Heparin/Heparan Sulfate Interactome. *J. Biol. Chem.* **286**, 19892–19904 (2011).

29. Ghiselli, G. Heparin Binding Proteins as Therapeutic Target: An Historical Account and Current Trends. *Medicines* **6**, 80 (2019).

30. Van Vactor, D., Wall, D. P. & Johnson, K. G. Heparan sulfate proteoglycans and the emergence of neuronal connectivity. *Current Opinion in Neurobiology* **16**, 40–51 (2006).

31. Höök, M., Björk, I., Hopwood, J. & Lindahl, U. Anticoagulant activity of heparin: Separation of high-activity and low-activity heparin species by affinity chromatography on immobilized antithrombin. *FEBS Lett.* **66**, 90–93 (1976).

32. Lam, L. H., Silbert, J. E. & Rosenberg, R. D. The separation of active and inactive forms of heparin. *Biochem. Biophys. Res. Commun.* **69**, 570–577 (1976).

33. Lindahl, U., Bäckström, G., Thunberg, L. & Leder, I. G. Evidence for a 3-O-sulfated D-glucosamine residue in the antithrombin-binding sequence of heparin. *Proc. Natl. Acad. Sci. U. S. A.* **77**, 6551–5 (1980).

34. Sinaÿ, P. *et al.* Total synthesis of a heparin pentasaccharide fragment having high affinity for antithrombin III. *Carbohydr. Res.* **132**, C5–C9 (1984).

35. Jin, L. *et al.* The anticoagulant activation of antithrombin by heparin. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 14683–8 (1997).

36. Desai, U. R., Petitou, M., Björk, I. & Olson, S. T. Mechanism of Heparin Activation of Antithrombin. *J. Biol. Chem.* **273**, 7478–7487 (1998).

37. Olson, S. T., Björk, I. & Bock, S. C. Identification of critical molecular interactions mediating heparin activation of antithrombin: implications for the design of improved heparin anticoagulants. *Trends Cardiovasc. Med.* **12**, 198–205 (2002).

38. Olson, S. T. *et al.* Role of the antithrombin-binding pentasaccharide in heparin acceleration of antithrombin-proteinase reactions. Resolution of the antithrombin conformational change contribution to heparin rate enhancement. *J. Biol. Chem.* **267**, 12528–38 (1992).

39. and, D. J. D. J. & Huntington*, J. A. Crystal Structure of Antithrombin in a Heparin-Bound Intermediate State†,‡. (2003). doi:10.1021/BI034524Y

40. Guerrini, M. *et al.* Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction: Evidence for very high affinity induced by an unusual glucuronic acid residue. *J. Biol. Chem.* **283**, 26662–26675 (2008).

41. Guerrini, M. *et al.* Conformational transitions induced in heparin octasaccharides by binding with antithrombin III. *Biochem. J.* **399**, 191 (2006).

42. Guerrini, M. *et al.* Antithrombin-binding octasaccharides and role of extensions of the active pentasaccharide sequence in the specificity and strength of interaction. Evidence for very high affinity induced by an unusual glucuronic acid residue. *J. Biol. Chem.* **283**, 26662–75 (2008).

43. Guerrini, M. *et al.* Effects on Molecular Conformation and Anticoagulant Activities of 1,6-Anhydrosugars at the Reducing Terminal of Antithrombin-Binding Octasaccharides Isolated from Low-Molecular-Weight Heparin Enoxaparin. *J. Med. Chem.* **53**, 8030–8040 (2010).

44. Guerrini, M. *et al.* An unusual antithrombin-binding heparin octasaccharide with an additional 3-O-sulfated glucosamine in the active pentasaccharide sequence. *Biochem. J.* **449**, 343–351 (2013).

45. Viskov, C. *et al.* Heparin Dodecasaccharide Containing Two Antithrombin-binding Pentasaccharides. *J. Biol. Chem.* **288**, 25895–25907 (2013).

46. Crowther, M. A., Berry, L. R., Monagle, P. T. & Chan, A. K. C. Mechanisms responsible for the failure of protamine to inactivate low-molecular-weight heparin. *Br. J. Haematol.* **116**, 178–186 (2002).

47. Blossom, D. B. *et al.* Outbreak of Adverse Reactions Associated with Contaminated Heparin. *N. Engl. J. Med.* **359**, 2674–2684 (2008).

48. Kishimoto, T. K. *et al.* Contaminated Heparin Associated with Adverse Clinical Events and Activation of the Contact System. *N. Engl. J. Med.* **358**, 2457–2467 (2008).

49. Guerrini, M. *et al.* Oversulfated chondroitin sulfate is a contaminant in heparin associated with adverse clinical events. *Nat. Biotechnol.* **26**, 669–675 (2008).

50. Xu, Y. *et al.* Chemoenzymatic synthesis of homogeneous ultralow molecular weight heparins.

Science **334**, 498–501 (2011).

51. Yates, E. A., Guimond, S. E. & Turnbull, J. E. Highly Diverse Heparan Sulfate Analogue Libraries: Providing Access to Expanded Areas of Sequence Space for Bioactivity Screening. *J. Med. Chem.* **47**, 277–280 (2004).

52. Xu, Y. *et al.* Homogeneous low-molecular-weight heparins with reversible anticoagulant activity. *Nat. Chem. Biol.* **10**, 248–250 (2014).

53. Guerrini, M., Bisio, A. & Torri, G. Combined Quantitative 1H and 13C Nuclear Magnetic Resonance Spectroscopy for Characterization of Heparin Preparations. *Semin. Thromb. Hemost.* **27**, 473–482 (2001).

54. Mauri, L. *et al.* Qualification of HSQC methods for quantitative composition of heparin and low molecular weight heparins. *J. Pharm. Biomed. Anal.* **136**, 92–105 (2017).

55. Peng, J. W., Moore, J. & Abdul-Manan, N. NMR experiments for lead generation in drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **44**, 225–256 (2004).

56. Calle, L. P., Cañada, F. J. & Jiménez-Barbero, J. Application of NMR methods to the study of the interaction of natural products with biomolecular receptors. *Nat. Prod. Rep.* **28**, 1118 (2011).

57. Christopher A. Lepre, †,‡, Jonathan M. Moore, *,† and & Jeffrey W. Peng‡, §. Theory and Applications of NMR-Based Screening in Pharmaceutical Research. (2004). doi:10.1021/CR030409H

58. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie Int. Ed.* **38**, 1784–1788 (1999).

59. Carlomagno, T. Ligand-Target Interactions: What Can We Learn from NMR? *Annu. Rev. Biophys. Biomol. Struct.* **34**, 245–266 (2005).

60. Ni, F. Recent developments in transferred NOE methods. *Prog. Nucl. Magn. Reson. Spectrosc.* **26**, 517–606 (1994).

61. Meragelman, T. L. Basic One- and Two-Dimensional NMR Spectroscopy, 4th Edition. By Horst Friebolin (Organisch-Chemisches Institut der Universität, Heidelberg). Wiley-VCH, Weinheim. 2005. xxiv + 406 pp. 6 1/2 × 9 1/2 in. $69.95. ISBN 3-527-31233-1. (2005). doi:10.1021/NP058259O

62. Moseley, H. N. B., Curto, E. V. & Krishna, N. R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of NOESY Spectra of Interacting Systems; Two-Dimensional Transferred NOESY. *J. Magn. Reson. Ser. B* **108**, 243–261 (1995).

63. Roccatano, D. A Short Introduction to the Molecular Dynamics Simulation of Nanomaterials. in *Micro and Nanomanufacturing Volume II* 123–155 (Springer International Publishing, 2018). doi:10.1007/978-3-319-67132-1_6

64. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967).

65. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

66. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

67. Kreuger, J., Spillmann, D., Li, J. & Lindahl, U. Interactions between heparan sulfate and proteins: the concept of specificity. *J. Cell Biol.* **174**, 323–327 (2006).

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

69. Hsieh, P.-H., Thieker, D. F., Guerrini, M., Woods, R. J. & Liu, J. Uncovering the Relationship between Sulphation Patterns and Conformation of Iduronic Acid in Heparan Sulphate. **6**, (2016).

70. Muñoz-García, J. C. *et al.* Effect of the Substituents of the Neighboring Ring in the Conformational Equilibrium of Iduronate in Heparin-like Trisaccharides. *Chem. - A Eur. J.* **18**, 16319–16331 (2012).

71. Hricovíni, M. *et al.* Conformation of heparin pentasaccharide bound to antithrombin III. *Biochem. J.* **359**, 265–72 (2001).

72. van Boeckel, C. A. A., Grootenhuis, P. D. J. & Visser, A. A mechanism for heparin-induced potentiation of antithrombin III. *Nat. Struct. Mol. Biol.* **1**, 423–425 (1994).

73. Petitou, M., Barzu, T., Herault, J.-P. & Herbert, J.-M. A unique trisaccharide sequence in heparin mediates the early step of antithrombin III activation. *Glycobiology* **7**, 323–327 (1997).

74. PETITOU, M., LORMEAU, J.-C. & CHOAY, J. Interaction of heparin and antithrombin III. The role of O-sulfate groups. *Eur. J. Biochem.* **176**, 637–640 (1988).

75. Wang, Z. *et al.* Synthesis of 3- *O* -Sulfated Oligosaccharides to Understand the Relationship between Structures and Functions of Heparan Sulfate. *J. Am. Chem. Soc.* **139**, 5249–5256 (2017).

76. Parish, C. R., Freeman, C., Brown, K. J., Francis, D. J. & Cowden, W. B. Identification of sulfated oligosaccharide-based inhibitors of tumor growth and metastasis using novel in vitro assays for angiogenesis and heparanase activity. *Cancer Res.* **59**, 3433–41 (1999).

77. Kudchadkar, R., Gonzalez, R. & Lewis, K. D. PI-88: a novel inhibitor of angiogenesis. *Expert Opin. Investig. Drugs* **17**, 1769–1776 (2008).

78. Lewis, K. D. *et al.* A phase II study of the heparanase inhibitor PI-88 in patients with advanced melanoma. *Invest. New Drugs* **26**, 89–94 (2008).

79. Liu, C.-J. *et al.* Heparanase inhibitor PI-88 as adjuvant therapy for hepatocellular carcinoma after curative resection: A randomized phase II trial for safety and optimal dosage. *J. Hepatol.* **50**, 958–968 (2009).

80. Vlodavsky, I. *et al.* Heparanase: From basic research to therapeutic applications in cancer and inflammation. *Drug Resist. Updat.* **29**, 54–75 (2016).

81. Ferro, V., Fewings, K., Palermo, M. C. & Li, C. Large-scale preparation of the oligosaccharide phosphate fraction of Pichia holstii NRRL Y-2448 phosphomannan for use in the manufacture of PI-88. *Carbohydr. Res.* **332**, 183–9 (2001).

82. Yu, G. *et al.* Preparation and anticoagulant activity of the phosphosulfomannan PI-88. *Eur. J. Med. Chem.* **37**, 783–791 (2002).

83. Ferro, V. *et al.* Determination of the composition of the oligosaccharide phosphate fraction of Pichia (Hansenula) holstii NRRL Y-2448 phosphomannan by capillary electrophoresis and HPLC. *Carbohydr. Res.* **337**, 139–146 (2002).

84. Handley, P. N., Carroll, A. & Ferro, V. New structural insights into the oligosaccharide phosphate fraction of Pichia (Hansenula) holstii NRRL Y2448 phosphomannan. *Carbohydr. Res.* **446**–**447**, 68–75 (2017).

85. Martin, N. H., Loveless, D. M., Main, K. L. & Pyles, A. K. Computation of through-space NMR shielding effects by functional groups common to peptides. *J. Mol. Graph. Model.* **25**, 1–9 (2006).

86. Cochran, S. *et al.* Probing the Interactions of Phosphosulfomannans with Angiogenic Growth Factors by Surface Plasmon Resonance. *J. Med. Chem.* **46**, 4601–4608 (2003).

87. Almond, A., Bunkenborg, J., Franch, T., Gotfredsen, C. H. & Duus, J. Ø. Comparison of Aqueous Molecular Dynamics with NMR Relaxation and Residual Dipolar Couplings Favors Internal Motion in a Mannose Oligosaccharide. *J. Am. Chem. Soc.* **123**, 4792–4802 (2001).

88. Rivara, S., Milazzo, F. M. & Giannini, G. Heparanase: a rainbow pharmacological target associated to multiple pathologies including rare diseases. *Future Med. Chem.* **8**, 647–680 (2016).

89. Sheng, J., Xu, Y., Dulaney, S. B., Huang, X. & Liu, J. Uncovering biphasic catalytic mode of C5-epimerase in heparan sulfate biosynthesis. *J. Biol. Chem.* **287**, 20996–1002 (2012).

11. Ringraziamenti

“Sei tutti i limiti che superi”

Alla fine di un lungo percorso come questo è doveroso ringraziare le persone che hanno contributo all’ottenimento di un risultato così importante.

Il primo, più grande e più sincero grazie va a mia madre e a mio fratello che, seppure a distanza, mi hanno sempre fatto sentire “a casa”. Hanno creduto in me, non dubitando mai neppure quando io stesso ho vacillato. È grazie a loro se sono quello che sono.

Non posso non menzionare i miei amici che, in maniera più o meno velata, hanno sempre fatto leva sul mio orgoglio spingendomi sempre in avanti, aiutandomi a superare tutte le difficoltà.

Come dimenticare poi le persone che, con la loro professionalità hanno reso possibile il raggiungimento di questo traguardo: i miei colleghi del Ronzoni, tutti, senza dimenticare nessuno! Un particolare ringraziamento va però a Marco e Stefano, miei tutor in questo lavoro, grazie ai quali ho imparato come approcciarmi e come condurre un progetto di ricerca così lungo e difficoltoso.

Infine un grazie speciale va a Cesare, mio mentore all’NMR, mio amico adesso.

Finché il tempo non passa
non sai cosa è stato
importante nella tua vita.

[Cesare Cosentino Photographer]