UNIVERSITÀ DEGLI STUDI DI PAVIA

DOTTORATO IN SCIENZE CHIMICHE E FARMACEUTICHE E INNOVAZIONE INDUSTRIALE

(XXXVI Ciclo)

Coordinatore: Chiar.mo Prof. Giorgio Colombo

NMR AND IN SILICO APPROACHES FOR INVESTIGATING PROTEIN-GLYCOSAMINOGLYCAN INTERACTIONS

Tesi di Dottorato di

Michela Parafioriti

AA 2022/2023

Tutor:

Chiar.ma Prof.ssa Simona Collina

Dr. Marco Guerrini

Co-tutor:

Dr. Stefano Elli

“What you get by achieving your goals is not as important as what you become by achieving your goals.”

Henry David Thoreau

Short abstract

Heparin and heparan sulfate (HS) are highly sulfated polysaccharides belonging to the glycosaminoglycan (GAG) family. They consist of repeating disaccharide units that include uronic acid and glucosamine. Exogenous heparin, derived from porcine or bovine tissues, is used as an anticoagulant drug and its therapeutic activity is primarily attributed to the binding and activation of antithrombin (AT). Heparan sulfate (HS), expressed on the cell surface and in the extracellular matrix, is involved in numerous physiological and pathological processes. It interacts with several proteins, including the spike (S) protein that drives the early stage of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and the heparanase (HPSE) enzyme, whose overexpression induces tumor growth and spread. In this thesis, nuclear magnetic resonance (NMR) experiments and computational methods were applied to gain molecular insights into the recognition process of heparin by AT. The combined use of NMR spectroscopy and in silico methodologies also allowed to define the structural and conformational features underlying both the interaction of HS with the SARS-CoV-2 S protein and the binding of HS mimetics to HPSE.

Riassunto breve

L'eparina e l'eparan solfato (HS) sono polisaccharidi solfatati appartenenti alla famiglia dei glicosaminoglicani (GAG). Sono composti da unità disaccaridiche ripetute di acido uronico e glucosammina. L'eparina esogena, estratta dalla mucosa di intestino suino o di polmone bovino, è usata come farmaco anticoagulante e la sua attività è principalmente correlata al legame e all'attivazione dell'antitrombina (AT). L'HS presente sulla superficie cellulare e nella matrice extracellulare è coinvolto in numerosi processi fisiologici e patologici. Esso interagisce con diverse proteine, tra cui la proteina spike implicata nella infezione da SARS-CoV-2 e l'enzima eparanasi, la cui sovraespressione è associata allo sviluppo e alla crescita di neoplasie. In questo lavoro di tesi, esperimenti di risonanza magnetica nucleare (NMR) e metodi computazionali sono stati applicati al fine di approfondire il meccanismo di riconoscimento molecolare dell'eparina da parte dell'AT. L'uso combinato di spettroscopia NMR e metodologie in silico ha inoltre permesso di analizzare gli aspetti strutturali e conformazionali alla base dell'interazione dell'HS con la proteina spike del SARS-CoV-2 e del legame tra mimetici dell'HS e l'eparanasi.

Abstract

Heparin and heparan sulfate (HS) are highly sulfated polysaccharides belonging to the glycosaminoglycan (GAG) family. They consist of repeating disaccharide units that include a uronic acid (β-D-glucuronic acid or α-L-iduronic acid) (1-4)-linked to a α-D-glucosamine. Their structural heterogeneity, including the size of the polysaccharide chain, the ratio between glucuronic acid and iduronic acid units, the quantity and distribution of sulfate groups along the glycan backbone, arises from variations in their biosynthesis. Exogenous heparin, derived from porcine or bovine mucosa, is used as an anticoagulant drug and its therapeutic activity is primarily attributed to the binding and activation of antithrombin (AT). Heparan sulfate (HS), expressed on the cell surface and in the extracellular matrix (ECM), binds to a plethora of proteins, modulating various physiological and pathological processes. Nuclear magnetic resonance (NMR) spectroscopy, in combination with molecular modelling, serves as a versatile strategy for studying the molecular mechanisms underlying the recognition of GAGs by proteins.

This thesis deals with the application of NMR techniques and computational methods to investigate the interaction between HS and the spike (S) protein of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), to analyse the binding of HS mimetics to heparanase (HPSE) and to deepen the molecular recognition of heparin by AT.

The S protein of SARS-CoV-2 plays a pivotal role in the recognition of the virus by host cells and its subsequent internalization. Specifically, SARS-CoV-2 infection begins with the interaction between HS and the receptor binding domain (S1-RBD), corresponding to the distal part of the S protein. This event facilitates the attachment of the viral particles to the cell surface and the binding of S1-RBD to the angiotensin converting enzyme 2 (ACE2) receptor, the preliminary step toward fusion of the virus-host cell membrane. Three synthetic HS oligosaccharides were employed to explore the binding of HS by S1-RBD. A molecular description of the interaction was achieved by a combined use of saturation transfer difference (STD) NMR, transferred nuclear Overhauser effect spectroscopy (tr-NOESY), docking and molecular dynamics (MD) simulations. Our studies allowed to define the GAG binding site on the S1-RBD surface and to determine that HS oligosaccharides bind S1-RBD in two binding modes characterized by opposite orientations of the ligand on the protein surface. Our results suggested that the binding of HS with S1-RBD has low specificity, promoting the co-receptor role of HS in the interaction between S1-RBD and the cell surface ACE2 receptor.

Heparanase (HPSE) is an endo-β-D-glucuronidase that cleaves HS chains, participating in the degradation and remodeling of the ECM. In inflammatory and oncological diseases, HPSE is often overexpressed, contributing to the release of growth factors and promoting angiogenesis. Exogenous heparin, structurally similar to HS, is able to inhibit the HPSE activity; however, besides being a substrate of HPSE, its anticoagulant properties limit the use in anticancer therapies. On the contrary, chemically modified heparins, including glycol-split heparins obtained by periodate oxidation followed by borohydride reduction and characterized by low anticoagulant activity, are innovative and promising HPSE inhibitors. Two synthetic glycol-split trisaccharide models were selected to study the recognition process of glycol-split heparins by HPSE. STD NMR and docking analysis revealed details on the molecular mechanism by which glycol-split oligosaccharides inhibit HPSE.

Saturation transfer difference (STD) NMR is employed to determine the ligand moieties involved in the recognition process and their proximity to the protein surface (*i.e.*, the binding epitope), but it gives no information about the nature and architecture of the binding site. The novel differential epitope mapping (DEEP)-STD NMR protocol, performed in D_2O and H_2O, can be used to identify polar residues (*i.e.*, arginine) in GAG-binding proteins. The combined use of STD NMR, DEEP-STD NMR and MD simulations allowed to analyse the differences in the interaction between two structurally similar heparin pentasaccharides and AT. Our studies indicated that, despite the remarkable structural similarity, they possess different affinity for AT and are characterized by a slightly different orientation in the AT-binding site.

Riassunto

L'eparina e l'eparan solfato (HS) sono polisaccaridi solfatati appartenenti alla famiglia dei glicosaminoglicani (GAG). Sono composti da unità disaccaridiche ripetute costituite da un acido glucuronico o iduronico legato ad una glucosamina. La loro eterogeneità strutturale (la lunghezza della catena polisaccaridica, il rapporto tra le unità di acido glucuronico e acido iduronico, la quantità e la distribuzione dei gruppi solfato) è il risultato di variazioni nella loro biosintesi. L'eparina esogena, estratta dalla mucosa di intestino suino o di polmone bovino, è utilizzata come farmaco anticoagulante. La sua attività è principalmente correlata al legame e all'attivazione dell'antitrombina (AT). L'HS presente sulla superficie cellulare e nella matrice extracellulare (ECM) interagisce con molteplici proteine, modulando vari processi fisiologici e patologici.

La spettroscopia di risonanza magnetica nucleare (NMR), abbinata a metodi computazionali, rappresenta una strategia efficace per studiare i meccanismi alla base del processo di riconoscimento dei GAG da parte delle proteine.

Questa tesi si concentra sull'applicazione di tecniche NMR e metodi computazionali per esplorare l'interazione tra l'HS e la proteina spike del SARS-CoV-2, per analizzare il legame tra mimetici dell'HS e l'enzima eparanasi (HPSE) e per approfondire i meccanismi molecolari alla base del riconoscimento dell'eparina da parte dell'AT.

La proteina spike del SARS-CoV-2 svolge un ruolo cruciale nel riconoscimento e nell'internalizzazione del virus da parte della cellula ospite. L'infezione da SARS-CoV-2 inizia con l'interazione tra l'HS e l'S1-RBD, corrispondente alla parte distale della proteina spike. Questo evento promuove l'accumulo delle particelle virali sulla superficie cellulare e il legame dell'S1-RBD con il recettore ACE2 che porta alla fusione delle membrane di virus e cellula ospite. Oligosaccaridi sintetici di HS sono stati utilizzati per studiare il legame dell'HS con l'S1-RBD. La descrizione molecolare dell'interazione è stata ottenuta attraverso l'uso combinato di esperimenti NMR (STD NMR e tr-NOESY), docking e simulazioni di dinamica molecolare (MD). I nostri studi hanno permesso di identificare il sito di legame preferenziale dei GAG sulla superficie dell'S1-RBD. Inoltre, hanno rivelato che oligosaccaridi di HS legano l'S1-RBD in due modi caratterizzati da un'orientazione opposta del ligando rispetto alla superficie della proteina. Questi risultati hanno evidenziato una bassa specificità nel legame tra gli oligosaccaridi di HS e l'S1-RBD, supportando il ruolo di co-recettore dell'HS nell'interazione tra l'S1-RBD e l'ACE2.

L'HPSE è una endo-β-D-glucuronidasi che idrolizza le catene di HS, partecipando alla degradazione e al rimodellamento della ECM. La sua sovraespressione è spesso associata a condizioni patologiche, tra cui infiammazione e cancro. L'eparina esogena, strutturalmente simile all'HS, è in grado di inibire l'attività dell'HPSE; tuttavia, il suo utilizzo nelle terapie antitumorali è limitato dal fatto che rappresenta un substrato per l'enzima e possiede proprietà anticoagulanti. Le eparine modificate, tra cui le eparine glycol-split ottenute dall'eparina mediante una reazione di ossidazione seguita da una reazione di riduzione e caratterizzate da una bassa attività anticoagulante, sono invece promettenti inibitori dell'HPSE. Trisaccaridi glycol-split sintetici sono stati utilizzati per analizzare il riconoscimento delle eparine glycol-split da parte dell'HPSE mediante esperimenti NMR (STD NMR) e calcoli di docking. I risultati ottenuti hanno rivelato dettagli del meccanismo molecolare attraverso cui gli oligosaccaridi glycol-split inibiscono l'HPSE.

La spettroscopia NMR è utile a caratterizzare i complessi GAG-proteina a livello molecolare in soluzione. La tecnica STD NMR viene utilizzata per determinare le porzioni del ligando coinvolte nell'interazione con un recettore, ma non fornisce informazioni sulla natura e sull'architettura del sito di legame di quest'ultimo. Il nuovo protocollo DEEP-STD NMR, eseguito in D_2O e H_2O, può essere sfruttato per identificare i residui polari (ad esempio, l'arginina) nelle proteine che legano i GAG. Inoltre, se la struttura 3D del recettore è nota, è possibile ottenere informazioni sull'orientamento del ligando nel sito di legame. L'uso combinato di STD NMR, DEEP-STD NMR e simulazioni di MD ha permesso di analizzare le differenze nell'interazione tra due pentasaccaridi di eparina e l'AT. I nostri studi hanno rivelato che i suddetti oligosaccaridi, nonostante la notevole somiglianza strutturale, possiedono una diversa affinità per l'AT e presentano un distinto orientamento nella tasca dell'AT.

Abbreviations and acronyms

2OST	2-O-sulfotransferase
3OST	3-O-sulfotransferase
6OST	6-O-sulfotransferase
ACE2	Angiotensin converting enzyme 2
ARDS	Acute respiratory distress syndrome
AT	Antithrombin
CORCEMA	Complete relaxation and conformational exchange matrix analysis
CORCEMA-ST	Complete relaxation and conformational exchange matrix analysis of saturation transfer
COSY	Correlation spectroscopy
COVID-19	Coronavirus disease 2019
Cryo-EM	Cryo-electron microscopy
DEEP-STD	Differential epitope mapping by saturation transfer difference
DIC	Diffuse intravascular coagulation
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EXT1/EXT2	Exostosin-1/Exostosin-2
FID	Free induction decay
GAG	Glycosaminoglycan
Gal	Galactose
GalN	Galactosamine
GalT	Galactosyltransferase
GlcA	Glucuronic acid
GlcAT	Glucuronyltransferase
GLCE	Glucuronic acid epimerase
GlcN	Glucosamine
GlcNAc	N-acetyl-glucosamine
GlcNS	N-sulfated-glucosamine
HBD	Heparin binding domain
HBP	Heparin binding protein
HEPES	N-(2-hydroxyethyl) piperazine-N'-2-ethane sulfonic acid

HPSE	Heparanase
HP-SEC/TDA	High performance size exclusion chromatography with triple detector array
HS	Heparan sulfate
HSQC	Heteronuclear single quantum correlation
IdoA	Iduronic acid
INEPT	Insensitive nuclei enhancement by polarization transfer
ITC	Isothermal titration calorimetry
LMWH	Low molecular weight heparin
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
MD	Molecular dynamics
MM	Molecular mechanics
NDST	N-deacetylase/N-sulfotransferase
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PAPS	3'-phosphoadenosine 5'-phosphosulfate
PDB	Protein data bank
PME	Particle mesh Ewald
RBD	Receptor binding domain
RedMat	Reduced relaxation matrix
RMSD	Root-mean-square deviation
ROE	Rotating frame nuclear Overhauser effect
ROESY	Rotating frame Overhauser effect spectroscopy
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SEC	Size exclusion chromatography
SNFG	Symbol nomenclature for glycans
STD	Saturation transfer difference
Sulf	Endosulfatase
TOCSY	Total correlation spectroscopy
tr-NOE	Transferred nuclear Overhauser effect
tr-NOESY	Transferred nuclear Overhauser effect spectroscopy
tr-ROE	Transferred rotating frame nuclear Overhauser effect

tr-ROESY	Transferred rotating frame nuclear Overhauser effect spectroscopy
UDP	Uridine diphosphate
UFH	Unfractionated heparin
VDW	Van der Waals
Water-LOGSY	Water-ligand observed by gradient spectroscopy
WHO	World Health Organization
Xyl	Xylose
XylT	Xylosyltransferase

List of figures

List of tables

Index

Chapter 1

1. General introduction and objectives

1.1. Glycosaminoglycans: heparin and heparan sulfate

Glycosaminoglycans (GAGs) are long linear negatively charged polysaccharides composed of repeating disaccharide units.[1,2] Each unit consists of a hexuronic acid (or galactose) linked to a hexosamine (glucosamine or galactosamine) through a α or β (1-3 or 1-4) glycosidic bond. These monosaccharides can be differently decorated with acetyl and/or sulfate groups. On the basis of the composition and sulfation pattern of the disaccharide building blocks, GAGs can be classified into four groups: heparin/heparan sulfate, chondroitin sulfate/dermatan sulfate, keratan sulfate and hyaluronic acid (Figure 1.1). Except for the latter, they are covalently attached to various core proteins to form proteoglycans.[3,4]

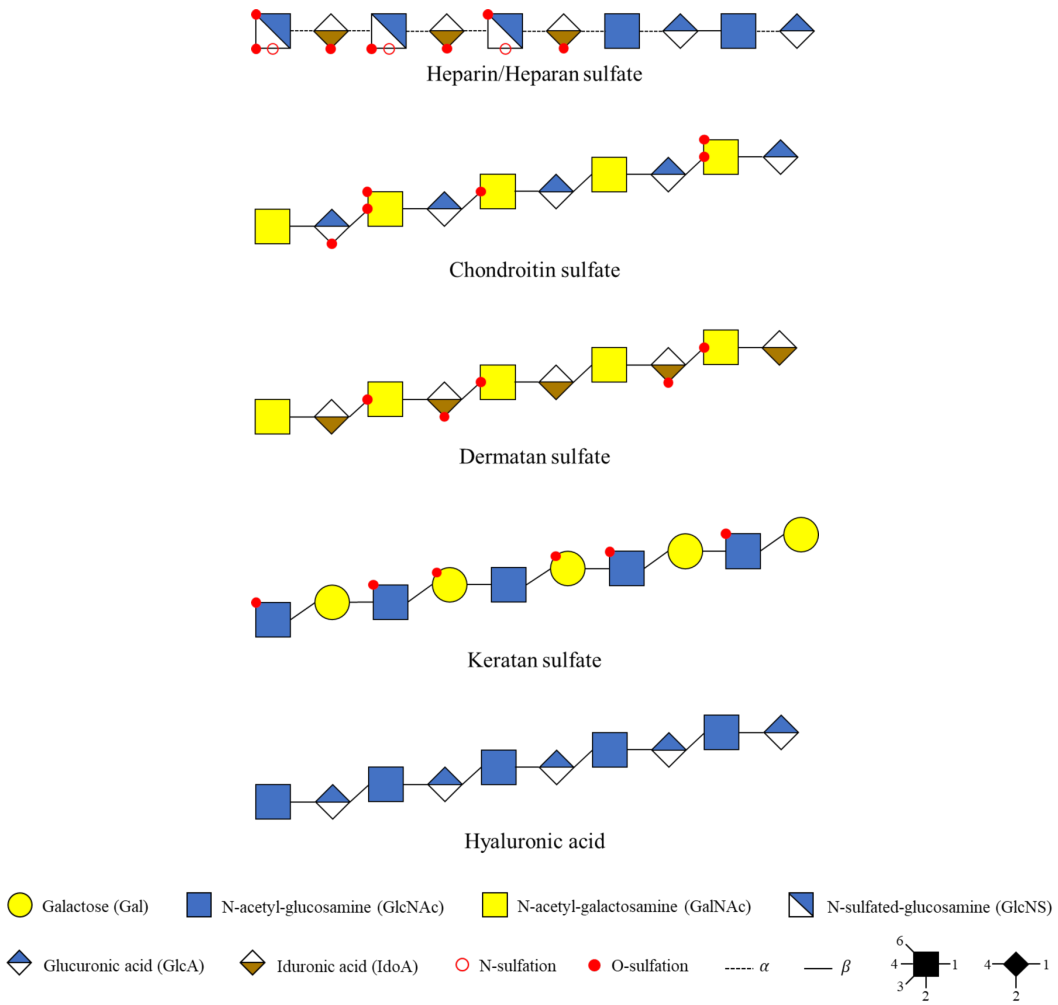

Figure 1.1. General structures of GAGs with their monosaccharide components. The above illustration is based on the symbol nomenclature for glycans (SNFG) representation.[5,6]

This research focuses on the GAG family of heparin and heparan sulfate (HS). Heparin is located in mast cell granules, while HS is ubiquitously expressed on cell surfaces and in the extracellular matrix (ECM). They are composed of repeating disaccharides that comprise a uronic acid (β-D-glucuronic acid or α-L-iduronic acid) (1-4)-linked to a α-D-glucosamine (Figure 1.2).[7-9] The uronic acids can be 2-O-sulfated, while the glucosamines are N-acetylated or N-sulfated and/or 6-O-sulfated and, more rarely, 3-O-sulfated.

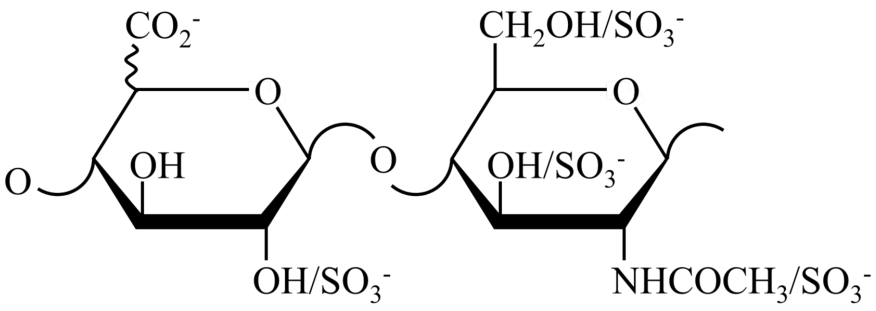

Figure 1.2. Disaccharide building block characterizing the GAG family of heparin and HS.

In solution, both glucosamine (GlcN) and glucuronic acid (GlcA) residues adopt a $^4C_1$ chair conformation, while the iduronic acid (IdoA) units exist in equilibrium between the $^4C_1$ chair, $^1C_4$ chair and $^2S_0$ skew-boat forms (Figure 1.3).[10-12] The population of each conformer depends on the degree of sulfation that characterizes both the iduronic acid itself and the adjacent glucosamines.[13]

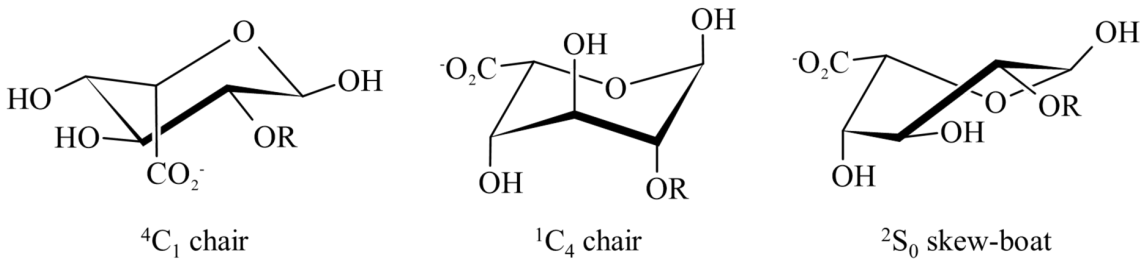

Figure 1.3. Conformations of the iduronate ring in the heparin/HS chains. IdoA exists in conformational equilibrium between the $^4C_1$ chair, $^1C_4$ chair and $^2S_0$ skew-boat forms.

While heparin possesses a high degree of sulfation, homogeneously distributed along its chain, HS presents highly sulfated clusters (S domains or NS domains) alternated to N-acetylated (NA domains) and poorly sulfated (NA/NS domains) regions.[14] The structural heterogeneity of heparin and HS, including the size of the polysaccharide chain, the ratio between GlcA and IdoA units and the amount/distribution of sulfate groups along the glycan backbone, results from variations in their biosynthesis.[15,16]

The biosynthesis of both heparin and HS, schematically outlined in Figure 1.4, starts in the endoplasmic reticulum and proceeds through the Golgi apparatus.[17] It initiates with the formation of a tetrasaccharide linkage region. A xylose (Xyl) unit is transferred to a serine residue on a core protein by xylosyltransferase (two similar isoforms, XylT1 and XylT2, exist). This attachment is followed by a stepwise transfer of two galactose (Gal) units and one GlcA unit mediated by galactosyltransferase-I (GalT1), galactosyltransferase-II (GalT2) and glucuronyltransferase (GlcAT1). The Xyl and Gal units can be 2-O-phosphorylated and 6-O-sulfated, respectively. Upon the addition of the first N-acetyl-glucosamine (GlcNAc) residue to the tetrasaccharide linkage region (Xyl-Gal-Gal-GlcA), the heparin/HS chain is elongated by adding alternating GlcA and GlcNAc residues. The polymerization process is carried out by the exostosin-1 (EXT1) and exostosin-2 (EXT2) complex; the final product is an extended polysaccharide. Since the formation of the glycosidic linkage is an energetically unfavorable process, the biosynthetic pathway requires uridine diphosphate (UDP) monosaccharides as sugar donors. As the HS/heparin chain grows, a number of modifications undergo through the action of specific enzymes, including sulfotransferases and epimerase. The availability of the sulfate donor 3'-phosphoadenosine 5'-phosphosulfate (PAPS) is fundamental for the sulfotransferase reactions. Both N-deacetylation and N-sulfation of GlcNAc to N-sulfated-glucosamine (GlcNS) are performed by N-deacetylase/N-sulfotransferase (NDST). Among the four isoforms of NDST, NDST1 is mainly responsible for HS synthesis, whereas NDST2 plays the major role in heparin synthesis. The C5-epimerization of GlcA to IdoA is mediated by GlcA epimerase (GLCE). Only the GlcA units linked at the non-reducing end to GlcNS are recognized as substrates.[18] This modification is tightly coupled to the 2-O-sulfation of the newly formed IdoA residue catalyzed by 2-O-sulfotransferase (2OST).[19] The 6-O-sulfation and 3-O-sulfation of the GlcNAc and GlcNS residues are carried out by 6-O-sulfotransferase (6OST) and 3-O-sulfotransferase (3OST), respectively. Three isoforms of 6OST and seven isoforms of 3OST exist. The 3-O-sulfate group is a crucial component of the heparin pentasaccharide sequence that specifically binds antithrombin (AT).[20]

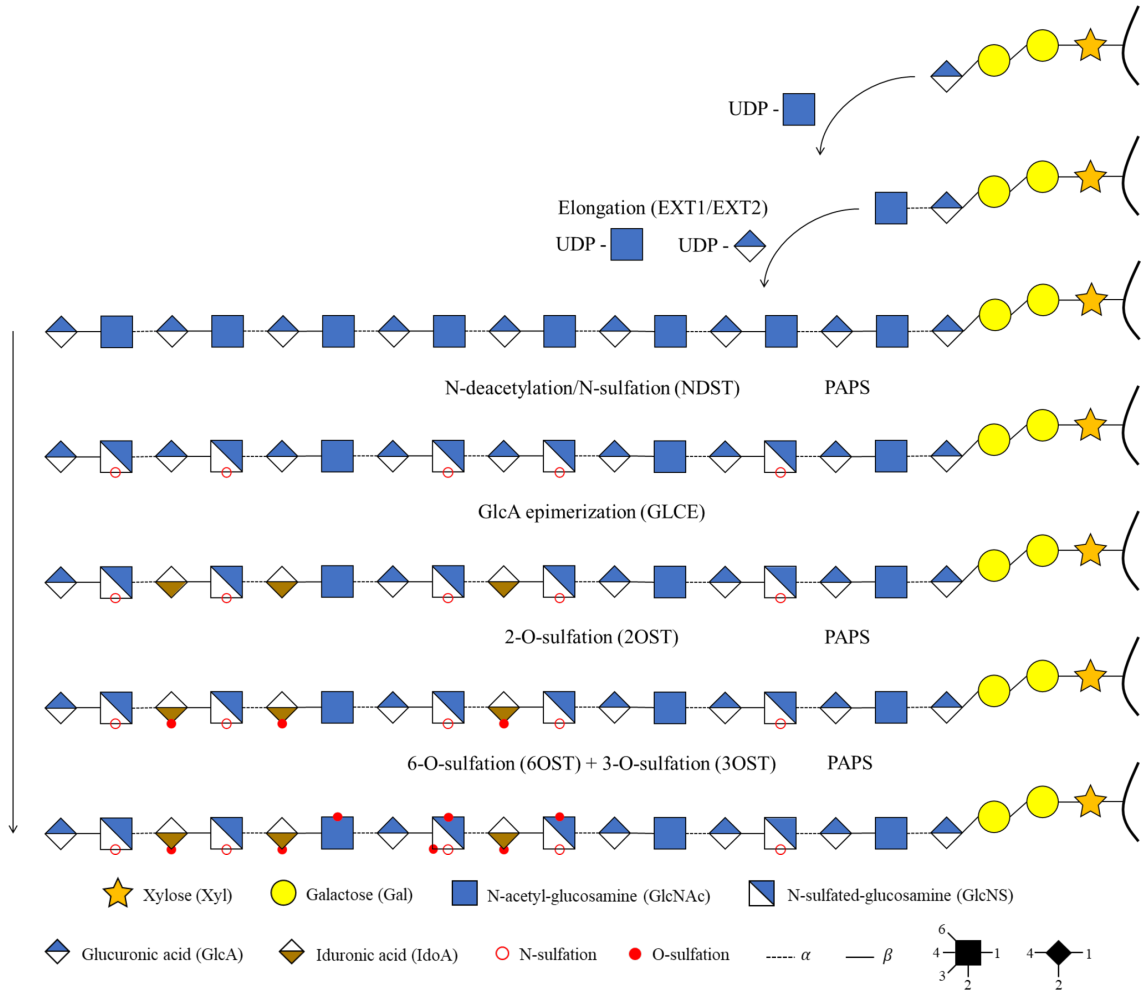

Figure 1.4. Heparin/HS biosynthesis. The growing polysaccharide is attached to a serine residue in a core protein. After the formation of the linkage region (Xyl-Gal-Gal-GlcA) and the attachment of the first GlcNAc, the polymerase complex (EXT1 and EXT2) adds alternating units of GlcA and GlcNAc from the respective uridine diphosphate (UDP) derivatives. In the presence of the sulfate donor PAPS, a series of modifications takes place: N-deacetylation and N-sulfation of the GlcNAc units, C5 epimerization of GlcA to IdoA, 2-O-sulfation of the uronic acids, 6-O-sulfation and 3-O-sulfation of the glucosamines.

Post-synthetic modifications, performed by two endosulfatases (Sulf1 and Sulf2) located at the cell surface, also occur. These enzymes act after the extracellular translocation of the heparin/HS chains by removing the 6-O-sulfate groups from the GlcNS residues.

Heparin and HS modulate key biological process by interacting with a plethora of proteins.[21,22] Endogenous heparin interacts with histamine, proteases and inflammatory mediators, controlling their storage, retention and activation. Exogenous heparin, derived from porcine and bovine mucosa, is used as an anticoagulant drug, due to its ability to bind and regulate the activity of several factors involved in the blood clotting cascade (Figure 1.5).[23,24] Specifically, it binds to AT and accelerates its inhibitory effect on blood clotting enzymes, particularly factor IIa (thrombin) and factor Xa.[25,26] This interaction, avoiding the

excessive formation of blood clots, is crucial in the prevention and treatment of thromboembolic disorders.

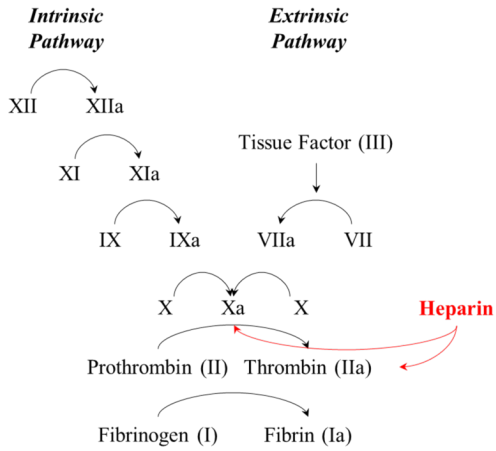

Figure 1.5. A simplified representation of the coagulation cascade and its pathways. Clotting factors are reported as Roman numerals; a lowercase "a" is added to indicate their active form.

Heparan sulfate (HS) interacts with growth factors, morphogens, chemokines, enzymes, ECM proteins, regulating their availability, distribution and signaling activity.[27,28] It modulates key biological processes such as cell growth, differentiation, adhesion, migration and apoptosis, affecting development and tissue homeostasis. The cleavage of the HS chains and the release of the HS-sequestered compounds can be correlated to diverse pathological conditions, such as inflammatory processes and cancer. This glycan serves also as a co-receptor for various pathogens, mediating the early stages of infection.
The multifaceted roles of these GAGs underscore their significance in maintaining normal physiological functions and offer potential avenues for therapeutic interventions.

1.2. Glycosaminoglycan-protein interactions

As mentioned in the previous paragraph, heparin and HS play a key role in various physiological and pathological processes by interacting with a wide array of proteins.[29,30] At the basis of the molecular recognition of heparin/HS by heparin binding proteins (HBPs), there are the electrostatic interactions between the negatively charged sulfate and carboxyl groups of heparin/HS and the positively charged arginine and lysine side chains of the corresponding HBPs.[31,32] It also involves van der Waals (VDW) forces, hydrogen bonds and hydrophobic interactions. Factors of the cellular environment, such as pH and ionic strength, can affect the stability of the heparin/HS-HBP complexes by altering the electrostatic

interactions.[33] The binding of heparin/HS to HBPs depends on several structural features characterizing these GAGs, such as disaccharide unit composition, monosaccharide ring conformation, glycosidic linkage, sulfation pattern and chain length.[34] These factors impact on the specificity, affinity and functional outcomes of the heparin/HS-HBP binding. The interaction of heparin/HS with HBPs is also influenced by the nature and architecture of the HBP active sites. Specific motifs on the HBP surface recognize distinct sequences and sulfation patterns on the polysaccharide chains, enhancing the specificity and affinity for certain heparin/HS structures. Heparin/HS and HBPs can undergo conformational changes to optimize their contacts. Specifically, the flexibility of the IdoA units allows heparin/HS chains to adopt various conformations, facilitating their interaction with different HBP sites.[35] The flexibility of the HBP surface influences the ability to accommodate heparin/HS structures (proteins with flexible regions are more likely to interact with GAGs that have high conformational freedom). Both heparin/HS and HBPs can have multiple binding sites. Multivalent interactions increase the overall strength of the binding and allow the formation of stable complexes. Understanding the mechanisms behind the molecular recognition of heparin/HS by HBPs is crucial to define the noteworthy role of heparin/HS as well as to lay the groundwork for the design of innovative glycomimetic drugs.

Nuclear magnetic resonance (NMR) spectroscopy is a versatile technique for studying GAG-protein interactions.[36] A plethora of NMR experiments can be employed to investigate the behaviour of a ligand before and upon its binding to a specific protein. In particular, ligand-based NMR approaches, including saturation transfer difference (STD) NMR and transferred nuclear Overhauser effect spectroscopy (tr-NOESY), allow to identify the binding epitope and the bioactive conformation of the ligand. Additionally, novel multifrequency and multisolvent STD NMR methodologies permit to gain information about the protein binding site, detecting the nature of the amino acid residues involved in the molecular recognition process. The combined use of NMR and molecular modelling enables the depiction of an accurate 3D model of the GAG-protein complexes, underlying the contacts and intermolecular forces that drive the interaction.[37] Among the computational methods, docking calculations can predict the conformation and orientation of GAGs in complex with proteins, while molecular dynamics (MD) simulations can provide insights into the dynamic behaviour of the GAG-protein complex over time. Advanced techniques, such as free energy calculations, can estimate binding affinities, contributing to the understanding of the thermodynamics of the interaction. NMR and computational studies involve an iterative process, where NMR data refine computational models and insights obtained from

simulations drive further experimental investigations. In general, NMR spectroscopy combined with molecular modelling can offer a holistic view of heparin/HS-HBP interactions.

1.3. Purposes of the project

This thesis dealt with the study of the molecular mechanisms underlying the interactions between heparin/HS and three HBPs, including the spike (S) protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), heparanase (HPSE) and AT. Key structural and conformational features of the recognition processes were obtained by applying NMR spectroscopy and molecular modelling.

The spike (S) protein plays a pivotal role in the early stage of SARS-CoV-2 infection.[38] A better understanding of the molecular aspects that drive the initial phase of infection is critical to enable the design of new antiviral drugs. It was recently demonstrated that SARS-CoV-2 infection begins with the recognition of HS by the receptor binding domain (S1-RBD), corresponding to the distal part of the S protein.[39,40] This event promotes the recruitment of viral particles to the cell surface and the binding of S1-RBD to the angiotensin converting enzyme 2 (ACE2) receptor, the preliminary step toward fusion of the virus-host cell membrane. Combining ligand-based NMR experiments and in silico methodologies, synthetic HS oligosaccharides with different length and composition were used as molecular probes to explore the binding of HS to S1-RBD. For an effective and prompt response to the coronavirus disease 2019 (COVID-19) public health emergency, the technical time to develop and commercialize ad hoc drugs could not be invested. Thus, already known drugs, used for other medical issues, found place against SARS-CoV-2 [*i.e.*, heparin and low molecular weight heparins (LMWHs)].[41] Among the physiopathological mechanisms of the new coronavirus, its capabilities to cause coagulopathy and pro-inflammatory state are the most important. As proven by several clinical trials, heparin and LMWHs find a role in both these processes.[42,43] Heparin and LMWHs, clinically approved as anticoagulant/antithrombotic agents, are also underexploited antiviral drugs with a broad-spectrum activity against Coronaviridae.[44] As potential antiviral agents, our interest was to pave the way for the possible enlargement of the therapeutic indications of heparin-related drugs against viral infections in general, and SARS-CoV-2 in particular.

Heparanase (HPSE) is an endo-β-D-glucuronidase that cleaves HS chains. The enzyme degrades HS in physiological conditions (HS turnover) and it is overexpressed in

pathological conditions such as inflammation, tumor angiogenesis and metastasis.[45] Glycol-split heparins, obtained by periodate oxidation followed by borohydride reduction and characterized by low anticoagulant activity, are promising inhibitors of HPSE activity.[46] An integrated STD NMR and molecular modelling approach was conducted to investigate the molecular recognition of synthetic glycol-split heparin oligosaccharide models by HPSE. Antithrombin (AT), a widely studied HBP, was chosen as a model protein and two synthetic structurally similar heparin pentasaccharides were selected as ligand probes to evaluate the effectiveness of the multisolvent STD NMR approach in determining the presence of polar residues (*i.e.*, arginine) in the binding sites of HBPs.[47,48] Additionally, ligand-based NMR techniques and computational methods were applied to analyse the differences in the interaction between the two heparin pentasaccharides and AT.

Chapter 2

2. Techniques

2.1. Nuclear magnetic resonance (NMR) spectroscopy

2.1.1. Basic concepts

The phenomenon of nuclear magnetic resonance (NMR) arises from the intrinsic property of spin exhibited by certain atomic nuclei.[49] This spin is quantified by a nuclear spin quantum number (I), which can assume values equal to or greater than zero in increments of 1/2 (I = 1/2, 1, 3/2, 2). Nuclei with I = 0 possess no spin; therefore, they are not observable by NMR spectroscopy. All nuclei carry an electric charge and the nuclei with non-zero spin also own an angular momentum (P). The rotational motion of this charge gives rise to a magnetic moment (μ) (Figure 2.1).

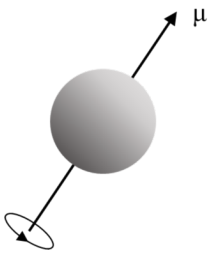

Figure 2.1. Nuclear magnetic moment (μ) of a spinning nucleus.

These nuclei, introduced into a static magnetic field, undergo a motion referred to as Larmor precession (Figure 2.2). The latter occurs at the Larmor frequency, which is directly proportional to the strength of the magnetic field (B_0) and depends on the gyromagnetic ratio (γ) of the nucleus. Mathematically, the Larmor frequency (ν) is expressed as:

$$\nu = \frac{\gamma B_0}{2\pi} Hz \qquad \text{(Equation 2.1)}$$

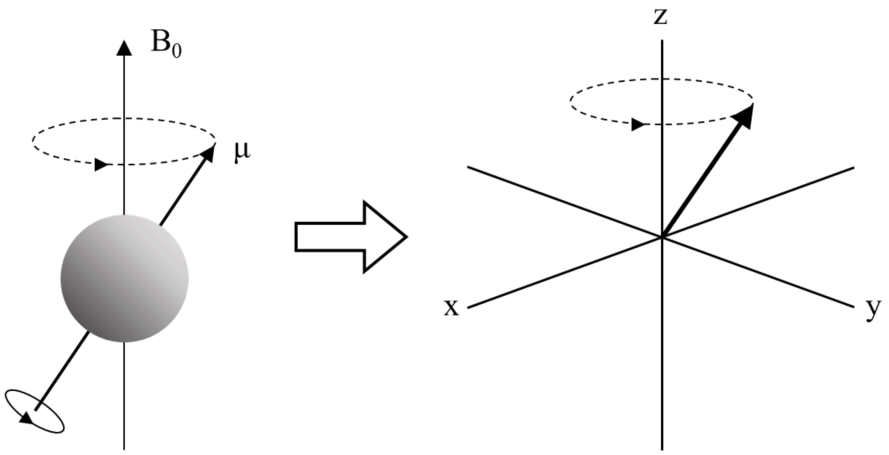

Figure 2.2. Larmor precession of a spinning nucleus placed into a static magnetic field (B_0). Conventionally, the magnetic field is applied along the z-axis of a Cartesian coordinate system and the motion of the nucleus is represented as a vector moving on the surface of a cone.

Since the energy states are quantized, the nuclei within the magnetic field (B_0) can adopt 2I+1 possible orientations. Those with I = 1/2 can align themselves in two ways, parallel or anti-parallel to the field B_0. The parallel orientation or α state is characterized by a lower energy than the anti-parallel orientation or β state. By applying an electromagnetic radiation oscillating at an appropriate frequency (the Larmor frequency), the lower energy α state can be excited to the higher energy β state, inducing the NMR phenomenon. Likewise, the β level can lose its excess energy and return to the α level (a process known as relaxation). The involved energy (ΔE) is given by:

$$\Delta E = h\nu = \frac{h\gamma B_0}{2\pi} \qquad \text{(Equation 2.2)}$$

where *h* is the Planck constant. At equilibrium, the population difference between the α and β states is determined by the Boltzmann equation:

$$\frac{N_\alpha}{N_\beta} = e^{\frac{\Delta E}{k_B T}} \qquad \text{(Equation 2.3)}$$

where N_α and N_β indicate the number of the nuclei in the α and β states, respectively, k_B the Boltzmann constant and T the temperature. Due to the small energy and population difference, NMR is a relatively insensitive spectroscopic method.

The vector model proves to be valuable in explaining basic NMR experiments. In the classical representation, a nucleus with I = 1/2 precesses around the z-axis, which corresponds to the direction of the static magnetic field (B_0), on the surface of a double cone

(Figure 2.3). Summing the z-components of all nuclei within a sample yields a bulk magnetization vector aligned along the +z-axis, which is parallel to B_0 as N_α is greater than N_β. This is referred to as longitudinal magnetization (M).

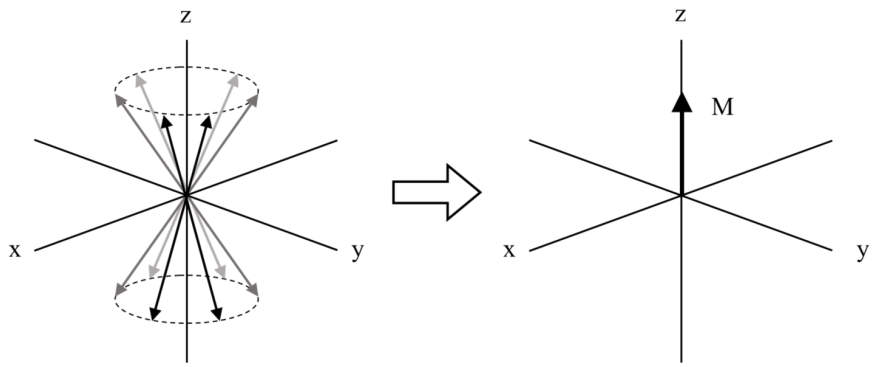

Figure 2.3. Bulk magnetization vector (M). In the vector model, many spins are represented by a bulk magnetization vector (M). At equilibrium, the excess of spins in the α state aligns this vector in parallel to the +z-axis.

The application of a second magnetic field (B_1), associated with a radiofrequency pulse (perpendicular to B_0), carries M towards the x-y plane, producing the transverse magnetization, which induces an oscillating voltage in the coil (Figure 2.4).

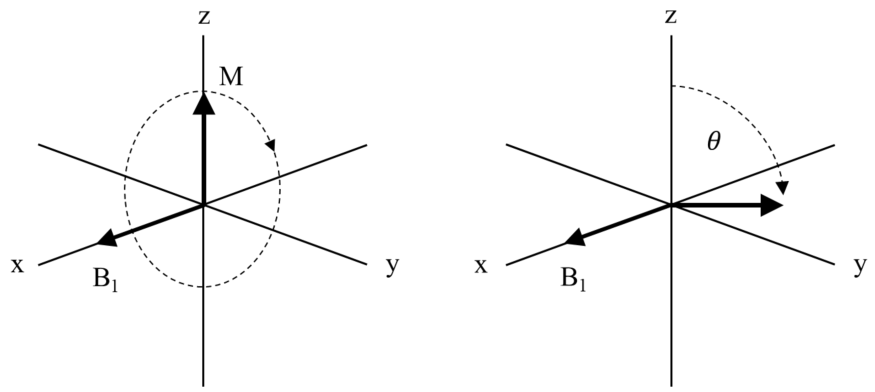

Figure 2.4. Radiofrequency (rf) pulse. A rf pulse applies a torque to the bulk magnetization vector (M), driving it towards the x-y plane. B_1 along the x-axis is the magnetic field associated with the rf pulse; θ is the pulse angle.

The transverse magnetization returns to the +z-axis by losing its excess energy. Therefore, the oscillating voltage decays away with time, determining the free induction decay (FID). The Fourier transform of the time domain FID produces the corresponding frequency domain spectrum, namely the NMR spectrum.

2.1.2. NMR techniques for structure elucidation

One-dimensional (1D) and two-dimensional (2D) NMR techniques are commonly used to elucidate the structure and dynamics of molecular systems.

In 1D NMR experiments, the signal is recorded as a function of one time variable and then Fourier transformed to obtain a spectrum which is a function of one frequency variable. The basic elements of a 1D NMR experiment are shown in Figure 2.5.

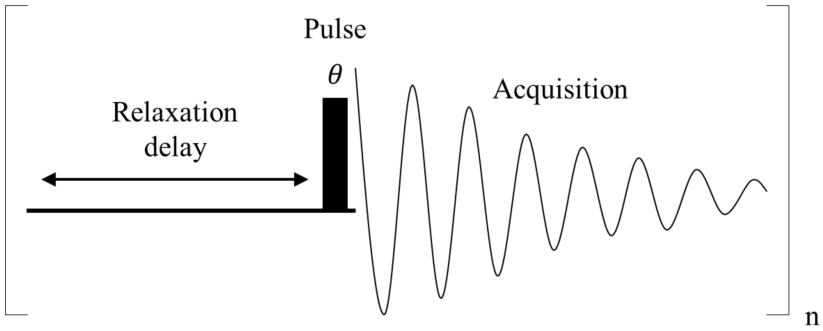

Figure 2.5. The essential elements of a 1D NMR experiment.

In 2D NMR experiments, the signal is acquired as a function of two time variables (t_1 and t_2) and then Fourier transformed to obtain a spectrum which is a function of two frequency variables. The general scheme for any 2D NMR experiment is displayed in Figure 2.6.

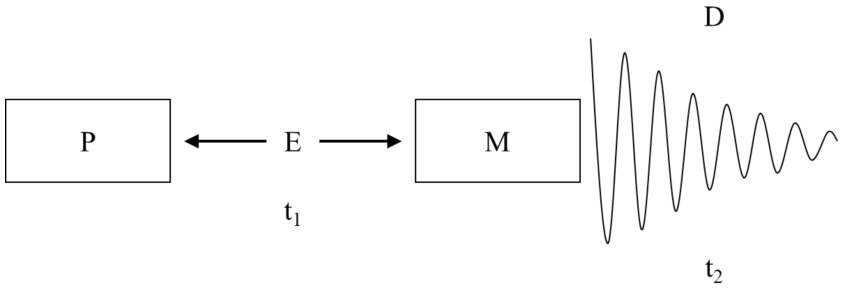

Figure 2.6. The general scheme for any 2D NMR experiment. The sequence comprises the preparation period (P), the evolution period (E), the mixing period (M) and the detection period (D); t_1 and t_2 are two time variables.

During the preparation time, one or more pulses are applied to excite the sample; the resulting magnetization evolves over a time interval denoted as t_1. Subsequently, during the mixing time, further pulse or pulses are employed. After this period, the signal is recorded as a function of a second time variable indicated as t_2. The peculiar features of the preparation and mixing phases dictate what information can be extracted from the 2D spectrum. A 2D NMR experiment allows to establish scalar or dipolar correlations between the same nucleus

(homonuclear correlation experiments) or different nuclei (heteronuclear correlation experiments).

This section provides a basic description of the 1D and 2D NMR techniques employed in this thesis to resolve the structure and dynamics of the molecules under study.

The proton (^{1}H) experiment is utilized to define the structural and conformational features of chemical compounds. This is accomplished by taking into account three essential variables: 1) chemical shifts (δ), measured in ppm, that reflect the local chemical environment of the proton; 2) spin-spin coupling constants (J), measured in Hz, that represent the interaction of one proton with others within a molecule through chemical bonds; 3) resonance intensities (integrals) that reveal the number of protons giving rise to each signal.

The correlation spectroscopy (COSY) experiment is recorded to identify nuclei which are coupled to each other. It is based on scalar coupling, also known as J-coupling. The basic COSY sequence consists of two 90° pulses interspaced by a specific evolution time (t_1) and followed by a detection period (t_2) (Figure 2.7).

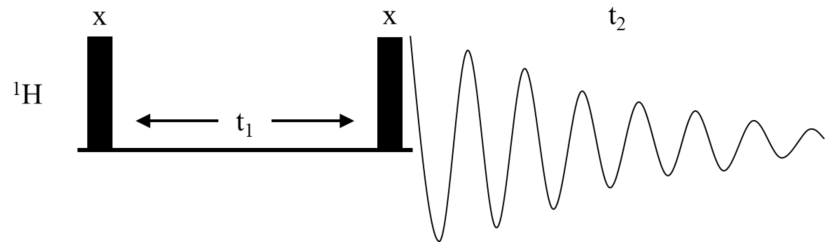

Figure 2.7. The basic COSY sequence. x indicates a 90° pulse; t_1 and t_2 are time variables.

The 2D COSY spectrum shows the frequencies of a single nucleus, usually hydrogen, along both axes. It displays two types of peaks: diagonal peaks and cross-peaks. Diagonal peaks correspond to the signals in the 1D spectrum; they appear along the diagonal of the plot and share the same frequency coordinates on each axis. A cross-peak occurs when there is a correlation between two signals of the spectrum along each of the two axes. Each coupling gives two symmetrical cross-peaks above and below the diagonal. Identifying cross-peaks among distinct signals leads to the assignment of adjacent protons.

The total correlation spectroscopy (TOCSY) experiment is similar to the COSY experiment. Unlike the latter, cross-peaks are observed for directly coupled nuclei connected by several covalent bonds. This capability is achieved by inserting a repetitive series of pulses, called spin-lock, which causes an isotropic mixing during the mixing time. Figure 2.8 reports the TOCSY sequence.

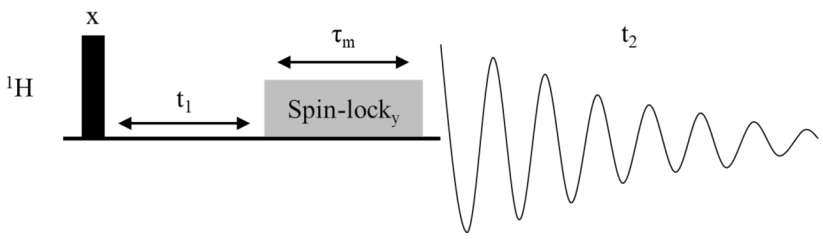

Figure 2.8. The TOCSY sequence. The spin-lock mixing time (τ_m) replaces the single mixing pulse of the COSY experiment (Figure 2.7).

The heteronuclear single quantum correlation (HSQC) experiment is acquired to detect correlations between two different nuclei which are directly bound. It works by transferring magnetization from the nucleus I (usually ^{1}H) to the nucleus X (usually a heteroatom) using the insensitive nuclei enhanced by polarization transfer (INEPT) sequence. After a time delay (t_1), the magnetization is transferred back to the first nucleus via an INEPT step in reverse to be detected. The HSQC sequence is displayed in Figure 2.9.

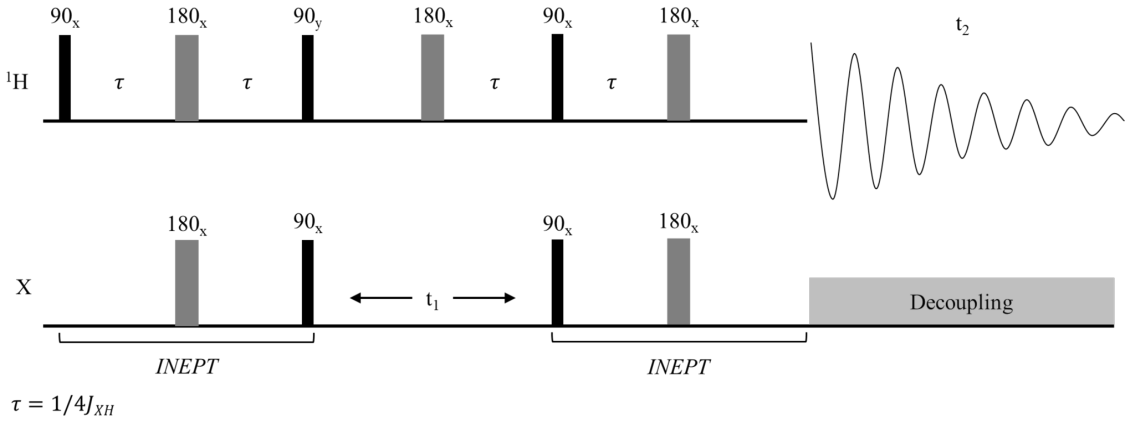

Figure 2.9. The HSQC experiment with the INEPT sequence to generate a transverse magnetization X, which first evolves and then is transferred back to ^{1}H by an INEPT step in reverse.

The HSQC spectrum shows one peak per pair of coupled nuclei, whose two coordinates are the chemical shifts of the two connected atoms.

The nuclear Overhauser effect spectroscopy (NOESY) experiment provides information on conformation and three-dimensional (3D) structure. It is carried out to detect and quantify dipolar interactions, also known as nuclear Overhauser effects (NOEs), between pairs of nuclei that are in close spatial proximity, typically with a distance less than 5 Angstroms (Å). The NOESY sequence is similar to the COSY sequence, except for the mixing time (τ_m), during which the NOEs develop, and an additional 90° pulse (Figure 2.10).

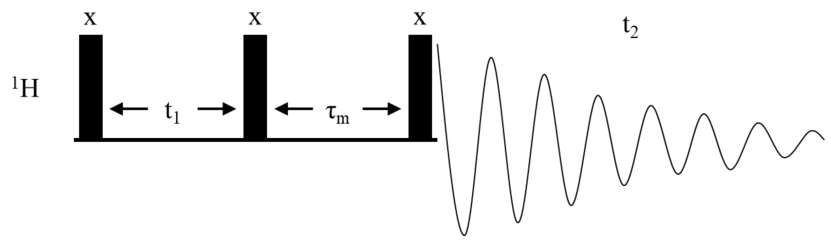

Figure 2.10. The NOESY sequence. x indicates a 90° pulse; t_1, t_2 and τ_m are time variables.

The NOE phenomenon arises when a specific resonance is saturated by a radiofrequency pulse and the saturation leads to changes in the intensities of other resonances associated with nearby protons. The magnitude of the NOE, observed for proton A when proton B is perturbed ($\eta_A\{B\}$), can be quantified by comparing the intensity of proton A in absence (I_0) and in presence (I) of the NOE:

$$\eta_A\{B\} = \frac{I - I_0}{I_0} \cdot 100(\%) \qquad \text{(Equation 2.4)}$$

Notably, it is inversely proportional to the 6th power of the distance between the interacting protons. An unknown distance between two protons A and B (r_{AB}) can be estimated from the NOE intensity ($\eta_A\{B\}$) using a known reference distance (r_{XY}) and the corresponding NOE intensity ($\eta_X\{Y\}$):

$$\frac{\eta_A\{B\}}{\eta_X\{Y\}} = \frac{r_{AB}^{-6}}{r_{XY}^{-6}} \qquad \text{(Equation 2.5)}$$

This formula is valid for short build-up times, when spin diffusion is insignificant, and if the molecule tumbles rigidly and isotropically. The sign of the NOE can be positive or negative, depending on the molecular tumbling rate, which is correlated to the molecular weight characterizing the molecule. Specifically, small molecules tumble rapidly and produce weak positive NOEs, whereas large molecules rotate slowly and induce strong negative NOEs (Figure 2.11). For medium-sized molecules with NOE values close to zero, the rotating frame Overhauser effect spectroscopy (ROESY) experiment is preferred because the rotating frame NOE (ROE) is always positive (Figure 2.11).

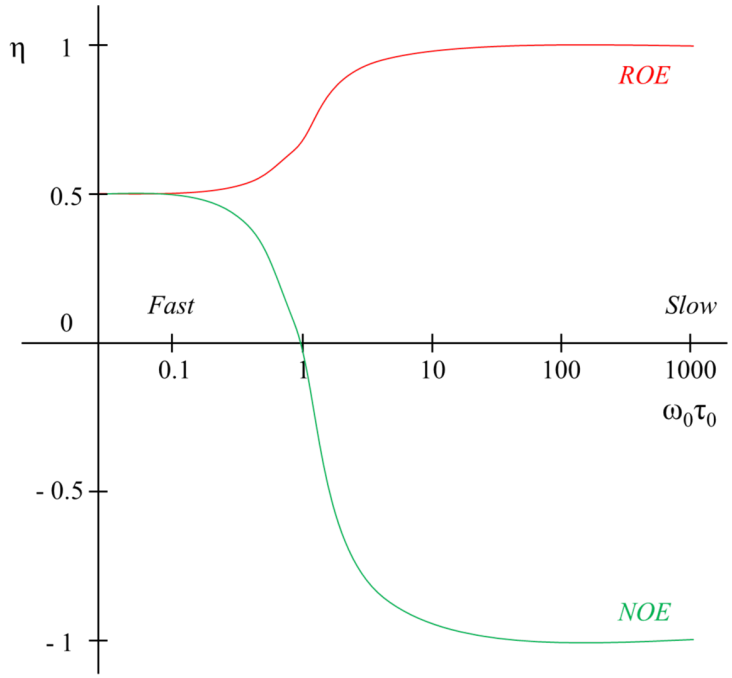

Figure 2.11. Variation in the NOE/ROE intensities (η) as a function of the molecular tumbling rate ($\omega_0\tau_c$).

2.1.3. NMR methods for studying protein-ligand interactions

Protein-ligand interactions are essential for several biological processes as the proper functioning of a protein often requires the recognition of a specific ligand at a particular binding site on its surface. NMR spectroscopy is a powerful tool for characterizing protein-ligand interactions in solution, providing a wide set of experiments that are optimized for various objectives (*i.e.*, identification of protein and ligand binding epitopes, characterization of structural rearrangements induced by binding).[50,51] Additionally, it is not limited to high affinity systems, but it can also be applied to study weak interactions. NMR experiments for studying protein-ligand interactions are divided into two categories: protein-based and ligand-based NMR methods. Protein-based NMR approaches are based on understanding the interaction mechanism from the macromolecule point of view. Ligand-based NMR techniques provide insights into the binding event from the perspective of the small molecule. They rely on changes of the NMR parameters of the ligand when it binds to a protein. In the free state, a ligand exhibits small transverse relaxation rates, weakly positive cross-relaxation rates and large diffusion constants. Upon binding to a protein, it transiently acquires the NMR properties of the receptor, including high transverse relaxation rates, strong negative cross-relaxation rates and small diffusion constants. Ligand-based NMR experiments are useful for investigating moderate or weak binding processes, where the off-rate is fast on the NMR timescale. Their primary advantage is that they require small

quantities of unlabeled proteins. Ligand-based NMR methods include saturation transfer difference (STD) NMR, differential epitope mapping by STD NMR (DEEP-STD NMR), water-ligand observed by gradient spectroscopy (water-LOGSY) and transferred NOESY (tr-NOESY). These techniques focus on the NMR signals of the ligand and permit to identify the binding epitope and the bioactive conformation of the ligand. The following paragraphs describe the basics of the ligand-based NMR approaches used in this thesis.

2.1.3.1. Saturation transfer difference (STD) NMR

Saturation transfer difference (STD) NMR technique is employed to verify if the interaction of a ligand to its receptor occurs.[52,53] It is also carried out to determine the ligand moieties in contact with the receptor and their proximity to the protein surface, defining the so-called binding epitope. Additionally, it can be used to obtain the dissociation constant (K_d).[54] The method is applicable to fast exchange systems, such as ligands with medium to weak affinity to their receptors, which exhibit a dissociation constant (K_d) in the millimolar to micromolar range. In the STD NMR experiment, two spectra are acquired on a sample containing a protein:ligand molar ratio of 1:10 to 1:1000 (Figure 2.12). The off-resonance experiment is performed by irradiating far from protein and ligand frequencies. In the off-resonance or reference spectrum, the ligand signals do not show a decrease in their intensities. In the on-resonance experiment, the protein is selectively saturated. The saturation spreads over the entire protein via intramolecular NOE and is transferred to the bound ligand via intermolecular NOE. Ligand protons in contact to the binding site receive saturation with a consequent decrease in signal intensity; while ligand protons far from the protein surface do not receive magnetization and show no reduction in signal intensity. To overcome possible artifacts in the difference spectrum, the on-resonance and off-resonance experiments are acquired in an interleaving manner to reduce the impact of equipment instabilities. The STD NMR spectrum, obtained by subtracting the on-resonance spectrum from the off-resonance spectrum, shows only the signals of the ligand protons involved in the binding process. Interestingly, the saturation degree received by protons of the ligand is not equal but depends on their proximity to the protein. Specifically, the larger the STD response, the closer protein-ligand contact. The STD intensities (I_{STD}) are calculated as follows:

$$I_{STD} = \frac{I_0 - I_{sat}}{I_0} \qquad \text{(Equation 2.6)}$$

where I_0 and I_{sat} are the intensities of a signal in the off-resonance and on-resonance spectra, respectively. The normalization of the STD intensities against the most intense signal, which is assigned a value of 100%, provides the relative STD percentages. The binding epitope can be estimated by exploiting the differences in the STD intensities among different protons of the ligand. The initial growth rate approach is recommended for the epitope mapping because it allows to define the STD intensities at the saturation time of 0 (STD_0), minimizing artifacts associated to re-binding phenomena and differences in the relaxation time of the ligand protons. It is based on measuring the STD intensities at different saturation times and fitting the STD build-up curves as follows:

$$STD(t_{sat}) = STD_{max} \cdot (1 - e^{-k_{sat} \cdot t_{sat}}) \qquad \text{(Equation 2.7)}$$

where $STD(t_{sat})$ is the STD intensity at a specific saturation time (t_{sat}), STD_{max} is the asymptotic maximum of the curve and k_{sat} is a rate constant that quantifies the speed of the curve. After determining the STD_{max} and k_{sat} values, the initial slope of the STD build-up curves (STD_0) is calculated by the following function:

$$STD_0 = \lim_{t_{sat} \to 0} \frac{\delta STD(t_{sat})}{\delta t_{sat}} = STD_{max} \cdot k_{sat} \qquad \text{(Equation 2.8)}$$

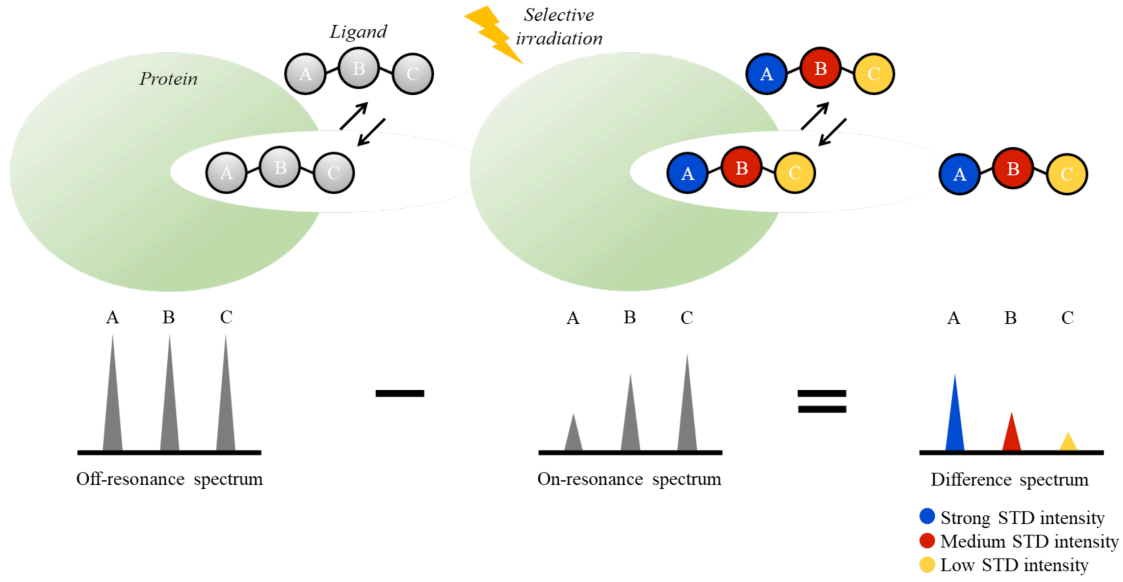

Figure 2.12. Scheme of the STD NMR experiment.

Complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST) and reduced relaxation matrix algorithm (RedMat) can be applied to

predict the theoretical STD values of a protein-ligand complex.[55,56] By comparing calculated and measured STD NMR data, it is possible to validate 3D models.

2.1.3.2. Differential epitope mapping by STD NMR (DEEP-STD NMR)

A novel STD NMR methodology, called differential epitope mapping by STD NMR (DEEP-STD NMR), was developed by Angulo and co-workers to gain information about the type of amino acids (aliphatic, aromatic or polar) surrounding the ligand in the binding site.[47] The new approach exploits the differences in the ligand epitope mapping that arise from acquiring a pair of STD NMR experiments under two different conditions, such as different frequencies or solvents. The multifrequency STD NMR, conducted by irradiating the aliphatic and aromatic residues of the protein, allows to define the aliphatic or aromatic nature of the amino acid side chains in contact with the ligand. Ligand protons close to directly irradiated protein residues receive a higher degree of saturation, showing a relative increase in their STD intensities. The multisolvent STD NMR, performed in D_2O and H_2O, permits to determine the presence of polar residues in the binding site (Figure 2.13). In D_2O, the polar residues present their exchangeable protons replaced by deuterium, which is unable to transfer saturation. In H_2O, the aforementioned protons may allow an additional transfer of saturation, depending on their exchange rate with the bulk water (Figure 2.14).[57] Specifically, slow exchanging protein protons (*i.e.*, protons belonging to the guanidinium group of arginine residues) may contribute to transfer saturation, leading to a relative increase in the STD intensities of the ligand protons in contact with them in H_2O. If fast exchanging protein protons are isolated, they have similar effect on the STD intensities of the ligand protons adjacent to them in D_2O and H_2O. Instead, when they are close to non-exchangeable protein protons, the magnetization of the latter may be lost due to an exchange-mediated leakage, determining a relative decrease in the STD intensities of the ligand protons in proximity to the above-mentioned non-exchangeable protein protons in H_2O. In general, this approach could provide the nature of the binding site and, if the 3D structure of the protein is known, the orientation of the ligand in the binding pocket.

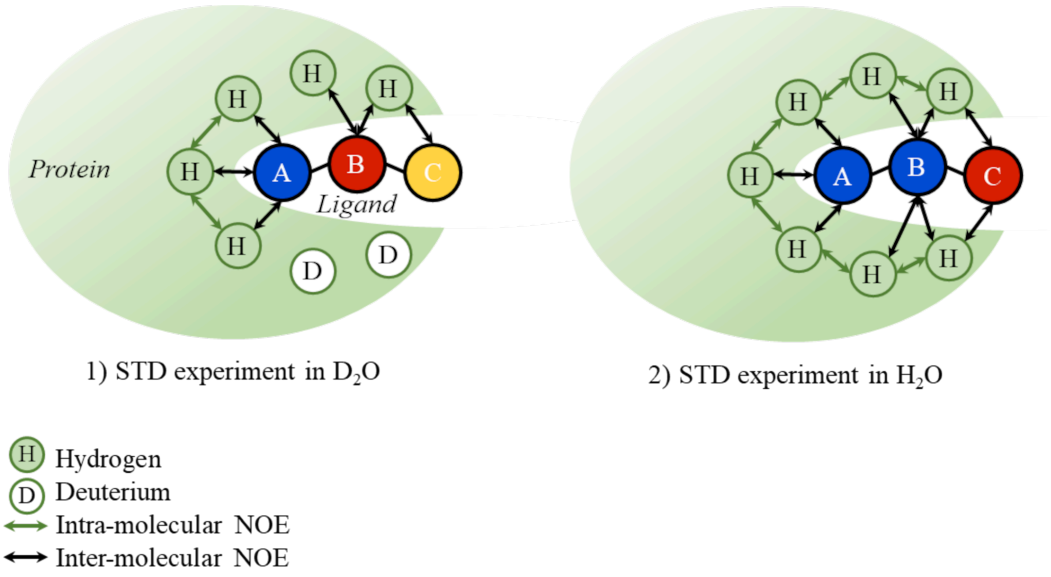

Figure 2.13. Illustration of the DEEP-STD NMR protocol in different solvents (D_2O and H_2O).

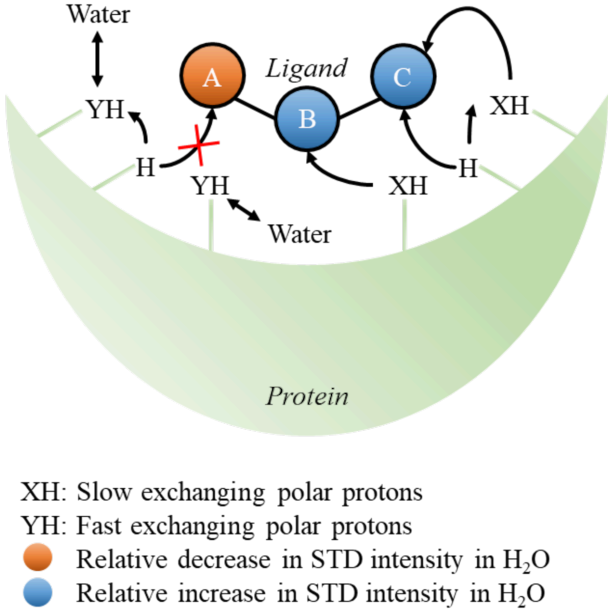

Figure 2.14. Cartoon representing the different pathways for saturation transfer from slow and fast exchanging polar protons of the protein to the bound ligand in H_2O.

2.1.3.3. Water-ligand observed by gradient spectroscopy (water-LOGSY)

Water-ligand observed by gradient spectroscopy (water-LOGSY) technique is a variant of STD NMR exploited to detect the interaction of ligands for their receptors.[58] It also enables the evaluation of the solvent accessibility of a ligand bound to its receptor and the identification of resident water molecules within the protein binding site. As STD NMR, water-LOGSY is suitable to analyse weak affinity ligands in fast exchange with their receptors. The method is especially useful for studying complexes where either ligand or

receptor are strongly hydrated because the magnetization originates from the bulk water. After the selective irradiation of the bulk water, the magnetization is transferred to the protein, via intermolecular NOE and chemical exchange with labile hydrogens, and consequently to the bound ligands. The water-LOGSY spectrum shows the proton signals of the interacting and non-interacting compounds with opposite sign (Figure 2.15).

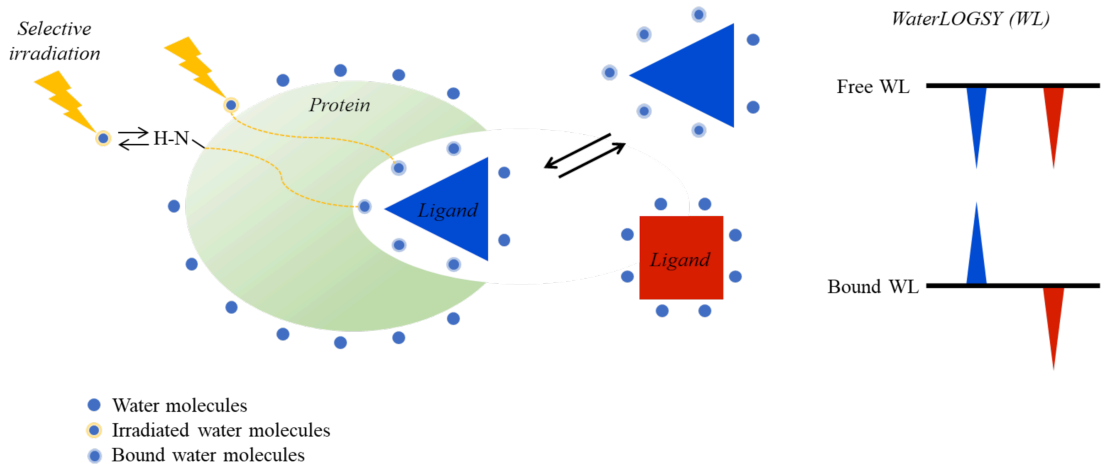

Figure 2.15. Scheme of the water-LOGSY experiment.

2.1.3.4. Transferred nuclear Overhauser effect spectroscopy (tr-NOESY)

Transferred NOESY (tr-NOESY) experiment is employed to investigate the conformational arrangement of bound ligands, contributing to the determination of the bioactive conformation.[59] The observation of transferred NOE (tr-NOE) relies on the different correlation times (τ_c) exhibited by ligands in their free and bound states. Depending on molecular weight, shape and magnetic field strength, small/medium entities possess a short τ_c, leading to positive, absent or minimal negative NOEs. Conversely, large molecules display negative NOEs. When a small molecule binds to a macromolecule, it assumes the behaviour of the latter, showing strong negative NOEs, defined as tr-NOEs. These reflect the conformation of the ligand in its bound state. Protein-ligand interactions can be detected by looking at both sign and size of the observed tr-NOEs. Additionally, the discrimination between NOEs of a free ligand in solution and tr-NOEs originating from a bound ligand can also be achieved by the build-up rate (the time required to achieve the maximum intensity), which is longer for NOEs than tr-NOEs. The method is schematically represented in Figure 2.16. Transferred ROESY (tr-ROESY) experiment can be carried out as an alternative to tr-NOESY. It is used to analyse cross-peaks affected by spin diffusion.

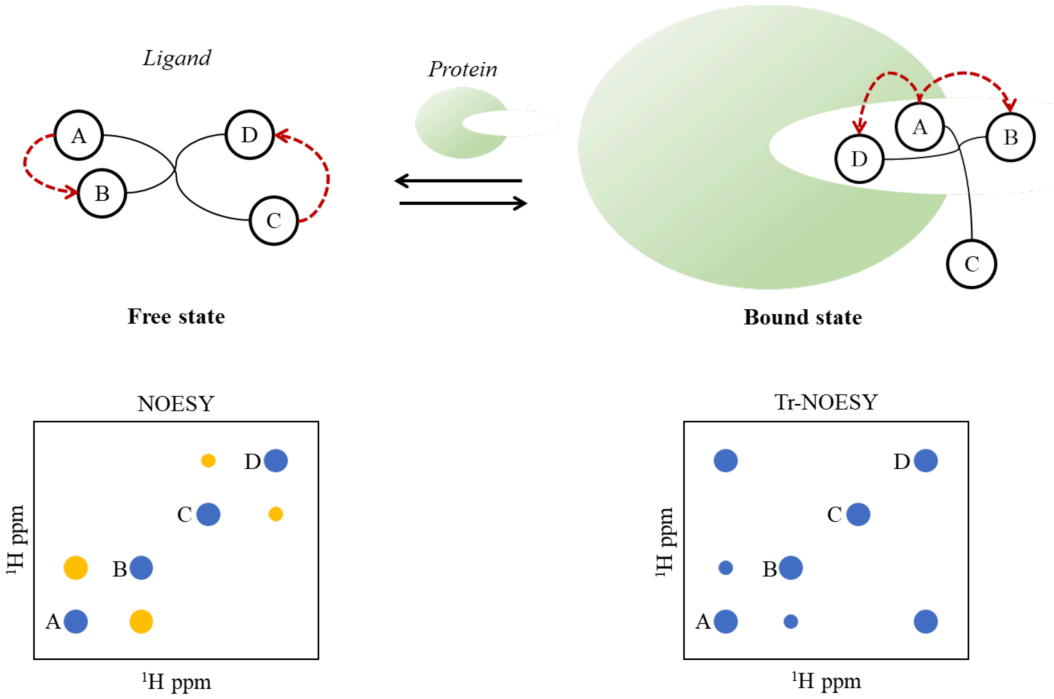

Figure 2.16. Schematic representation of NOE (left) and tr-NOE (right) effects. In the free state, small molecules exhibit positive NOEs. In a NOESY spectrum, cross-peaks have opposite sign to the diagonal peaks. In the bound state, small molecules assume the behaviour of the large receptor, displaying negative NOEs. In a tr-NOESY spectrum, the NOE turns negative.

Complete relaxation and conformational exchange matrix analysis (CORCEMA) can be employed to simulate NOE/tr-NOE values of a ligand before and upon binding to a specific protein.[60]

2.2. Molecular modelling

2.2.1. Fundamental principles of molecular mechanics (MM) and molecular dynamics (MD)

Molecular mechanics (MM) and molecular dynamics (MD) are computational techniques exploited to investigate the structural, conformational and dynamic features of molecular systems.[61]

Molecular mechanics (MM) employs force fields to describe the covalent and non-covalent bond forces acting on a molecular system. These force fields are empirical equations that describe the potential energy $[U(\vec{R})]$ of a system as a function of the coordinates $(\vec{R})$ of its atoms:

$$U(\vec{R}) = \sum_{bonds} k_b(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\phi[1 + \cos(n\phi + \delta)]^2 + \sum_{Coul_{i,j}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \sum_{LJ\ i,j} 4\varepsilon_{i,j}\left[\left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{12} - \left(\frac{\sigma_{i,j}}{r_{i,j}}\right)^{6}\right] \quad \text{(Equation 2.9)}$$

In this description, $\vec{R}$ = (x_1, y_1, z_1; x_2, y_2, z_2…x_N, y_N, z_N) is the position vector of the atoms within the system. The molecular forces $(\vec{F})$ are obtained by the gradient of the potential energy:

$$\vec{F} = -\vec{\nabla}U(\vec{R}) \quad \text{(Equation 2.10)}$$

The functional form of a force field includes both bonded terms (for interactions between covalently bonded atoms) and non-bonded terms (for interactions between non-covalently bonded atoms). The former reproduce the stretching, bending and torsional degrees of freedom (stretch terms, bend terms and dihedral terms). The latter are related to electrostatic forces and van der Waals (VDW) interactions. Electrostatic forces ($Coul_{i,j}$) result from the interaction between fully or partially charged groups and are described by Coulomb's law. In Equation 2.9, q_i and q_j are force field parameters describing the partial charges associated to the atoms *i* and *j*. Van der Waals (VDW) interactions ($LJ_{i,j}$) are short-range forces between electrons clouds generated by instant fluctuations in electron densities. They are classically interpreted as attractive/repulsive interactions between dynamic induced-dipoles. The Lennard-Jones potential, such as the 12-6 potential, is empirically used to describe these interactions; ε_{ij} and σ_{ij} are force field parameters representing the potential energy deep and the VDW radius acting between the atom pair *i* and *j*. An accurate representation of the molecular systems depends on a proper parameterization of the force fields and the number of terms included in Equation 2.9. Force field parameters, including equilibrium bond length (r_0), bond angle (θ_0), force constants (k_b, k_θ, k_ϕ), partial charges (q_i, q_j) and VDW parameters (ε_{ij}, σ_{ij}), are determined through a combination of experimental data and quantum mechanical calculations. The selection of an appropriate force field relies on the molecular system under investigation and the surrounding environment (solution, amorphous or crystalline solid state). MM is frequently used to optimize the geometry of a system by

minimizing its energy. This procedure involves guessing an initial molecular geometry and iteratively adjusting the coordinates of the atoms until a local energy minimum is found.
Molecular dynamics (MD) simulations are based on Newton's second law, which relates the forces acting on atoms $(\vec{F})$ to their accelerations $(\vec{a})$:

$$\vec{F} = m \cdot \vec{a} \qquad \text{(Equation 2.11)}$$

Given the acceleration at time 0 (t_0), the position of the atoms within a system at a sufficiently small subsequent time ($t_0 + \Delta t$) can be predicted through the velocity Verlet algorithm:

$$\vec{R}(t_0 + \Delta t) = 2\vec{R}(t_0) - \vec{R}(t_0 - \Delta t) + \frac{d^2\vec{R}(t_0)}{d^2 t_0}\Delta t^2 \qquad \text{(Equation 2.12)}$$

After setting the temperature, the simulations begin with the assignment of the initial velocities to the atoms of the system under investigation. These velocities are sampled according to the Maxwell distributions, depending to the set initial temperature. The equations of motion are then solved iteratively to predict atoms positions and velocities at subsequent time steps ($t_0 + \Delta t$) by numerical integrations; the result is the numerical trajectory in positions r(t) and velocities v(t) of the atoms of the system. Several numerical integration algorithms are available to update atom positions r(t) and velocities v(t) at each time step. MD simulations can be conducted under different statistical ensembles, such as the NVE ensemble (constant number of particles, volume and energy), the NVT ensemble (constant number of particles, volume and temperature) and the NPT ensemble (constant number of particles, pressure and temperature). MD provides insights into the dynamic behavior of molecules, including molecular motions, conformational changes and thermodynamic properties.
Molecular mechanics (MM) and molecular dynamics (MD) have their own strengths and limitations. Both MM and MD relies on classical approximations and may not accurately consider quantum effects or electronic properties (*i.e.*, the partial charges depend on the conformation and cannot be established a priori as described in Equation 2.9). The MD simulations are computationally demanding and limited by the time scales that can be realistically simulated within a reasonable timeframe.

Chapter 3

3. Role of glycosaminoglycans in SARS-CoV-2 infection

3.1. Molecular recognition of cell surface glycans by SARS-CoV-2 spike protein

3.1.1. Introduction

3.1.1.1. SARS-CoV-2 infection

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was the etiological agent responsible for the coronavirus disease 2019 (COVID-19), emerged in China at the end of 2019 and declared as a pandemic by the World Health Organization (WHO) in March 2020.[38,62] COVID-19 is characterized by a wide spectrum of clinical severity, ranging from asymptomatic or pre-symptomatic infection to severe or critical illness.[63] Primarily, COVID-19 can affect the respiratory tract, leading to pneumonia and acute respiratory distress syndrome (ARDS). Due to these complications, patients often require intensive medical care. Additionally, COVID-19 can extend beyond the respiratory system; indeed, it can induce hyper-inflammation which, when combined with hypoxia and diffuse intravascular coagulation (DIC), increases the vulnerability of patients to both venous and arterial thromboembolic diseases.[64] These symptoms can persist after the active phase of infection, giving rise to what is commonly known as Long COVID.[65] Vaccines have not a complete efficacy; moreover, their production and distribution may not be fast enough to counteract the multiple variants that regularly occur. Therefore, the design and development of new antiviral drugs represents an alternative strategy for both treatment and prophylaxis of the disease. In this context, a better understanding of the molecular aspects that drive the initial stage of SARS-CoV-2 infection is fundamental.

3.1.1.2. SARS-CoV-2 spike protein

SARS-CoV-2 infection begins with the interaction between the spike (S) protein and the angiotensin converting enzyme 2 (ACE2) receptor located on the host cell surface (Figure 3.1.1).[66,67] This event initiates the viral internalization mechanism, which leads to fusion between the virus and host cell membranes. Evidence suggests that heparan sulfate (HS) acts as a co-receptor, promoting the binding of the S protein to ACE2.[39]

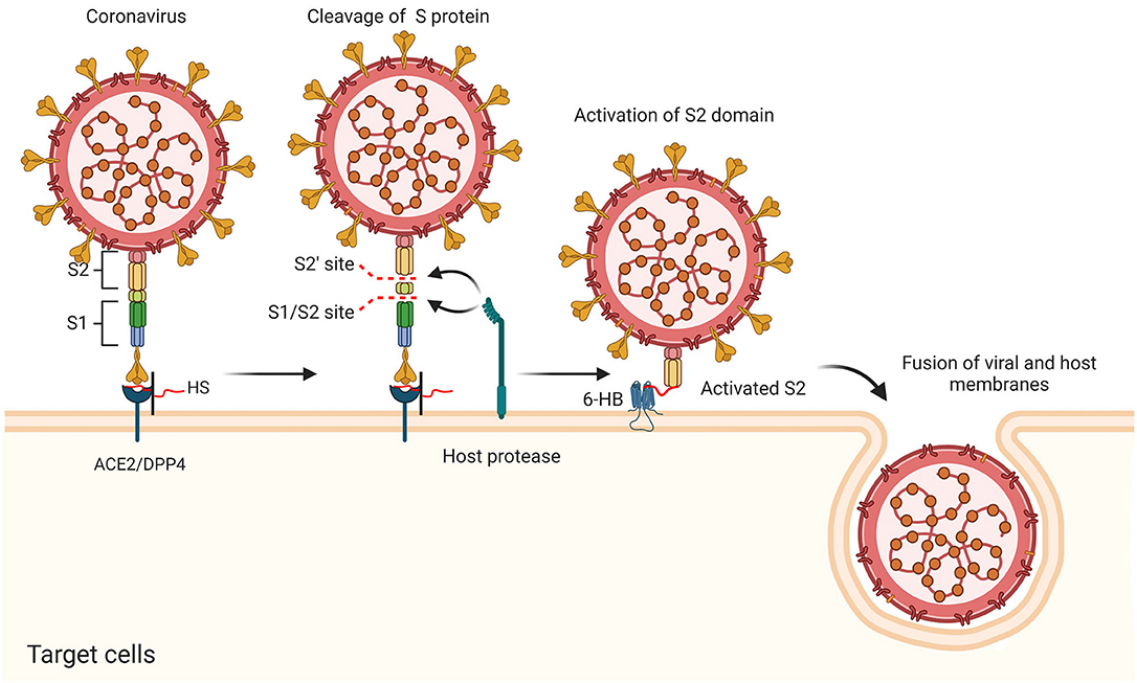

Figure 3.1.1. Initial steps of SARS-CoV-2 infection.[68]

The S protein of SARS-CoV-2 is a class I membrane fusion protein, consisting of a large N-terminal ectodomain, a single-pass transmembrane anchor and a short C-terminal tail located in the interior of the viral particle.[69,70] It is a glycoprotein that assembles to form homotrimers on the surface of the virus, giving it a characteristic crown-like morphology. The coat of sugars which covers the S protein hides it to the immune system of the host organism. Each protomer is composed by two subunits: the subunit 1 (S1), which allows the binding of the virus to the host cell, and the subunit 2 (S2), which drives the fusion process between viral and cellular membranes. The subunit S1 includes the N-terminal domain (NTD), the receptor binding domain (RBD) and two subdomains (SD1 and SD2); it forms a globular head protruding from the envelope. The subunit S2 comprises the N-terminal hydrophobic fusion peptide (FP), two heptad repeats (HR1 and HR2), the transmembrane domain (TM) and the cytoplasmic tail (CT); it constitutes a stalk-like region connecting S1 to the intracellular domain. The receptor binding domain of S1 (S1-RBD), corresponding to the distal part of the S protein, is mainly involved in the interaction with the ACE2 receptor.[71] It can assume two conformations: a down or receptor-inaccessible conformation and an up or receptor-accessible conformation (Figure 3.1.2). The binding to the HS co-receptor leads the conformational change from the down to the up state, where the ACE2 recognition interface is exposed, allowing the viral protein to interact with the host cell receptor.[39,40,72]

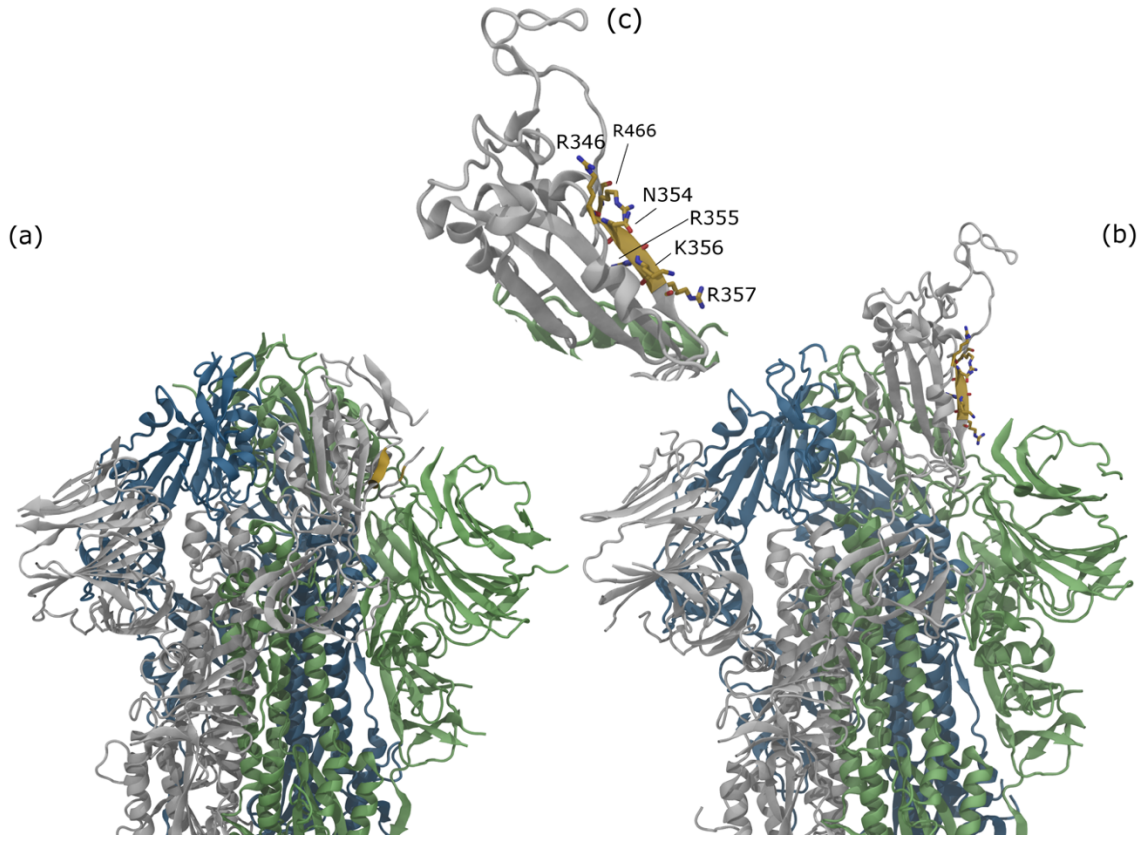

Figure 3.1.2. Panel (a): S protein in the closed state (PDB ID: 6VXX). Panel (b): S protein in the open state (PDB ID: VYB). Panel (c): Zoomed-in view of S1-RBD. The three monomers of the S protein are shown as grey, green and blue ribbons. The key residues of S1-RBD (R346, N354, R355, K356, R357, R466) are displayed as orange ribbon in the closed state and as orange ribbon and tubes in the open state.

3.1.1.3. Involvement of heparan sulfate in SARS-CoV-2 infection

The role of HS as co-receptor mirrors several coronaviruses, such as murine coronavirus and human coronavirus CoV-NL63; moreover, it has parallels with many HS-dependent growth factors, which have an analogous dual receptor system.[73-76] HS serves as a co-receptor in SARS-CoV-2 infection, facilitating the initial attachment of the virus to the host cell (the interaction of the S protein with the ACE2 receptor may not be sufficient for an efficient entry of the virus).[77] Moreover, it binds the S protein and triggers a conformational change in S1-RBD, promoting the exposure of the ACE2 recognition interface that enables the virus to specifically interact with the ACE2 receptor on the surface of the host cell. Since the first observations supporting that the conformational change in S1-RBD is induced by exogenous heparin, considerable efforts have been done to determine the molecular keys underpinning this interaction.[40,78,79] Several progresses have been made in identifying putative binding sites on S1-RBD and potential HS-based structures that engage the protein.[80,81] However, a

considerable gap remains in understanding the atomic basis of the interaction between S1-RBD and HS.

3.1.2. Aim of the work

The purpose of this research was to build and validate a model of the binding between S1-RBD and HS by applying ligand-based NMR techniques and computational methods. Three synthetic HS-like oligosaccharides characterized by different length and composition - a hexasaccharide [L-IdoA2S α (1-4) D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4) OMe], a pentasaccharide [D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS3S6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4) OMe] and a nonasaccharide [D-GlcA β (1-4) D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS3S6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4) D-GlcA β-$pPhNO_2$] - were used as molecular probes to study how HS interacts with S1-RBD (Figure 3.1.3). The hexasaccharide (1), representing a short fragment of HS, was selected for exploring the glycosaminoglycan (GAG) binding site on the S1-RBD surface. The pentasaccharide (2), known to specifically bind and activate the antithrombin protein (AT), was chosen as an alternative molecular probe because of the presence of the rare trisulfated glucosamine moiety, a key residue in driving the GAG-AT interaction. The longest oligosaccharide, namely nonasaccharide (3), was included for identifying the minimum binding epitope.

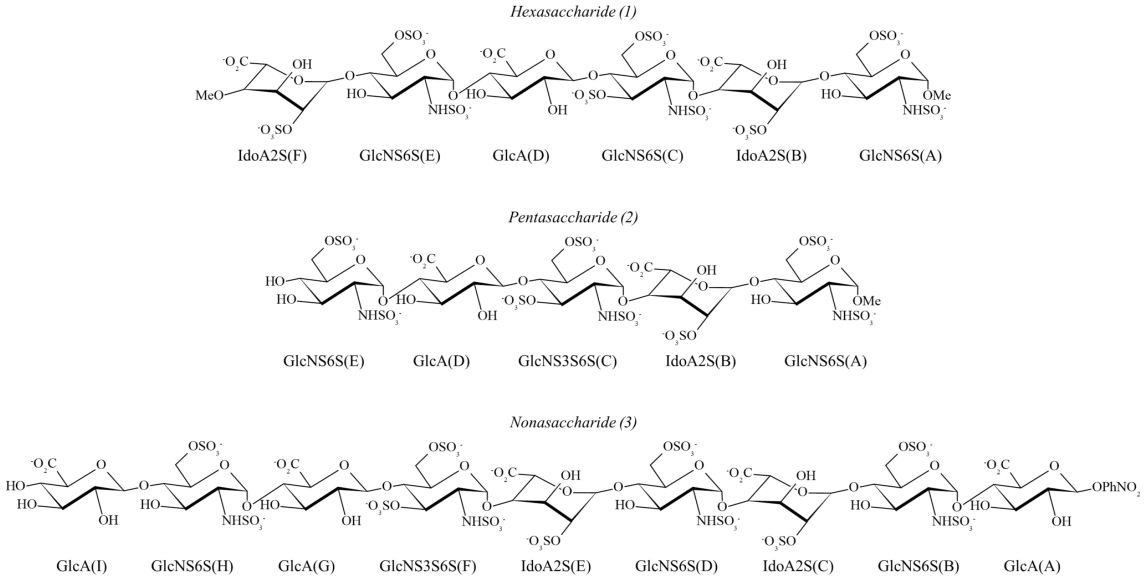

Figure 3.1.3. Chemical structure of the synthetic HS oligosaccharides used as molecular probes. The sugar residues are depicted in their prevalent conformation in the unbound state (L-IdoA2S residues can be either in $^{1}C_4$ chair or $^{2}S_0$ skew-boat conformation, depending on the sulfation degree of the neighbouring glucosamines).

The interaction between these oligosaccharides and S1-RBD was analysed by saturation transfer difference (STD) NMR and transferred nuclear Overhauser effect spectroscopy (tr-NOESY) experiments. The STD NMR spectra were recorded to determine the binding epitope of the glycans (1), (2) and (3), while NOE and tr-NOE data were compared to define the unbound and bound conformations of the oligosaccharide (1). Docking and molecular dynamics (MD) simulations were employed to delineate the GAG binding site on the S1-RBD surface and to build 3D models of the (1)-S1-RBD and (2)-S1-RBD complexes depicting the contacts and intermolecular forces that drive the interaction. Selected (1)-S1-RBD geometries were validated by comparing simulated and experimental intra-residue and inter-glycosidic tr-NOE values, while the (2)-S1-RBD poses were supported by STD NMR data.

3.1.3. Results

3.1.3.1. Mapping the binding epitope of three glycan probes on S1-RBD by STD NMR

Hexasaccharide (1) was characterized by 1D and 2D NMR experiments. All proton (^{1}H) and carbon (^{13}C) chemical shifts were assigned unambiguously (Table A3.1.1). The ^{1}H-^{13}C HSQC spectrum is reported in Figure A3.1.1. The NMR characterization of pentasaccharide (2) is published.[82,83] The ^{1}H-^{13}C HSQC spectrum of nonasaccharide (3) with the assignment of the anomeric region is shown in Figure A3.1.2.

STD NMR experiments were performed to test if the interaction of the oligosaccharides (1), (2) and (3) with S1-RBD occurs. Additionally, they were carried out to determine the glycan residues in contact with S1-RBD and their proximity to the receptor surface. For the non-overlapping ligand protons (*i.e.*, anomeric protons), the STD percentages were calculated as reported in the experimental section and mapped onto the ligand structure to give the interacting epitope. Since these values reflect the relative amount of saturation transferred from the protein to the ligand, the higher the STD percentage, the closer the ligand proton is to the binding site.

The STD NMR analysis conducted on (1) in the presence of S1-RBD allowed to delineate the epitope map of (1) for its interaction with S1-RBD (Table 3.1.1 and Figure 3.1.4). The comparison of the STD NMR spectrum with its reference showed that all the anomeric protons of (1) are affected by a magnetization transfer from S1-RBD, indicating that all the sugar units are involved in the molecular recognition process. Specifically, the iduronate moiety at the non-reducing end [IdoA2S(F)] displays the lowest STD NMR signals, suggesting a longer distance from the protein surface than the other residues, whose STD values are in the range of 76-100%.

Table 3.1.1. STD percentages (STD %), calculated as described in the experimental section, of the anomeric ligand protons.

Hexasaccharide (1)	
Proton	**STD %**
H1F	65
H1E	76
H1D	100
H1C	85
H1B	87
H1A	100

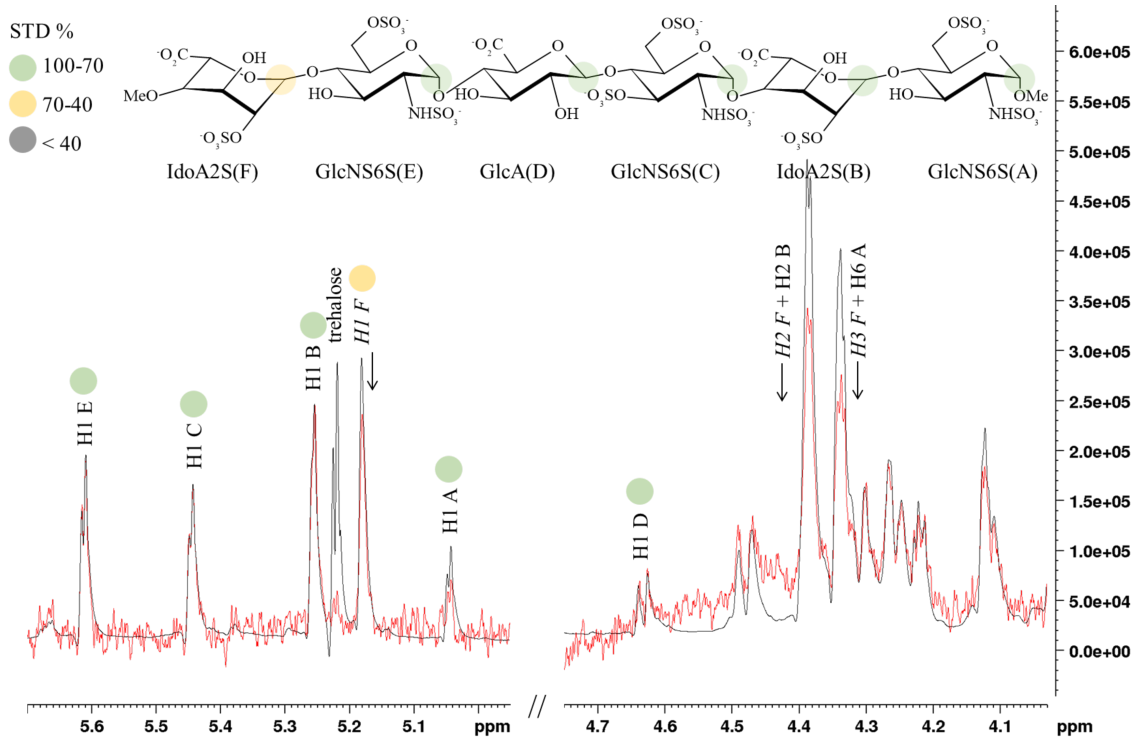

Figure 3.1.4. STD NMR analysis of hexasaccharide (1) in interaction with S1-RBD. Binding epitope map of hexasaccharide (1) as derived by STD NMR data. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of hexasaccharide (1)-S1-RBD mixture with a molecular ratio of 100:1, at 293 K.

Furthermore, the binding of (1) to S1-RBD was qualitatively monitored by analysing linewidths and chemical shifts of the signals in the proton spectrum of the ligand recorded in the presence of the protein (ligand-to-protein ratio of 7:1) (Figure 3.1.5). The increase in linewidths and the absence of chemical shift perturbations are correlated with an equilibrium regulated by an intermediate exchange rate between the free and bound states. The dissociation constant (K_d), estimated by isothermal titration calorimetry (ITC) experiments using heparin hexasaccharide fractions, is 4.5×10^{-7} M (see Chapter 3.2).

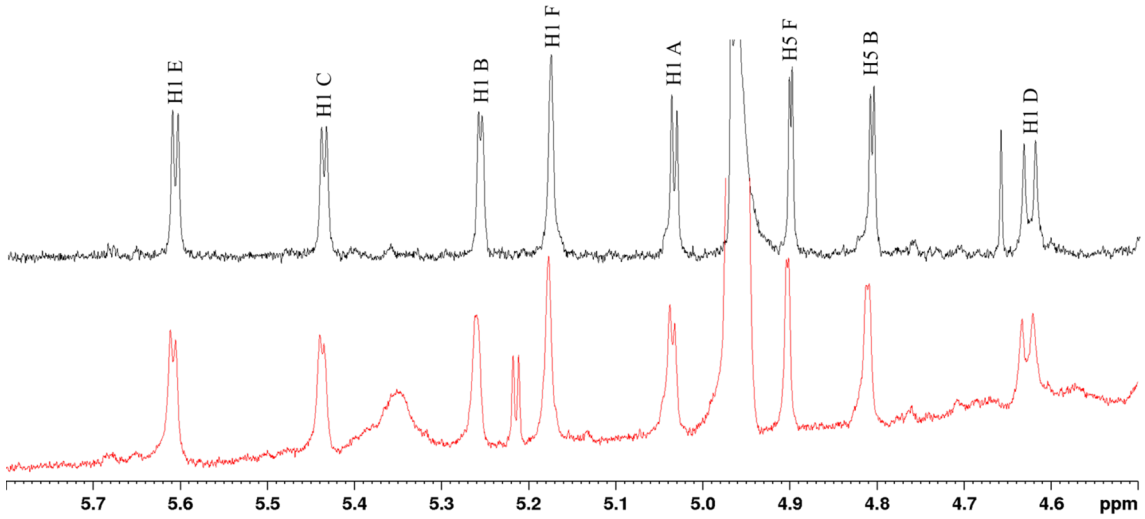

Figure 3.1.5. Anomeric region of the proton spectrum of hexasaccharide (1) in the free state (black line) and in the S1-RBD-bound state (ligand-to-protein ratio of 7:1), at 280 K.

The interaction between (2) and S1-RBD was also detected by means of the STD NMR technique. Analogously to (1), the binding epitope of (2) indicated that all the monosaccharides are recognized by S1-RBD (Table 3.1.2 and Figure 3.1.6). Indeed, all the H1 signals of (2) exhibit STD enhancements around 70-100%, except for those belonging to the GlcNS6S(E) and IdoA2S(B) residues, whose STD percentages are 57% and 63%, respectively.

Table 3.1.2. STD percentages (STD %), calculated as described in the experimental section, of the anomeric ligand protons.

Pentasaccharide (2)	
Proton	**STD %**
H1E	57
H1D	100
H1C	70
H1B	63
H1A	100

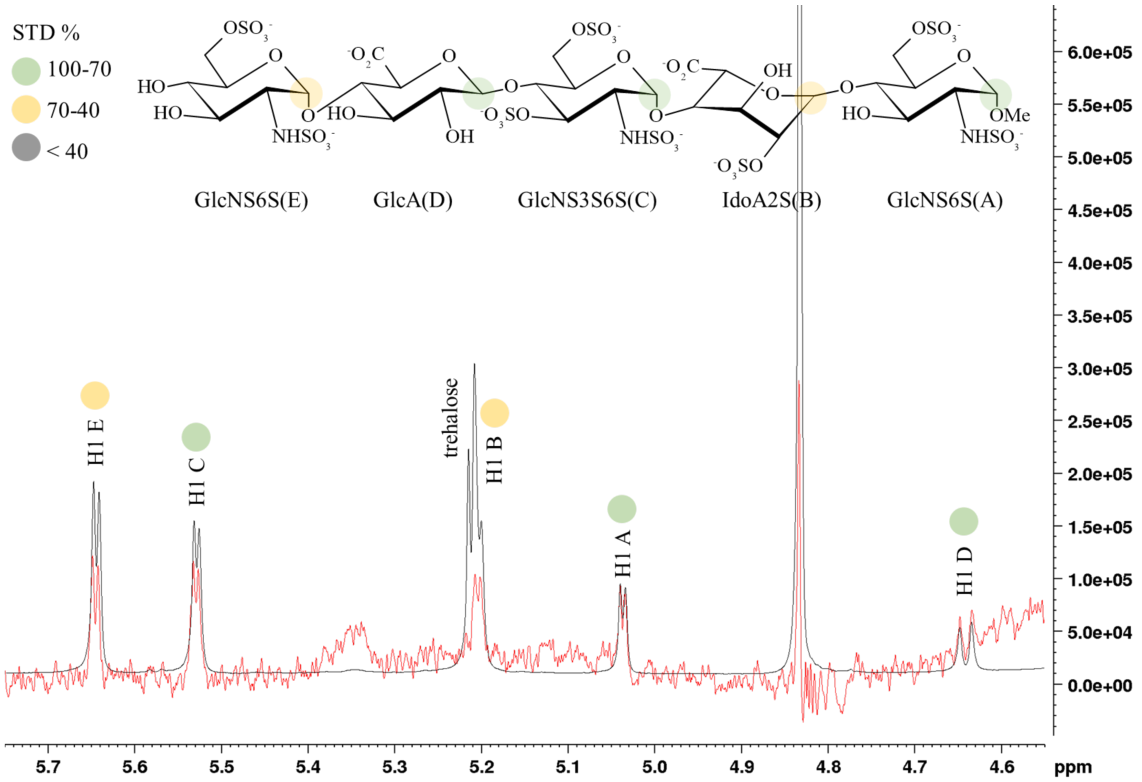

Figure 3.1.6. STD NMR analysis of pentasaccharide (2) in interaction with S1-RBD. Binding epitope map of pentasaccharide (2) as derived by STD NMR data. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of pentasaccharide (2)-S1-RBD mixture with a molecular ratio of 100:1, at 293 K.

The epitope mapping of (3) for its interaction with S1-RBD, obtained by STD NMR, revealed that only a limited portion of the glycan binds contiguously to the protein (Table 3.1.3 and Figure 3.1.7). The GlcNS6S(H) monosaccharide at the non-reducing end as well as the GlcNS6S(B)-GlcA(A) disaccharide at the reducing end show weaker STD intensities than the other residues, unveiling that they are less involved in the binding. Moreover, the aromatic ring linked to the glucuronic acid at the reducing end [GlcA(A)] displays no STD enhancement, implying that it is likely exposed to the solvent. Figure 3.1.7 does not show the signals of H1 GlcA(I) and H1 GlcA(G) because of their overlap with the H_2O signal. The STD NMR analysis performed at 280 K allowed to obtain information about both GlcA(I) and GlcA(G): the epitope mapping showed the lower involvement of GlcA(I) and the stronger contribution of GlcA(G) (data not reported). These results suggest the presence of a recognition site on the S1-RBD surface of less than six sugar residues.

Table 3.1.3. STD percentages (STD %), calculated as described in the experimental section, of selected ligand protons.

Nonasaccharide (3)	
Proton	**STD %**
H1H	68
H1F	98
H1E	100
H1D	92
H1C	73
H1B	68
H1A	50
$PhNO_2$	43

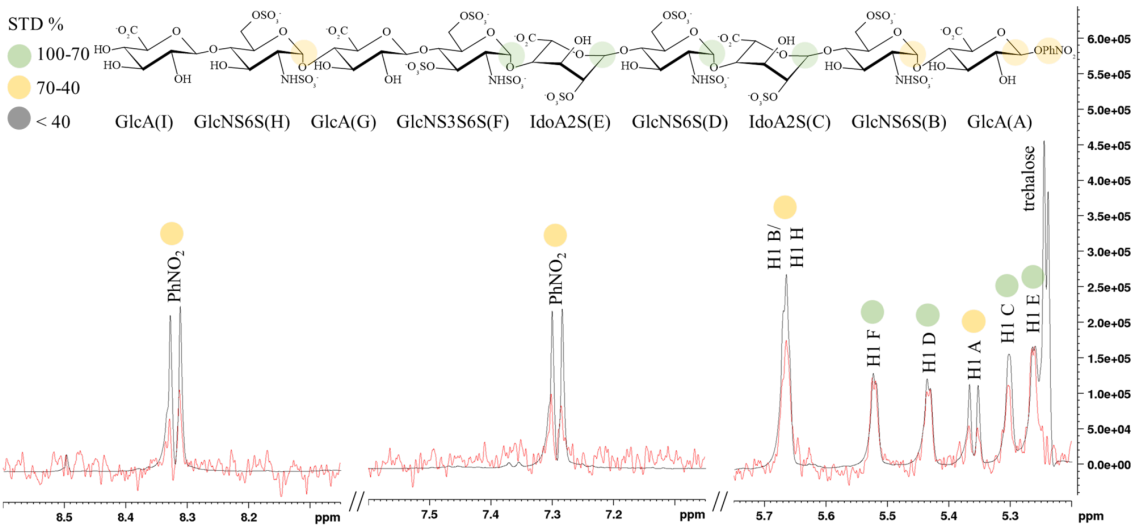

Figure 3.1.7. STD NMR analysis of nonasaccharide (3) in interaction with S1-RBD. Binding epitope map of nonasaccharide (3) as derived by STD NMR data. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of nonasaccharide (3)-S1-RBD mixture with a molecular ratio of 100:1, at 293 K.

A signal belonging to trehalose, a preservative found in the protein solution, is present in all the reference spectra and absent in the corresponding difference spectra (Figures 3.1.4, 3.1.6 and 3.1.7). This was a demonstration of the effectiveness of the conducted STD NMR experiments in discriminating between interacting and non-interacting molecules.

3.1.3.2. Determining the conformation of hexasaccharide (1) in the free and bound states by $^3J_{H-H}$ coupling constants and NOESY/tr-NOESY experiments

Three-bond proton-proton coupling constants ($^3J_{H-H}$) of (1) in the free state were measured to define the conformation of the sugar rings (Table 3.1.4). The analysis of the $^3J_{H-H}$ values showed that the glucosamines [GlcNS6S(E, C, A)] ($^3J_{H1-H2}$ ~3.5 Hz and $^3J_{H2-H3}$ ~10 Hz) and

the glucuronic acid [GlcA(D)] ($^3J_{H1-H2}$ and $^3J_{H2-H3} \geq 8$ Hz) adopt a $^4C_1$ chair conformation.[84] Additionally, the evaluation of the $^3J_{H-H}$ magnitudes allowed to establish the conformer populations of the iduronic acid units [IdoA2S(F, B)], which depend on the identity and sulfation pattern of the neighbouring glucosamines. The iduronate moiety at the non-reducing end [IdoA2S(F)] exists in the $^4C_1$ chair, $^1C_4$ chair and $^2S_0$ skew-boat conformations: the splitting of both H1 ($^3J_{H1-H2} < 1$ Hz) and H5 ($^3J_{H4-H5} = 1.9$ Hz) signals indicated the prevailing abundance of the $^1C_4$ form. In contrast, the complete set of the $^3J_{H-H}$ values pointed out that IdoA2S(B) is present in a conformational equilibrium between the $^1C_4$ (63%) and $^2S_0$ (37%) forms. Unfortunately, the line-broadening observed in the proton spectrum of the (1)-S1-RBD complex precluded the $^3J_{H-H}$ measurement.

Table 3.1.4. $^3J_{H-H}$ coupling constants (Hz) of the sugar units of hexasaccharide (1) in the unbound state.

$^3J_{H-H}$	IdoA2S(F)	GlcNS6S(E)	GlcA(D)	GlcNS6S(C)	IdoA2S(B)	GlcNS6S(A)
$^3J_{H1-H2}$	<1.0	3.8	7.9	3.5	2.4	3.6
$^3J_{H2-H3}$	-	10.3	9.3	10.6	6.0	10.8
$^3J_{H3-H4}$	-	-	-	-	3.5	-
$^3J_{H4-H5}$	1.9	-	-	-	2.5	-

NOESY and tr-NOESY experiments were carried out to prove the interaction and describe the intra-residue and inter-glycosidic conformational changes upon binding (Figure A3.1.3). The cross-relaxation rate of (1) in the free state differs from that of (1) bound to S1-RBD. This difference led to an increase in the intensity of the cross-peaks from the NOESY spectra acquired for the free ligand to the tr-NOESY spectra recorded in the presence of the protein, providing evidence of the interaction between (1) and S1-RBD (Tables 3.1.5, 3.1.6 and 3.1.7). There is no change in the sign of NOE because oligosaccharides characterized by more than four units as well as large molecules exhibit a negative NOE in the set experimental condition.

Table 3.1.5. H1-H2 NOEs and tr-NOEs of GlcNS6S(E, C, A) of hexasaccharide (1) in the unbound and S1-RBD-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage.

	GlcNS6S(E)		*GlcNS6S(C)*		*GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2
0.15	7.7	13.1	6.7	10.7	5.0	8.8
0.3	16.4	25.8	13.2	21.1	11.1	19.8
0.5	27.5	45.4	21.7	35.7	19.7	36.8

Table 3.1.6. H5-H2 and H5-H4 NOEs and tr-NOEs of IdoA2S(F, B) of hexasaccharide (1) in the unbound and S1-RBD-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage. The ratio H5-H2/H5-H4 is shown in brackets.

	IdoA2S(F)		*IdoA2S(B)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H5-H2/ H5-H4	H5-H2/ H5-H4	H5-H2/ H5-H4	H5-H2/ H5-H4
0.15	0.5/4.6 (0.1)	1.0/10.0 (0.1)	1.9/5.2 (0.4)	3.8/11.2 (0.3)
0.3	1.0/10.7 (0.1)	2.0/19.5 (0.1)	4.3/11.1 (0.4)	7.7/21.1 (0.4)
0.5	1.5/16.3 (0.1)	4.6/32.0 (0.1)	7.4/17.3 (0.4)	12.8/29.2 (0.4)

Table 3.1.7. H1-H6 (or H1-H3) and H1-H4 inter-glycosidic NOEs and tr-NOEs characterizing the backbone conformation of hexasaccharide (1) in the unbound and S1-RBD-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage. The ratios H1-H6/H1-H4 and H1-H3/H1-H4 are shown in brackets.

	IdoA2S(F)-GlcNS6S(E)		*GlcNS6S(E)-GlcA(D)*		*GlcA(D)-GlcNS6S(C)*		*GlcNS6S(C)-IdoA2S(B)*		*IdoA2S(B)-GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H6/ H1-H4	H1-H6/ H1-H4	H1-H4	H1-H4	H1-H6/ H1-H4	H1-H6/ H1-H4	H1-H3/ H1-H4	H1-H3/ H1-H4	H1-H6/ H1-H4	H1-H6/ H1-H4
0.15	1.6/4.9 (0.3)	3.1/7.9 (0.4)	5.1	9.5	4.4/7.1 (0.6)	4.9/14.8 (0.3)	4.7/5.0 (0.9)	7.7/9.8 (0.8)	5.1/5.4 (0.9)	8.1/9.7 (0.8)
0.3	2.9/9.8 (0.3)	6.9/14.9 (0.5)	10.4	18.3	5.8/10.4 (0.6)	8.8/28.9 (0.3)	9.4/9.6 (1.0)	13.8/ 17.3 (0.8)	10.4/ 10.3 (1.0)	15.8/ 16.9 (0.9)
0.5	6.5/16.5 (0.4)	15.0/ 26.2 (0.6)	17.5	32.8	9.2/19.6 (0.5)	14.1/ 41.7 (0.3)	15.7/ 16.8 (0.9)	23.7/ 29.2 (0.8)	16.4/ 17.7 (0.9)	26.9/ 28.6 (0.9)

The strong H3-H5 NOEs and tr-NOEs measured for GlcNS6S(E, C, A) and GlcA(D) confirmed that these units adopt a $^4C_1$ chair conformation in both free and bound states (data not reported). Since IdoA2S in $^1C_4$ chair and $^2S_0$ skew-boat conformations exhibits distinct H5-H2 distances (4.0 Å and 2.4 Å, respectively), the ratio between the magnitudes of the H5-H2 and H5-H4 NOEs qualitatively provides the proportions of the two conformers. The weak H5-H2 NOE of IdoA2S(F) is compatible with an almost pure $^1C_4$ chair conformation (H5-H2/H5-H4 ~0.1), while the strong H5-H2 NOE of IdoA2S(B) complies with an equilibrium between the $^1C_4$ chair and $^2S_0$ skew-boat conformations (H5-H2/H5-H4 ~0.4). This agreed with the $^3J_{H-H}$ measurements. The ratios between the H5-H2 and H5-H4 tr-NOEs for both IdoA2S(F) and IdoA2S(B) remain unaltered upon binding, implying that neither residue undergoes significant changes in its average conformation. The conformation of the oligosaccharide backbone in the free and bound states was described by a set of inter-glycosidic NOE and tr-NOE values. The top rotor dynamic of heparin/HS oligosaccharides longer than four residues introduces a dependency of the NOE cross-relaxation rate on the orientation of H-H vectors with the molecular axis. Nevertheless, the aforementioned effect

can be considered negligible in this case. Indeed, the ratios between the linewidth values of all the H1 signals in the free and bound states are approximately uniform (0.75, 0.80, 0.79, 0.78, 0.74, 0.65 from F to A residue), indicating only a weak anisotropic effect. Although the comparison between the NOE and tr-NOE values cannot be precisely correlated with a change in terms of distance upon binding, it highlighted significant conformational rearrangement. Docking and MD simulations showed minor modifications of the inter-glycosidic geometries, suggesting a low adaptation of the backbone conformation on the S1-RBD surface to optimize the contacts. The most remarkable adjustment concerns the GlcA(D)-GlcNS6S(C) glycosidic linkage, also observed experimentally by the significant variation of the ratio between H1-H6 and H1-H4 NOEs/tr-NOEs, that changes from 0.6 in the free state to 0.3 in the bound state.

3.1.3.3. Exploring the glycan binding sites on S1-RBD by molecular modelling

The analysis of the S1-RBD surface allowed three putative GAG binding sites to be detected: site I (R346, N354, K356, R357, R355, K444, R466), site II (K424, R454, R457, K458, K462, R466) and site III (R403, R408, K417, K444). These sites are not shielded by glycosylation and are characterized by arginine (R) and lysine (K) residues that, even if non-consecutive, are solvent exposed and form positively charged patches. Unlike sites I and II, site III partially overlaps with the ACE2 binding site. Therefore, even though site III may initially engage HS, it can be excluded since S1-RBD binds ACE2 in the presence of HS or heparin.[39] To define the favoured GAG binding site on the S1-RBD surface, these sites were ranked by a computational approach that combines docking and MD simulations, employing (1) as a molecular probe. Firstly, docking was performed to generate three clusters of geometries of the (1)-S1-RBD complex. Each cluster was obtained independently, targeting site I, site II or site III. Then, MD simulations were carried out to refine the obtained geometries and to estimate their free energy of binding. Moreover, the per-residue energy decomposition analysis allowed to quantify the contribution of each residue to the free energy of binding. The docking solutions were ranked, according to the AutoDock scoring function (ADT), the cluster population and the free energy of binding (ΔG_{PB}^{bind}) (Table 3.1.8). The comparison of the ADT and ΔG_{PB}^{bind} indicators pointed out that (1) preferentially binds site I of S1-RBD over sites II and III. Moreover, even if sites II and III are characterized by comparable binding energies, the interaction with site II is preferred than site III in terms of cluster population.

Table 3.1.8. Summary of the docking and MD simulation data. In the 6th column, the estimated error on the average value of the free energy of binding (ΔG_{PB}^{bind}) is reported in brackets on the last decimal digit.

Cluster Rank	ADT Lowest Binding Energy (Kcal mol^{-1})	Cluster Population	Equilibration Time (ns)	Average Time (ns)	ΔG_{PB}^{bind} (Kcal mol^{-1})
Site I					
1	–2.50	3	80	[80, 100]	–30.6(3)
2	–1.70	2	36	[36, 56]	–32.3(6)
3	–1.33	2	36	[36, 56]	–30.6(4)
Site II					
1	0.26	2	36	[36, 56]	–9.7(4)
2	1.12	7	58	[58, 78]	–10.4(5)
Site III					
1	0.10	4	40	[40, 60]	–12.1(3)
2	1.23	3	52	[52, 72]	–6.7(3)

Regarding the interaction between (1) and site I of S1-RBD, three clusters of docking solutions, namely clusters 1, 2 and 3, characterized by comparable free energy of binding, emerged. Clusters 1 and 2, being approximately similar in terms of contacts and binding energy, were treated together. Interestingly, the docking solutions belonging to clusters 1 and 3 show opposite orientations of (1). Cluster 1 is characterized by the non-reducing end of (1) oriented toward the ACE2 binding site (binding mode A), while the reverse orientation is found in cluster 3 (binding mode B) (Figure 3.1.8). The per-residue analysis of the free energy of binding showed that key residues - R346, K355, R356, R357, R466 - contribute to the binding in both modes A and B (Figure 3.1.8). Specifically, it displayed that the binding mode A presents a stronger involvement of R346 (–11 vs –6 Kcal mol^{-1}) and a weaker participation of R466 (–2 vs –11 Kcal mol^{-1}) than the binding mode B. This analysis also indicated that (1) interacting with S1-RBD in mode A exhibits tighter contacts with the residues located near site II (N450, Y451, R454, R457, K458 and K462). Furthermore, it revealed that the binding mode B has a stronger interaction with N354 (–3 vs 1.3 Kcal mol^{-1}) and greater repulsive contributions from GlcA(D), GlcNS6S(C) and D398 (5.8, 4.3 and 3 Kcal mol^{-1} vs 1.7, 2.9 and 1.9 Kcal mol^{-1}) than the binding mode A.

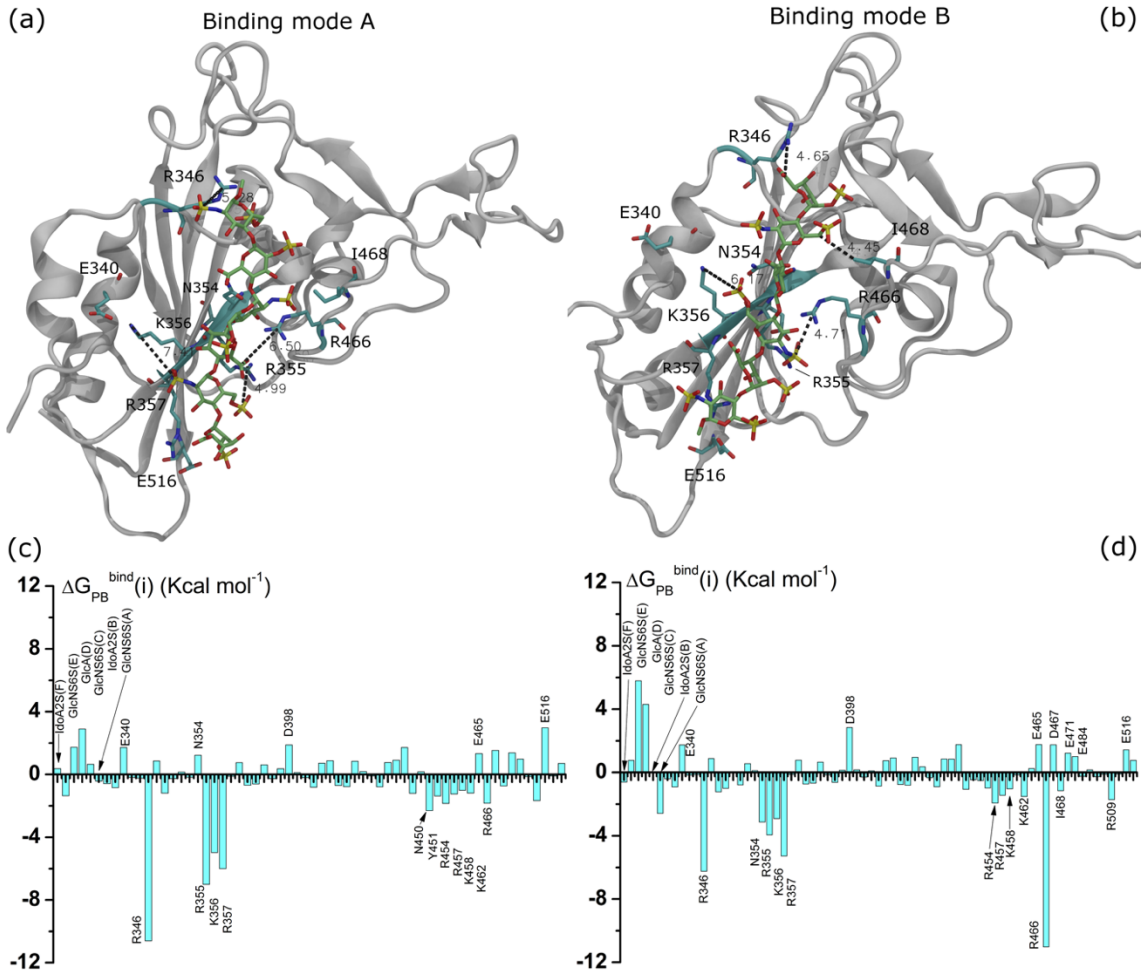

Figure 3.1.8. Analysis of the interaction between hexasaccharide (1) and site I of S1-RBD. Panel (a): 3D structure of (1) bound to S1-RBD in mode A sampled at the MD simulation time of 81.58 ns. Panel (b): 3D structure of (1) bound to S1-RBD in mode B sampled at the MD simulation time of 37.75 ns. Panels (c) and (d): histograms of the per-residue analysis of the free energy of binding [$\Delta G_{PB}^{bind}(i)$] of (1) interacting with S1-RBD in modes A and B. In panels (a) and (b), hexasaccharide (1) is represented by green tubes, S1-RBD is shown as grey ribbon and the amino acids of site I are displayed as cyan ribbon and tubes. Selected distances are underlined by dashed segments. Distances are in Angstrom (Å). In panels (c) and (d), the histograms are averaged on the production stage of the MD simulation trajectory. Energies are in Kcal mol^{-1}.

The analysis of the unfavourable contributions when (1) binds S1-RBD in mode B showed that GlcA(D) and GlcNS6S(C) are characterized by strong positive (unfavourable) de-solvation energies (ΔG_{sol}), not counterbalanced by negative (favourable) electrostatic and/or van der Waals terms (Table 3.1.9). At molecular level, the positive values of GlcA(D) and GlcNS6S(C) in the binding mode A [$\Delta G_{PB}^{bind}(i)$ = 1.7 and 2.9 Kcal mol^{-1}] correlate with a de-solvation penalty when GlcA(D) approaches the backbone of K355 and the N-sulfate group of GlcNS6S(C) becomes proximal to the hydrophobic side chain of I468. In both contacts, the solvation shell surrounding the ligand is locally lost upon binding. In the case

of the binding mode B, the carboxyl group of GlcA(D) loses its solvation shell, being oriented toward the S1-RBD surface (close to N354), without an electrostatic and/or van der Waals interaction to compensate this loss of energy. Analogously, GlcNS6S(C) presents unfavourable contacts localized on both its ring and the 6-O-sulfate group [$\Delta G_{PB}^{bind}(i)$ = 3.0 and 2.0 Kcal mol^{-1}]. Even though the N-sulfate group exhibits favourable contacts with R355 and R466 [$\Delta G_{PB}^{bind}(i)$ = –0.8 Kcal mol^{-1}], the contribution of GlcNS6S(C) to the free energy of binding is therefore unfavourable [$\Delta G_{PB}^{bind}(i)$ = 5.80 Kcal mol^{-1}].

Table 3.1.9. Van der Waals (ΔE_{vdw}), electrostatic (ΔE_{coul}) and solvation (ΔG_{sol}) energy terms for GlcA(D) and GlcNS6S(C). These contributions were calculated for the (1)-S1-RBD complex in modes A and B and averaged on the production period of the MD simulations reported in the 5th column of Table 3.1.8.

Residue	Binding Mode	ΔE_{vdw}	ΔE_{coul}	ΔG_{sol}	$\Delta E_{vdw} + \Delta E_{coul} + \Delta G_{sol}$
GlcA(D)	A	–2.56	–46.11	50.39	1.73
GlcNS6S(C)		–2.48	–58.87	64.24	2.89
GlcA(D)	B	–2.04	–49.57	57.40	5.80
GlcNS6S(C)		–5.94	–109.36	119.60	4.30

3.1.3.4. Analysis of the binding of hexasaccharide (1) on S1-RBD

Docking, MD simulations and NOEs/tr-NOEs analysis suggested that (1) binds site I of S1-RBD in two modes A and B. The glycosidic conformation of (1) interacting with site I of S1-RBD in modes A and B was described by Ramachandran plots (Figure 3.1.9). A density colour map was superposed on each Ramachandran plot. The colour gradient (from blue to red) is proportional to the density of the sampled φ_i/ψ_i states and qualitatively predicts the preferred conformation of each glycosidic linkage.

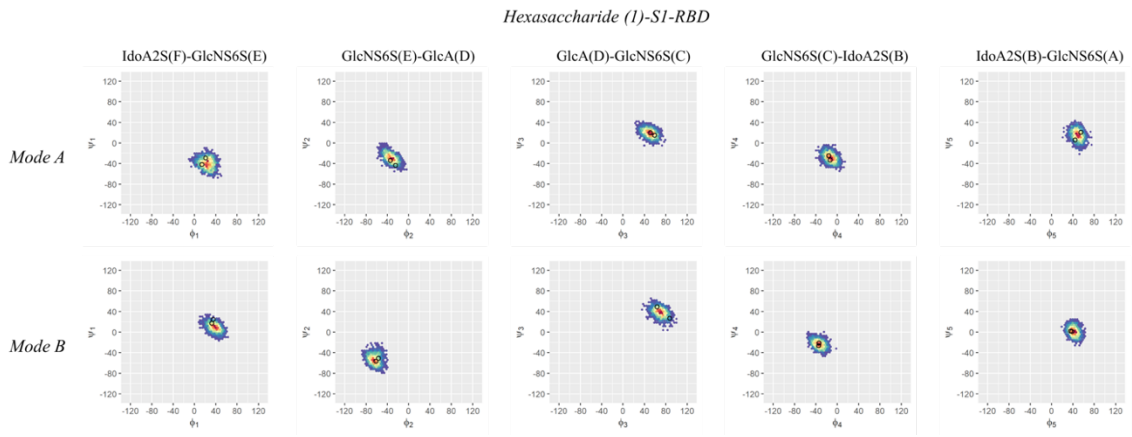

Figure 3.1.9. Ramachandran plots of the glycosidic dihedral angles φ_i/ψ_i of hexasaccharide (1) bound to site I of S1-RBD. Two possible orientations of (1), corresponding to modes A (upper panels) and B (bottom panels), are represented. The glycosidic dihedral states were sampled during the production stage of the MD simulations [80 to 100 ns (mode A) and 36 to 56 ns (mode B)]. A density colour map is superposed on each Ramachandran plot; the colour gradient (blue to red) is proportional to the density of the sampled φ_i/ψ_i states and qualitatively predicts the preferred conformation of each glycosidic linkage. The empty circles indicate the φ_i/ψ_i values of the selected structures which interpret the experimental tr-NOEs (Tables A3.1.2-A3.1.7).

Selected geometries of the (1)-S1-RBD complex were used to simulate intra-residue and inter-glycosidic tr-NOEs, which were compared with the corresponding experimental values. Their agreement is summarized in Tables A3.1.2-A3.1.7. Interestingly, it improves when both modes A and B are considered, as observed comparing the R-factor values. For IdoA2S(B), the MD simulations predicted a $^{2}S_{0}$ skew-boat form and a $^{1}C_{4}$ chair form when (1) binds site I of S1-RBD in modes A and B, respectively. This finding was further supported by the improvement of the fitting between the simulated and experimental tr-NOEs when the binding of (1) to site I of S1-RBD is described by both modes A and B (Tables A3.1.2-A3.1.7).

Compared to the interaction with site I, the binding of (1) to sites II and III turned out to be significantly weaker. Docking and MD simulations revealed that the binding of (1) to site II of S1-RBD is described by two complexes, labelled as clusters 1 and 2 and characterized by opposite orientations of (1). The per-residue analysis of the free energy of binding showed a lower efficiency in the interaction of (1) with site II in comparison to site I (Figure 3.1.10). The residues of site II involved in the recognition of (1) in mode A are R454, R457, K458 and K462 [$\Delta G_{PB}^{bind}(i)$ = –2.6, –2.6, –3.4 and –2.2 Kcal mol^{-1}]. Moreover, R355 and R466, belonging to site I, contribute to reinforce the interaction [$\Delta G_{PB}^{bind}(i)$ = 3.4 and –9.6 Kcal mol^{-1}]. Unfavourable contributions are observed for IdoA2S(B), E465, D467 and E471 [$\Delta G_{PB}^{bind}(i)$ = 10.6, 8.6, 3.1 and 3.4 Kcal mol^{-1}]. The analysis of the 3D structure of the

complex between (1) and site II of S1-RBD supported these data. The favourable contributions from R355 and R466 are correlated to the strict contact of these residues to the 6-O-sulfate group of GlcNS6S(A) (4.0 and 3.1 Å); the unfavourable contributions from IdoA2S(B), E465 and E471 are due to the proximity between the carboxyl group of IdoA2S(B) and E465 and between the 6-O-sulfate group of GlcNS6S(E) and E471. The recognition of (1) in mode B approximately involves the same set of residues and contributions. Specifically, R454, R457 and K458 stabilize the binding [$\Delta G_{PB}^{bind}(i)$ = –2.2, –4.7, –7.8 Kcal mol^{-1}]; while IdoA2S(B), GlcNS6S(A) and E465 destabilize the interaction [$\Delta G_{PB}^{bind}(i)$ = 3.1, 6.6, 6.0 Kcal mol^{-1}].

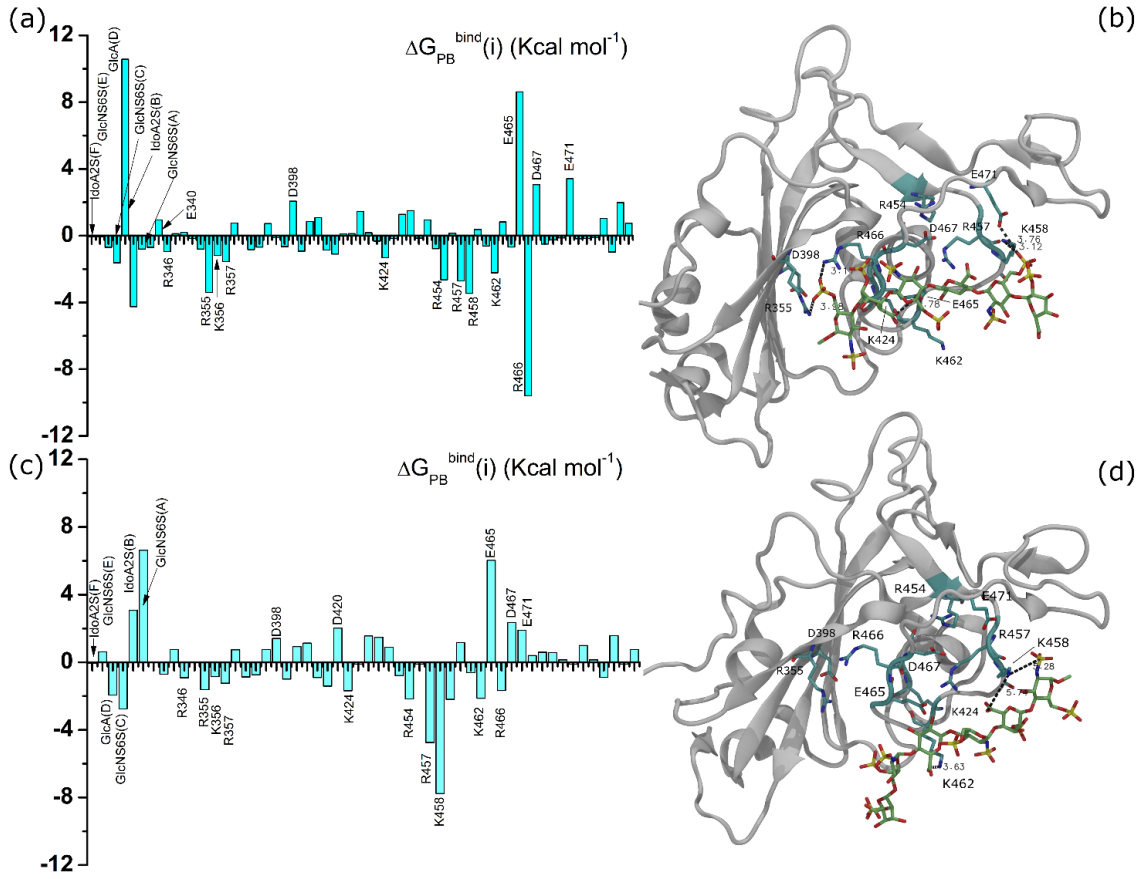

Figure 3.1.10. Analysis of the interaction between hexasaccharide (1) and site II of S1-RBD. Panel (a): histogram of the per-residue contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$] when (1) binds site II of S1-RBD in mode A. The free energy values plotted by histograms are averaged on the MD simulation trajectory upon the equilibration stage. Energies are in Kcal mol^{-1}. Panel (b): 3D structure of the (1)-S1-RBD complex generated by MD simulation and sampled in the production stage (production stage of 36 to 56 ns, snapshot sampled at simulation time of 44 ns). Panel (c): histogram of the per-residue contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$] when (1) binds site II of S1-RBD in mode B. Panel (d): 3D structure of the (1)-S1-RBD complex generated by MD simulation and sampled in the production stage (production stage of 58 to 78 ns, snapshot sampled at simulation time of 65 ns). In panels (b) and (d), hexasaccharide (1) is represented by green tubes, S1-RBD is shown as grey ribbon and the amino acids of site II are displayed as cyan ribbon and tubes. Selected distances are underlined by dashed segments and reported in Angstrom (Å).

The same procedure was employed to study the interaction of (1) with site III of S1-RBD, indicating putative binding regions that partially interfere with the ACE2 recognition site. The corresponding free energy of binding (ΔG_{PB}^{bind} = –12.1 Kcal mol^{-1}) is weaker than and comparable with the poses that involve sites I and II, respectively. The histogram of the per-residue contributions to the free energy of binding showed that the residues of site III contribute less efficiently to the interaction than those of site I. As shown in Figure 3.1.11, R403, R408 and Q414 present the strongest favourable contributions to the binding

[$\Delta G_{PB}^{bind}(i)$ = –2.1, –5.5, –3.8 Kcal mol^{-1}], while the highest unfavourable contributions are localized at IdoA2S(B), D405, E406 and D420 [$\Delta G_{PB}^{bind}(i)$ = 6.1, 2.1, 2.6, 2.7 Kcal mol^{-1}]. Additionally, Figure 3.1.11 displays the 3D structure of the complex between (1) and site III of S1-RBD sampled at MD simulation time of 44 ns, in which the most significant interactions are between IdoA2S(B), GlcNS6S(A) and R408 and between D420 and GlcNS6S(C), in agreement with the histogram of the free energy of binding.

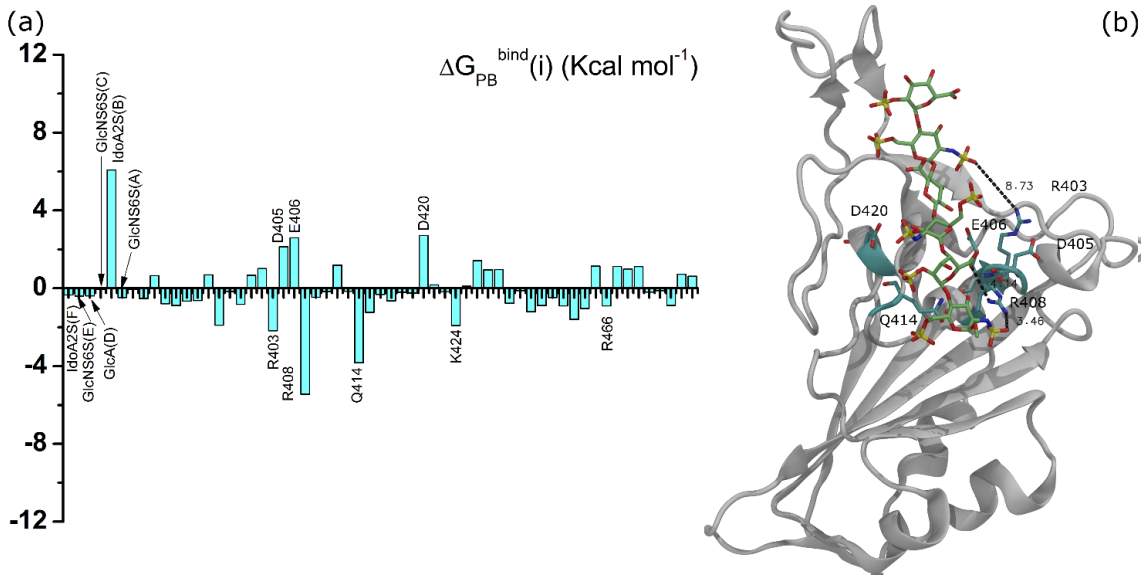

Figure 3.1.11. Analysis of the interaction between hexasaccharide (1) and site III of S1-RBD. The (1)-S1-RBD complex was built using the cluster 1 of the docking solutions (lowest ADT and ΔG_{PB}^{bind}, when site III was targeted). Panel (a): histogram of the per-residue contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$]. The free energy values plotted by histograms are averaged on the MD simulation trajectory upon the equilibration stage. Energies are in Kcal mol^{-1}. Panel (b): 3D structure of the (1)-S1-RBD complex generated by MD simulation and sampled in the production stage (production stage of 40 to 60 ns, snapshot sampled at simulation time of 44 ns). Hexasaccharide (1) is represented by green tubes, S1-RBD is shown as grey ribbon and the amino acids of site III are displayed as cyan ribbon and tubes. Selected distances are underlined by dashed segments and reported in Angstrom (Å).

Collectively, these descriptions suggested that sites II and III of S1-RBD recognize and bind (1) less efficiently than site I, despite they possess a comparable number of arginine (R) and lysine (K) residues. The presence of less homogeneously distributed positive patches on the surface of sites II and III, frequently interrupted by negatively charged residues, may provide an explanation for this observation. Positively charged side chains could be partially neutralized if surrounded by the carboxyl groups of aspartate (D) and/or glutamate (E) residues. In site I, the domain of positively charged patches, consisting of R346, R355, K356, R357 and R466, is not interrupted by aspartate and/or glutamate residues and the nearest

carboxyl groups belong to E340 and E516, outside of the site I area. Instead, in site II, E465, D467 and E471 are distributed between the positive patches of R466, K462, R454, R457 and K458, weakening their binding to (1). Analogously, in site III, D420 and D405 are in proximity of R408 and R403, disturbing the ability of site III to efficiently bind (1). Hypothetically replacing (1) with a longer HS oligosaccharide, positively charged areas are encountered in site I, which are distinct from those containing negatively charged residues in sites II and III (Figure 3.1.12). This reinforces our hypothesis that an hypothetical HS polysaccharide preferentially binds site I in comparison to site II or III.

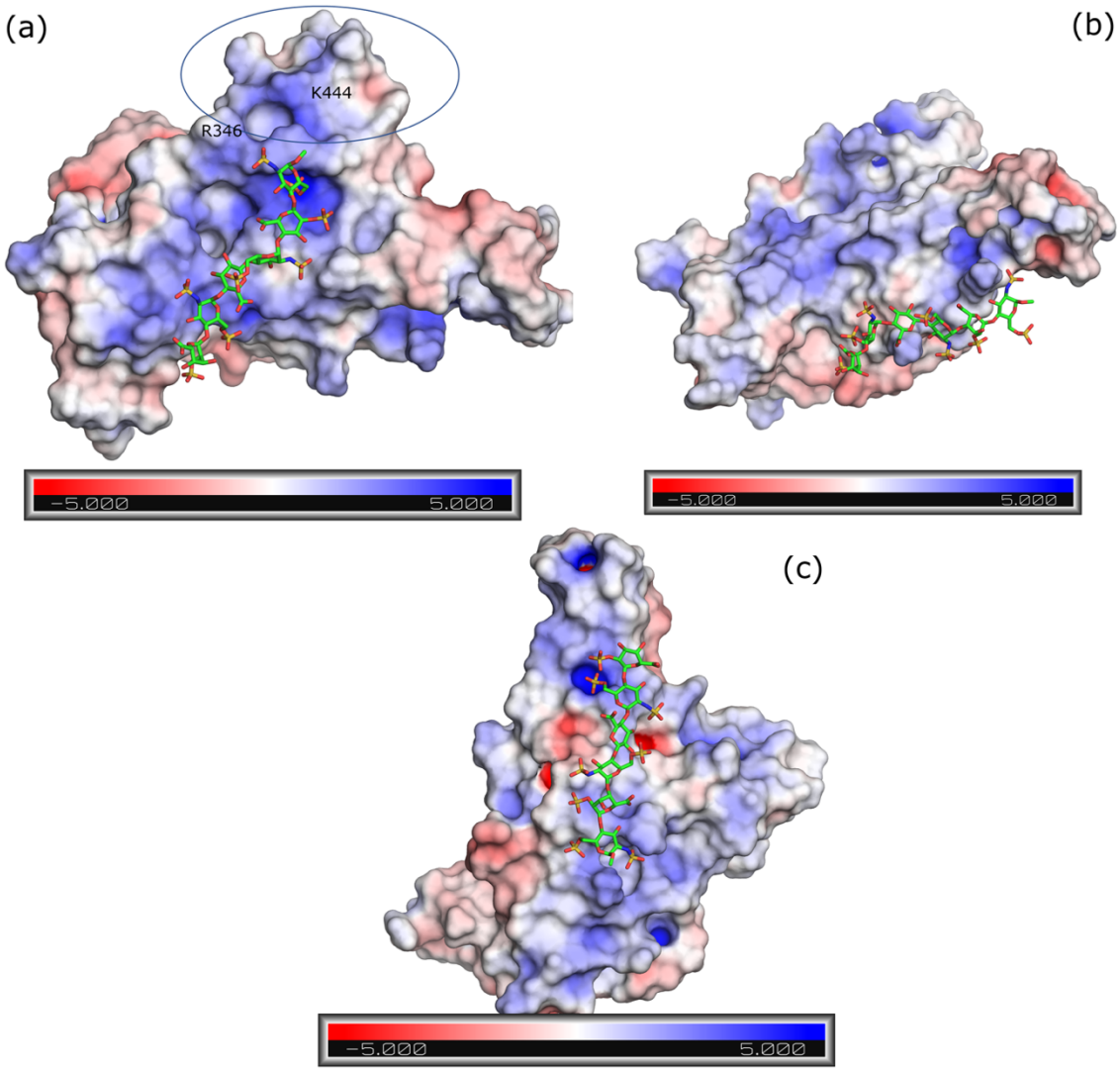

Figure 3.1.12. Electrostatic potential maps of the S1-RBD surface with hexasaccharide (1) in the bound state at site I [panel (a)], site II [panel (b)] and site III [panel (c)]. The electrostatic potential gradient (KT e^{-1}) increases from red (negative) to blue (positive). The complex of (1) with S1-RBD was sampled by MD simulation at 81.58 ns [panel (a)], 44 ns [panel (b)] and 44 ns [panel (c)].

3.1.3.5. Analysis of the binding of pentasaccharide (2) on S1-RBD

As previously described for (1), docking, MD simulations and STD NMR analysis showed that (2) binds site I of S1-RBD in two modes (Table 3.1.10). Interestingly, unlike (1), the free energy of binding and the unfavourable contacts localized on the glucosamine residues when (2) binds S1-RBD in mode A suggested that (2) preferentially interacts with S1-RBD in mode B (Figure 3.1.13). The de-stabilizing effect of the glucosamine units is correlated to significant de-solvation penalties not counterbalanced upon binding. When (2) binds S1-RBD in mode A, all the glucosamine residues (E, C, A) show strong unfavourable contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$ = 11.9, 6.1 and 5.5 Kcal mol^{-1}] since GlcNS6S(E) points its 6-O-sulfate group toward the hydrophobic cavity formed by P337 and I358, GlcNS3S6S(C) presents its 3-O-sulfate group close to the hydrophobic side chain of K356 and GlcNS6S(A) comes in contact with Y351, L452 and A348 through both N-sulfated and 6-O-sulfate groups. When (2) interacts with S1-RBD in mode B, the glucosamines E and C display less unfavourable contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$ = 5.5 and 2.9 Kcal mol^{-1}] because the N-sulfate group of GlcNS6S(E) is located in a hydrophobic area comprising Y351, L452, and A348, while the 6-O-sulfate group of GlcNS3S6S(C) makes contact with the backbone of N354. The site I residues - R346, N354, R355, K356, R357, R466 - contribute favourably to the molecular recognition of (2) by S1-RBD in both modes A [$\Delta G_{PB}^{bind}(i)$ = –1.1, –1.4, –3.2, –13.4, –2.1 and –3.1 Kcal mol^{-1}] and B [$\Delta G_{PB}^{bind}(i)$ = –2.9, –3.2, –3.5, –2.3, –7.5 and –4.3 Kcal mol^{-1}]. Few residues, including D398 and E516, destabilize the interaction between (2) and S1-RBD in both modes A [$\Delta G_{PB}^{bind}(i)$ = 3.1 and 1.3 Kcal mol^{-1}] and B [$\Delta G_{PB}^{bind}(i)$ = 2.4 and 1.5 Kcal mol^{-1}].

Table 3.1.10. Summary of the docking and MD simulation results for the (2)-S1-RBD complex. Each cluster corresponds to a docking solution and is represented by the geometry with the lowest binding energy (ADT: AutoDock scoring function). Selected docking solutions were submitted to MD simulations to calculate the Poisson Boltzmann free energy of binding (ΔG_{PB}^{bind}). In the 6th column, the estimated error on the average value of ΔG_{PB}^{bind} is reported in brackets on the last decimal digit.

Cluster Rank	ADT Lowest Binding Energy (Kcal mol^{-1})	Cluster Population	Equilibration Time (ns)	Average Time (ns)	ΔG_{PB}^{bind} (Kcal mol^{-1})
1	–4.88	3	42	[42, 56]	–11.6(4)
2	–2.90	7	40	[42, 56]	–22.4(4)

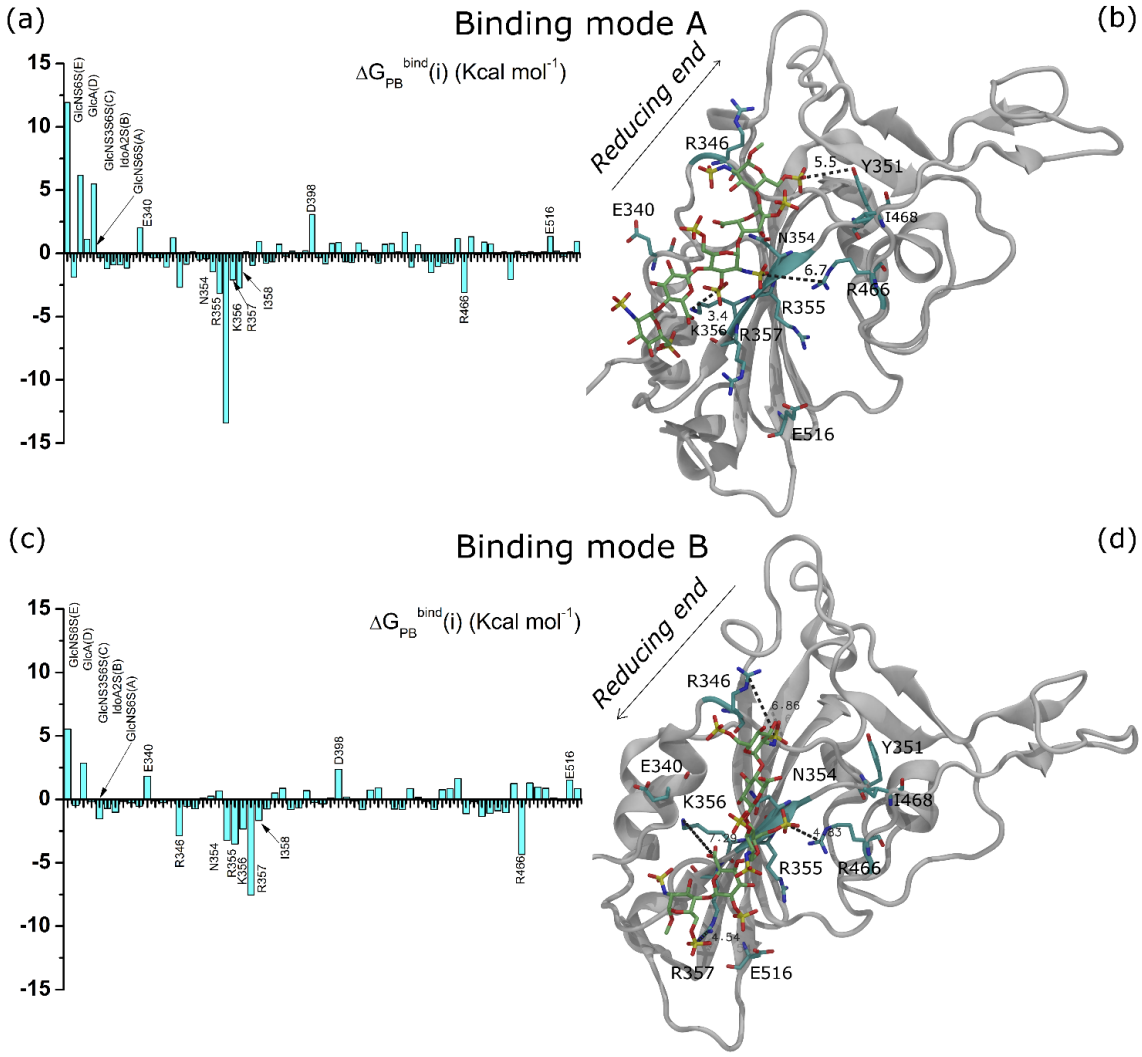

Figure 3.1.13. Analysis of the interaction between pentasaccharide (2) and site I of S1-RBD. Panels (a) and (c): histograms of the per-residue contributions to the free energy of binding [$\Delta G_{PB}^{bind}(i)$]. The free energy values plotted by histograms are averaged on the MD simulation trajectory upon the equilibration stage. Energies are in Kcal mol^{-1}. Panels (b) and (d): 3D structures of the (2)-S1-RBD complex generated by MD simulation and sampled in the production stage (42 to 56 ns). The reported snapshots were sampled at MD simulation time of 56 ns in both panels (b) and (d). Pentasaccharide (2) is represented as green tubes, S1-RBD is shown as grey ribbon and the amino acids of site I are displayed as cyan ribbon and tubes. Selected distances are underlined by dashed segments and reported in Angstrom (Å).

The presence of the STD NMR signal of H6 belonging to GlcNS3S6S(C) supports the mode B, where this proton is oriented to the S1-RBD surface within 4 Å of the carbonyl group of R355 (Figures 3.1.14 and 3.1.15). Despite that, the coexistence of the STD NMR signals of both HN and H6 of GlcNS3S6S(C), characterized by comparable intensity, suggests the possibility that (2) binds S1-RBD in both modes A and B (Figure 3.1.15).

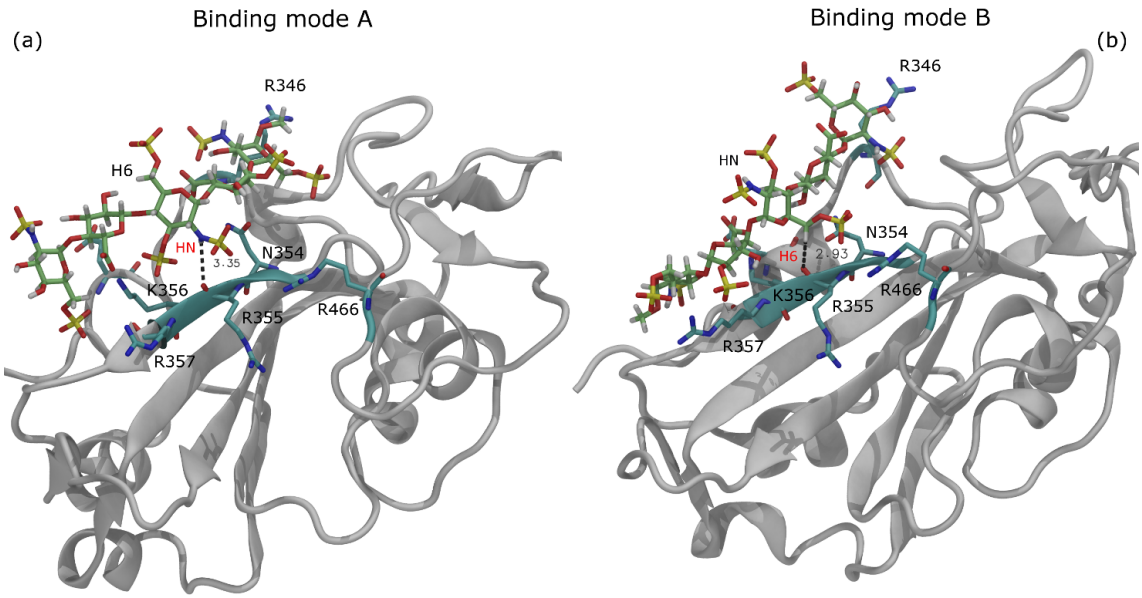

Figure 3.1.14. Insights into the contacts of GlcNS3S6S(C) when pentasaccharide (2) binds site I of S1-RBD in mode A [panel (a)] or B [panel (b)]. Pentasaccharide (2) is represented as green tubes, S1-RBD is shown as grey ribbon and the amino acids of site I are displayed as cyan ribbon and tubes. Selected distances are underlined by dashed segments and reported in Angstrom (Å).

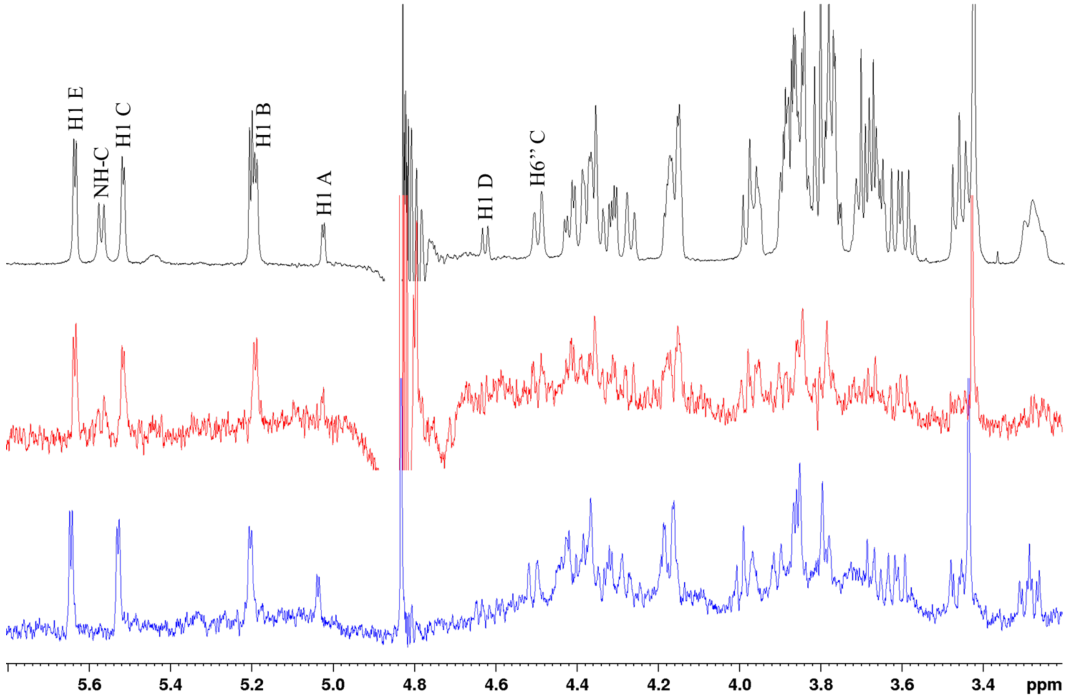

Figure 3.1.15. Reference spectrum of pentasaccharide (2) in H_2O buffer (black line). STD NMR spectra of pentasaccharide (2) interacting with S1-RBD in H_2O buffer (red line) and D_2O buffer (blue line). The H6 resonance of the trisulfated glucosamine GlcNS3S6S(C) is visible in both STD NMR spectra.

3.1.4. Conclusions

Several viruses employ HS as an initial means to attach to host cells and initiate the infectious process.[85] The extent of their specificity and affinity for certain structural features of HS is tuned to allow viral proteins to access receptors near the cell surface. SARS-CoV-2 exploits a low affinity contact between its S protein and HS to diffuse from the external to the internal layers of the glycocalyx and bind the cell surface ACE2 receptor. Our work aimed to build and validate in silico models of the interaction mechanisms between HS and S1-RBD, corresponding to the distal part of the S protein. Combining ligand-based NMR techniques and computational methods, synthetic HS oligosaccharides with different length and composition - a hexasaccharide (1), a pentasaccharide (2) and a nonasaccharide (3) - were used as molecular probes to explore the binding of HS to S1-RBD. The interaction between these oligosaccharides and S1-RBD was characterized by STD NMR spectra using (1), (2) and (3) and by NOE and tr-NOE data using (1). Three-dimensional models of the (1)-S1-RBD and (2)-S1-RBD complexes were built through docking and MD simulations and validated by comparing simulated and experimental intra-residue and inter-glycosidic tr-NOEs or by STD NMR data. The longest oligosaccharide (3) permitted to identify the minimum binding epitope (a recognition site on S1-RBD surface of no more than five or six residues). Our results support the involvement of the previously identified amino acids - R346, N354, R355, K356, R357, R466 and K444 (here labelled as site I) - as the principal site for the interaction with HS, leaving the specific site of S1-RBD characterized by flexible loops free to engage ACE2. The hypothetical extension of (1) in the complex between (1) and site I of S1-RBD to construct a longer HS oligosaccharide brings it into contact with positively charged regions on the S1-RBD surface. The theoretical macromolecular complex, obtained by superimposing the (1)-S1-RBD complex, the S1 subunit in its trimeric form and ACE2, allowed to understand that HS chains longer than six residues encounter contiguous areas of positive charge on the surface of the S1 trimer and do not interfere with the molecular recognition of ACE2 by S1-RBD (Figure 3.1.16). Our data indicate that the interaction between HS and S1-RBD lacks high specificity. Both oligosaccharides (1) and (2) possess several binding modes characterized by comparable affinities. Interestingly, these poses involve an opposite sugar orientation and reveal no detectable changes in the conformation of the iduronate residues or in the glycosidic linkage geometries that might enable a tighter binding. Evidence of multiple binding modes, while enhancing initial viral attachment, suggests a highly dynamic interaction of cellular and extracellular HS with S1-

RBD, which allows the virus to diffuse to the cell surface ACE2 receptor. Moreover, our hypothesis of the ternary complex between S1-RBD, ACE2 and HS supports the possibility for HS to bridge the complex between S1-RBD and ACE2, reinforcing their interaction and improving the chance of the virus to be internalized. These findings imply that highly specific inhibitors may not be optimal against the S protein and different HS-based structures, characterized by high affinity and including multidentate compounds, may be required.[86-88]

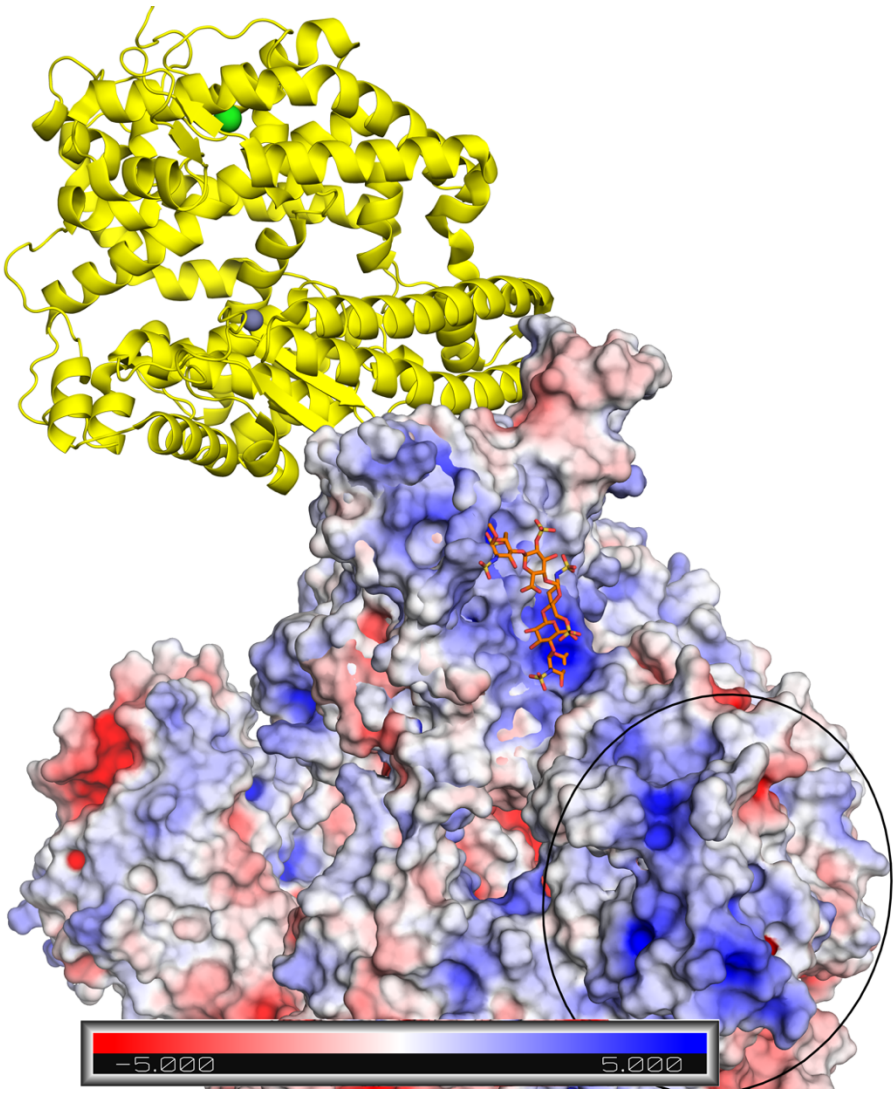

Figure 3.1.16. Hypothetical macromolecular complex obtained superimposing the (1)-S1-RBD system sampled by MD simulation and supported by tr-NOEs analysis on the S1-RBD subunit of the trimeric S protein (PDB ID: 6VYB) and the S1-RBD/ACE2 complex (PDB ID: 6M0J).

3.1.5. Materials and methods

3.1.5.1. Ligands and protein

Hexasaccharide (1) was synthetized by Dr. Minghong Ni (Ronzoni Institute), pentasaccharide (2) (Fondaparinux, Arixtra®) was supplied by Aspen France, nonasaccharide (3), obtained by chemoenzymatic synthesis, was provided by Prof. Dr. Jian Liu (University of North Carolina).

Recombinant SARS-CoV-2 spike protein (S1-RBD), expressed in HEK293 cells, was purchased from Sino Biological (Cat: 40592-V08H). Lyophilized S1-RBD aliquots (20 mM phosphate buffer pH 7.4, 300 mM NaCl, 10% glycerol, 5% trehalose, 5% mannitol and 0.01% Tween80) were first reconstituted with H_2O and then washed with 20 mM HEPES-d18 pH 7.2 with 200 mM NaCl (D_2O or H_2O) using VWR® centrifugal filters (10 kDa membrane, 0.5 mL).

3.1.5.2. NMR experiments

NMR experiments were performed with a Bruker Avance III spectrometer equipped with a high-sensitivity 5 mm TCI cryoprobe operating at 600.13 MHz.

For the NMR characterization of the oligosaccharides (1) and (3), ^{1}H, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC spectra were acquired.

For the STD NMR measurements, each oligosaccharide was dissolved in 0.2 mL of a purified S1-RBD solution, yielding a molar ratio of ligand:protein 100:1. The final concentrations were approximately 1 mM for the ligand and 10 μM for the protein. All spectra were recorded using the pulse sequence stddiffesgp.3 at 293 K. The number of scans were 2048 for hexasaccharide (1) and 1024 for both pentasaccharide (2) and nonasaccharide (3), the number of dummy scans 16, the recycle delay and saturation time 6 s and 3 s, respectively. A 10 ms spin-lock pulse was used to remove the broad resonances of the protein. The on-resonance irradiation was performed at 580 Hz, whereas the off-resonance control irradiation at 20000 Hz. The STD NMR spectrum was obtained by a phase cycling subtraction of the on-resonance and off-resonance data acquired in a interleaved mode. The STD intensities (I_{STD}) were calculated using Equation 2.6, considering only selected and well-separated signals (*i.e.*, anomeric protons). The normalization of the STD intensities against the most intense signal, which was assigned a value of 100%, provided the relative STD percentages. A control experiment was performed for each ligand without the protein

and no signal of direct irradiation was observed (Figure A3.1.4). The STD NMR experiment for the (2)-S1-RBD complex in D_2O buffer was run with both 3 s and 2 s saturation time to verify the spin diffusion effect. The obtained spectra were approximately superimposable, indicating that a slight decrease of the saturation time does not affect them (Figure A3.1.5). Saturation time smaller than 2 s did not allow the acquisition of STD NMR spectra with a reasonable signal-to-noise ratio. Long saturation times were required because the long distance of the glycan rings from the protein surface, for the presence of the sulfate groups decorating the GAG structure, determines a low efficiency in the magnetization transfer from the protein to the ligand. The STD NMR spectra of (2) in interaction with S1-RBD were also recorded in H_2O buffer to verify the binding of the amine groups. The STD NMR analysis of (3)-S1-RBD mixture was also acquired at 280 K to define the involvement in the recognition process of GlcA(I) and GlcA(G). This was not possible at 293 K because of the signals of H1 GlcA(I) and H1 GlcA(G) are overlapped to the H_2O signal.

For the NOESY experiment, 150 μg hexasaccharide (1) was solubilized in 0.2 mL of 20 mM HEPES-d18 pH 7.2 with 200 mM NaCl (D_2O); for the tr-NOESY experiment, the sample was prepared by dissolving 150 μg hexasaccharide (1) and 300 μg S1-RBD in 0.2 mL of the aforementioned buffer, reaching a molar ratio of ligand:protein 7:1. The final concentrations were about 400 μM for the ligand and 60 μM for the protein. All NOESY and tr-NOESY experiments were performed at 280 K. The lower temperature was used to avoid the overlap of the H_2O signal with both H5 IdoA2S(F, B) and H1 GlcA(D) signals and, thus, to measure the NOE and tr-NOE values for these protons. A total of 20 scans were collected for each free induction decay (2048×256 points), the data were zero-filled to 2048×1024 points before the Fourier transformation and the mixing times of 0.15, 0.3 and 0.5 s were used. The NOE and tr-NOE values are expressed as a percentage of the mean value of the diagonal peaks of H1 IdoA2S(F), H1 GlcNS6S(E), H1 GlcNS6S(C) and H1 IdoA2S(B).

3.1.5.3. Docking calculations

The AutoDock 4.2 software was used for the docking calculations.[89] The 3D model of S1-RBD was extracted from the crystal structure of S1-RBD/ACE2 (PDB ID: 6M0J). The model of (1) was built considering GlcNS6S(E, C, A) and GlcA(D) in $^4C_1$ chair conformation and IdoA2S(F, B) in $^1C_4$ chair conformation. From the non-reducing to the reducing end, the glycosidic dihedral angles (φ_i and ψ_i) were configured as follows: φ_1/ψ_1 = 41°/14°, φ_2/ψ_2 = –40°/–30°, φ_3/ψ_3 = 60°/30°, φ_4/ψ_4 = –39°/–33°, φ_5/ψ_5 = 41°/14°. The model

of (2) was constructed taking into account the following ring conformations: $^4C_1$ chair for GlcNS6S(E, C, A) and GlcA(D) and $^2S_0$ skew-boat for IdoA2S(B). The φ_i and ψ_i dihedral angles at the glycosidic linkages were defined as: φ_1/ψ_1 = –48°/–28°, φ_2/ψ_2 = 57°/10°, φ_3/ψ_3 = –36°/–45°, φ_4/ψ_4 = 45°/11°. The conformation of the sulfate groups was set up in agreement with the heparin model 1HPN reported in the Protein Data Bank (PDB).

Three docking simulations of (1) were run, targeting sites I, II and III of S1-RBD, to scan among the interactions of the ligand with different binding sites and to build the geometries of the complexes. The docking grid to target site I (Lx = Ly = 80, Lz = 100 points) was centred on Cα of the K356 residue. The centre was then manually translated (CMx = –40.272, CMy = 25.581, CMz = 31.746) to preserve enough free volume in the box, allowing (1) to explore efficiently the conformational space of the receptor. The docking grids to target site II (Lx = Ly = 80, Lz = 100 points) and site III (Lx = Ly = 90, Lz = 80 points) were centred on CMx = –29.378, CMy = 43.024, CMz = 22.511 and CMx = –26.181, CMy = –2.000, CMz = –4.806, respectively. All the torsional degrees of freedom of (1), except the glycosidic dihedral angles (φ_i and ψ_i), were free to be sampled. The receptor was set as rigid. In each docking simulation, the Lamarckian genetic algorithm search was run, setting the following parameters: number of runs = 100, population size = 2000 individuals, maximum number of energy evaluation = 2.5×10^7, maximum number of generations = 270000. The docking solutions were re-clustered at a root-mean-square deviation (RMSD) tolerance of 2 Å. For each targeted binding site, at least two docking solutions (or clusters) were selected, considering three criteria: the binding energy score [AutoDock scoring function (ADT)], the population of each cluster and the proximity to the putative binding site.

The (2)-S1-RBD complex geometries were built docking the ligand on the favourite binding site (site I). The docking grid (Lx = Ly = Lz = 100 points) was centred on CMx = –40.272, CMy = 25.585, CMz = 27.018. The parameters were set and the docking solutions were analysed as described for (1).

3.1.5.4. MD simulations

The geometries of the (1)-S1-RBD complexes obtained by docking calculations were refined by MD simulations in explicit solvent, until a suitable equilibration condition was achieved. The equilibration was monitored following two variables: the distance (RMSD) of the relative position of (1) from the corresponding initial value and the set of the dihedral angles φ_i/ψ_i. For each putative (1)-S1-RBD complex, these variables were plotted versus time, until

an oscillatory and stationary behaviour was reached. The shortest time required to achieve this condition was considered as the MD equilibration time of the system. The (1)-S1-RBD complex was described using the state-of-the-art of Amber (ff14SB) and Glycam06 force fields for the protein and the glycan, respectively.[90,91] Each simulated complex was surrounded by a 15 Å wide layer of the water model TIP3P to obtain an orthogonal simulation box, where a periodic boundary condition was applied (AmberTools14 software).[92] Non-bonded potential energy terms (electrostatic and Van der Waals terms) were cut-off at 12 Å; no counter ions were added. Each system was minimized using 200000 steps of the energy minimization algorithm included in the NAMD 2.14 software.[93] Each cell box was further submitted to MD simulation in the canonical NPT ensemble (constant number of particles, pressure and temperature). The temperature was set as 300 K by a Lowe-Andersen thermostat, while a Nosé-Hoover-Langevin piston algorithm was employed to control the pressure on the edges of the simulation box (1.01325 bar). During the first phase of the MD simulation, the geometry of the complex was restrained, applying a harmonic potential to each atom of the solute (K_i = 0.5 Kcal mol^{-1}); this allowed to equilibrate the density of the cell box in a period variable between 12 and 16 ns. After this stage, the harmonic restraint was removed, enabling the equilibration of the complex geometry, in which the protein-ligand relative position and contacts can evolve, following the intermolecular forces of the system. Finally, a MD simulation trajectory (between 36 and 80 ns) was sampled. The free energy of binding was estimated on the last 20 ns of the MD simulation production stage; a sampling frequency of 100 ps was applied.

The same procedure and settings described for the (1)-S1-RBD complexes were used to perform MD simulations, in explicit solvent for a period between 56 and 100 ns, on the (2)-S1-RBD complex.

The free energy of binding (ΔG_{PB}^{bind}) was estimated through the approximated method MMPBSA and the single trajectory scheme as implemented in AmberTools14 (MMPBSA.py application).[94-96] This approximation is accepted when the free energy of binding is employed to rank a set of predicted molecular recognition epitopes of a single protein-ligand pair.[97] The MMPBSA.py application was run setting the atoms radii for the Poisson Boltzmann calculation as included in the Amber (ff14SB)/Glycam06 force fields (radiopt = 0), while the total non-polar solvation free energy is modelled as a single term linearly proportional to the solvent-accessible surface area (inp = 1).

A reversible binding process between a ligand (L) and a receptor (R) could be summarized as:

$$L + R \rightleftarrows L - R \qquad \text{(Equation 3.1)}$$

The free energy of binding (ΔG^{bind}) could be estimated considering only the initial [G(L) and G(R)] and final [G(L–R)] states of the system:

$$\Delta G^{bind} = G(L - R) - G(L) - G(R) \qquad \text{(Equation 3.2)}$$

The free energy (G) of each state of the system could be further defined as:

$$G = \langle E_{MM} \rangle + \langle G_{sol} \rangle + \langle G_{noPol} \rangle + T\langle S_{MM} \rangle \qquad \text{(Equation 3.3)}$$

where E_{MM} is the molecular mechanic energy, G_{sol} is the polar solvation energy (estimated by the Poisson Boltzmann equation), G_{noPol} is the non-polar solvation energy (assessed by the solvent-accessible surface area of the molecule), T is the absolute equilibrium temperature and S_{MM} is the entropy of the molecule. The angle bracket indicates an averaging process on the overall visited states of the system (*i.e.*, selected set of snapshots generated by MD simulation). The last assumption allows the entropic term in Equation 3.3 to be neglected; indeed, in this case, all the tested protein-ligand pairs present a similar entropy change upon binding. In this study, the (1)-S1-RBD complexes were ranked in terms of the Poisson Boltzmann free energy of binding (ΔG_{PB}^{bind}) (Equation 3.4).

$$\Delta G_{PB}^{bind} = \Delta[\langle E_{MM} \rangle + \langle G_{sol} \rangle + \langle G_{noPol} \rangle] \qquad \text{(Equation 3.4)}$$

To quantify the contribution of the different residues characterizing both ligand and protein to the total free energy of binding (ΔG_{PB}^{bind}), the latter was decomposed according to Equation 3.5.

$$\Delta G_{PB}^{bind} = \Sigma_i \Delta G_{PB}^{bind}(i) \qquad \text{(Equation 3.5)}$$

In this description, the theoretical (1)-S1-RBD binding epitope was characterized by $\Delta G_{PB}^{bind}(i)$ values greater or lower than zero, indicating favourable or unfavourable contributions to the free energy of binding, respectively. An arbitrary cut-off criterion ($|\Delta G_{PB}^{bind}(i)| < 0.1$ Kcal mol^{-1}) was applied to limit the number of residues to be analysed.

The electrostatic potential (EP) map of selected macromolecules or macromolecular complexes were calculated by the APBS plugin included in Pymol 2.3.4, using the partial charges of the Amber force field (ff14SB). It was then plotted on the solvent-excluded

surface of the proteins (default option of the APBS plugin); the gradient colour from red to blue indicates areas with negative to positive values of the EP in KT e^{-1} units.

3.1.5.5. Complete relaxation and conformational exchange matrix analysis (CORCEMA)

The docking poses of (1) interacting with site I of S1-RBD in mode A or B, characterized by the lowest scores (ADT and ΔG_{PB}^{bind}), were submitted to MD simulations. A reduce set of geometries was sampled in the production stage and selected to mimic the average geometry of the ligand on the protein surface. The geometries were picked considering the smallest distances (RMSD) between the sampled snapshot of (1) at time *t* and the average structure of (1) (AVE) [RMSD (t - AVE)]. The average geometry of the complex, calculated on the production stage, cannot be used because it may not have physical meaning. The averaging process on a set of conformationally different chemical structures may not preserve the structural restraints (bond lengths, bond angles and dihedral angles). Three snapshots of the (1)-S1-RBD complex belonging to mode A at MD simulation times 81.58 ns, 91.89 ns and 92.13 ns [RMSD (81.58 - AVE) = 0.35 Å, RMSD (91.89 - AVE) = 0.36 Å, RMSD (92.13 - AVE) = 0.34 Å] and three ones belonging to mode B at MD simulation times 37.75 ns, 41.14 ns and 42.48 ns [RMSD (37.75 - AVE) = 0.35 Å, RMSD (41.14 - AVE) = 0.34 Å, RMSD (42.48 - AVE) = 0.34 Å] were chosen for the simulation of tr-NOEs.

The thermodynamic dissociation constant K_d, describing the interaction between (1) and S1-RBD, was set at 4.5×10^{-7} M (see Chapter 3.2). The kinetic constant k_{off} was estimated to be 30 s^{-1} by a trial-and-error procedure in which the H1-H2 tr-NOEs of GlcNS6S(E, C, A) were simulated and matched with the corresponding experimental values in a specific range of mixing time ($0.15 < T_{mix} < 0.5$ s) (data not reported).[98] Some saturation behaviours (spin diffusion) were observed for $T_{mix} > 0.5$ s. By guessing K_d and k_{off}, the kinetic constant k_{on} was set at 6.67×10^{7} s^{-1} M^{-1}, a value in agreement with the theoretical limit rate for a chemical reaction limited by the diffusion of the reagents (10^{7}-10^{9} s^{-1} M^{-1}). The correlation time constants of (1) and S1-RBD were defined using as a reference the interaction between heparin oligosaccharides and AT in similar experimental conditions. The isotropic correlation times for (1) and S1-RBD [or (1)-S1-RBD] were set to 0.6 ns and 21 ns, respectively.

The selected geometries and the estimated parameters were used to simulate intra-residue and inter-glycosidic tr-NOEs by the CORCEMA program. The calculated tr-NOEs were

compared to the experimental ones in a specific range of mixing time ($0.15 < T_{mix} < 0.5$ s). In this range, none of the tr-NOE signals present spin diffusion. Mixing time values smaller than 0.15 s do not give a reasonable signal-to-noise ratio in tr-NOESY experiments. The R-factor, calculated as reported in Equation 3.6, quantifies the agreement between the simulated and experimental tr-NOEs.

$$R = \sqrt{\frac{\sum(I^{exp,k} - I^{calc,k})^2}{\sum(I^{exp,k})^2}} \qquad \text{(Equation 3.6)}$$

In Equation 3.6, $I^{exp,k}$ is the experimental tr-NOE value for a proton pair k ($trNOE_{Tmix}^{exp,k}$) and $I^{calc,k}$ is the tr-NOE value simulated using the CORCEMA algorithm ($trNOE_{Tmix}^{calc,k}$).

Appendix

Table A3.1.1. $^{1}H/^{13}C$ chemical shift assignment of hexasaccharide (1).

Proton	IdoA2S(F)	GlcNS6S(E)	GlcA(D)	GlcNS6S(C)	IdoA2S(B)	GlcNS6S(A)
1	5.17/101.8	5.60/100.2	4.62/104.5	5.43/99.0	5.25/101.9	5.03/101.0
2	4.33/76.1	3.28/60.7	3.40/75.5	3.30/60.3	4.33/78.4	3.31/60.2
3	4.39/66.7	3.68/72.5	3.87/78.5	3.65/72.4	4.21/71.5	3.72/72.0
4	3.68/80.5	3.78/79.6	3.80/79.7	3.77/79.1	4.10/78.3	3.82/79.0
5	4.90/70.7	4.00/71.6	3.84/79.1	4.10/71.5	4.81/71.8	3.99/71.2
6	-	4.29/69.0	-	4.48/68.6- 4.24/68.7	-	4.38/69.5
OCH_3	3.44/60.4	-	-	-	-	3.43/58.1

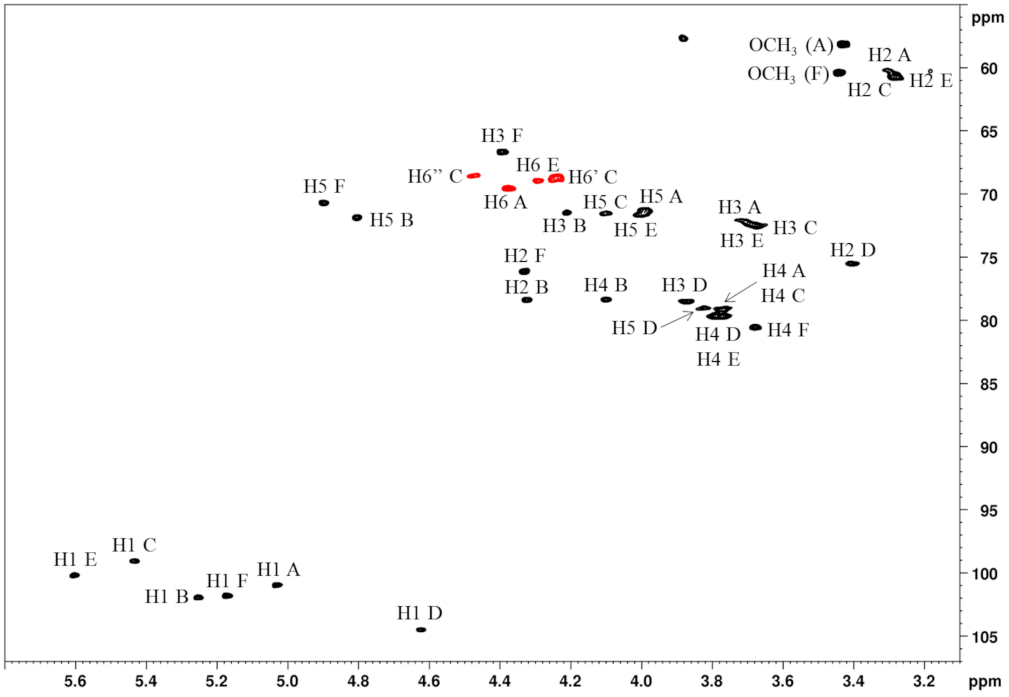

Figure A3.1.1. ^{1}H-^{13}C HSQC NMR spectrum of hexasaccharide (1).

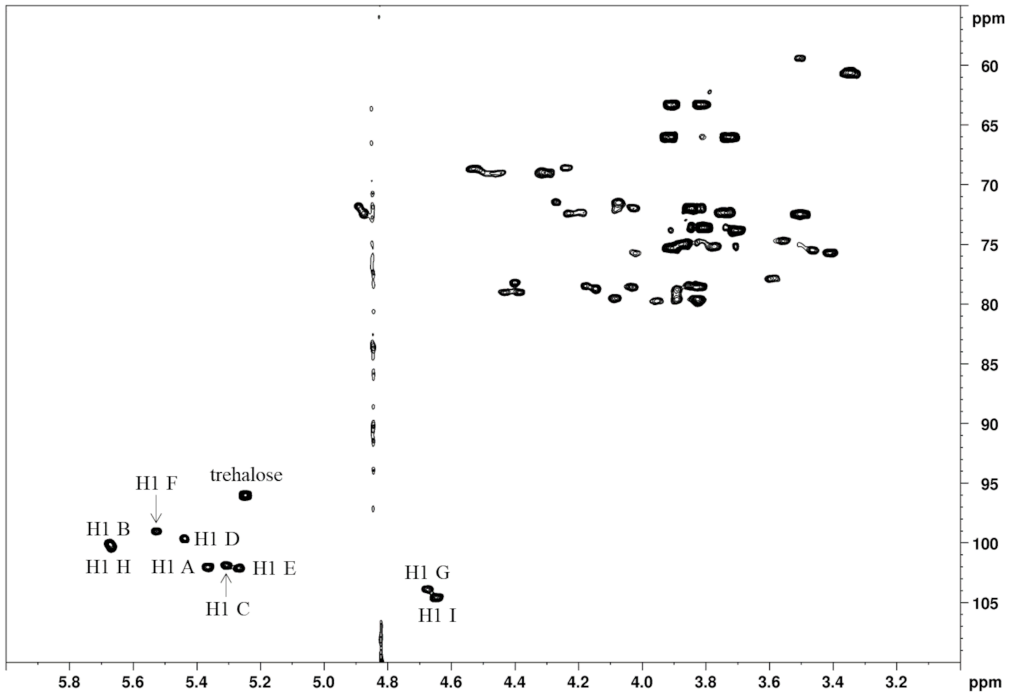

Figure A3.1.2. ^{1}H-^{13}C HSQC NMR spectrum of nonasaccharide (3).

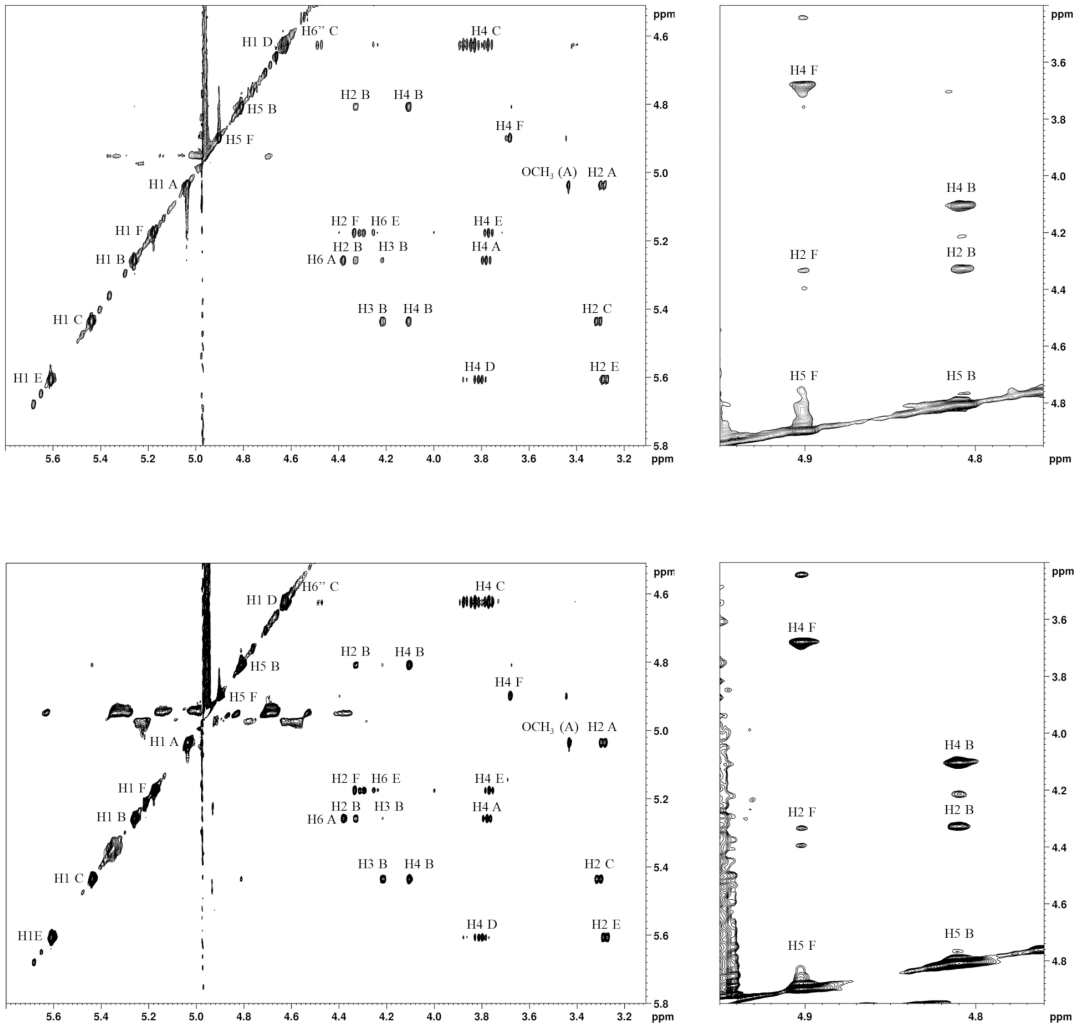

Figure A3.1.3. NOESY spectrum of hexasaccharide (1) (top panels). Tr-NOESY spectrum of hexasaccharide (1) interacting with S1-RBD (bottom panels).

Table A3.1.2. Simulated and experimental H5-H2 and H5-H4 tr-NOEs of IdoA2S(F, B) in the (1)-S1-RBD complex. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [81.58 ns (mode A) and 37.75 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)				*IdoA2S(B)*				
	H5-H2		H5-H4		H5-H2		H5-H4		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	1.37	1.03	7.89	9.97	3.28	3.83	10.76	11.15	
0.3	3.29	2.01	17.29	19.54	8.21	7.67	21.96	21.11	
0.5	7.05	4.58	33.65	32.01	18.12	12.77	39.22	29.15	
R (A+B)	0.30		0.01		0.12		0.07		ΣR=0.5
R (A)	0.38		0.32		1.91		0.91		ΣR=3.52
R (B)	0.27		0.30		0.55		0.22		ΣR=1.34

Table A3.1.3. Simulated and experimental H5-H2 and H5-H4 tr-NOEs of IdoA2S(F, B) in the (1)-S1-RBD complex. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [91.89 ns (mode A) and 41.14 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)				*IdoA2S(B)*				
	H5-H2		H5-H4		H5-H2		H5-H4		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	1.5	1.03	8.6	9.97	3.4	3.83	10.8	11.15	
0.3	3.6	2.01	18.9	19.54	8.2	7.67	22.2	21.11	
0.5	7.8	4.58	37.0	32.01	17.9	12.77	39.8	29.15	
R (A+B)	0.51		0.02		0.12		0.08		ΣR=0.73
R (A)	0.48		0.45		1.75		0.71		ΣR=3.39
R (B)	0.03		0.21		0.49		0.10		ΣR=0.83

Table A3.1.4. Simulated and experimental H5-H2 and H5-H4 tr-NOEs of IdoA2S(F, B) in the (1)-S1-RBD complex. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [92.13 ns (mode A) and 42.48 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)				*IdoA2S(B)*				
	H5-H2		H5-H4		H5-H2		H5-H4		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	1.48	1.03	8.30	9.97	2.72	3.83	9.47	11.15	
0.3	3.50	2.01	17.55	19.54	6.60	7.67	18.80	21.11	
0.5	7.29	4.58	32.79	32.01	14.21	12.77	32.67	29.15	
R (A+B)	0.37		0.00		0.02		0.01		ΣR=0.40
R (A)	0.53		0.16		0.67		0.23		ΣR=1.59
R (B)	0.27		0.17		0.53		0.14		ΣR=1.11

Table A3.1.5. Simulated and experimental inter-glycosidic tr-NOEs of (1) interacting with S1-RBD. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [81.58 ns (mode A) and 37.75 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)-GlcNS6S(E)				*GlcNS6S(E) -GlcA(D)*		*GlcA(D)-GlcNS6S(C)*				*GlcNS6S(C)-IdoA2S(B)*				*IdoA2S(B)-GlcNS6S(A)*				
	H1-H4		H1-H6		H1-H4		H1-H4		H1-H6		H1-H4		H1-H3		H1-H4		H1-H6		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	10.1	7.9	2.1	3.1	4.4	9.5	4.9	14.8	3.3	4.9	10.8	9.8	3.4	7.7	9.2	9.7	4.5	8.1	
0.3	19.0	14.9	3.8	6.9	9.4	18.3	10.2	28.9	6.9	8.8	20.7	17.3	6.8	13.8	16.6	16.9	9.1	15.8	
0.5	30.3	26.2	6.0	15.0	17.3	32.8	18.9	41.7	12.0	14.1	34.8	29.2	12.7	23.7	26.1	28.6	16.1	26.9	
R (A+B)	0.04		0.32		0.23		0.35		0.03		0.04		0.24		0.01		0.17		ΣR=1.43
R (A)	0.40		0.59		0.03		0.10		0.03		0.71		0.13		0.03		0.01		ΣR=2.03
R (B)	0.07		0.16		0.60		0.77		0.07		0.22		0.39		0.02		0.51		ΣR=2.81

Table A3.1.6. Simulated and experimental inter-glycosidic tr-NOEs of (1) interacting with S1-RBD. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [91.89 ns (mode A) and 41.14 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)-GlcNS6S(E)				*GlcNS6S(E) -GlcA(D)*		*GlcA(D)-GlcNS6S(C)*				*GlcNS6S(C)-IdoA2S(B)*				*IdoA2S(B)-GlcNS6S(A)*				
	H1-H4		H1-H6		H1-H4		H1-H4		H1-H6		H1-H4		H1-H3		H1-H4		H1-H6		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	15.0	7.9	1.3	3.1	5.3	9.5	5.4	14.8	4.3	4.9	9.2	9.8	5.2	7.7	12.2	9.7	3.6	8.1	
0.3	27.0	14.9	2.7	6.9	10.8	18.3	11.0	28.9	9.0	8.8	18.1	17.3	9.1	13.8	23.7	16.9	7.1	15.8	
0.5	41.9	26.2	5.0	15.0	19.0	32.8	19.3	41.7	15.8	14.1	30.9	29.2	14.2	23.7	39.5	28.6	12.1	26.9	
R (A+B)	0.45		0.43		0.18		0.33		0.01		0.00		0.15		0.14		0.30		ΣR=1.99
R (A)	0.74		0.49		0.09		0.08		0.13		0.32		0.28		0.66		0.18		ΣR=2.97
R (B)	0.30		0.38		0.29		0.77		0.03		0.21		0.08		0.01		0.46		ΣR=2.53

Table A3.1.7. Simulated and experimental inter-glycosidic tr-NOEs of (1) interacting with S1-RBD. The simulated tr-NOEs are reported as weighted averages of modes A and B (50% + 50%) represented by snapshots of the corresponding MD simulation [92.13 ns (mode A) and 42.48 ns (mode B)]. The R-factor quantifies the agreement between the simulated and experimental tr-NOEs generated by the weighted average of modes A and B or the pure mode A or B. The sum of the R-factor values (ΣR) shows that the quality of the fitting increases when both modes A and B are considered.

	IdoA2S(F)-GlcNS6S(E)				*GlcNS6S(E) -GlcA(D)*		*GlcA(D)-GlcNS6S(C)*				*GlcNS6S(C)-IdoA2S(B)*				*IdoA2S(B)-GlcNS6S(A)*				
	H1-H4		H1-H6		H1-H4		H1-H4		H1-H6		H1-H4		H1-H3		H1-H4		H1-H6		
T_{mix}	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	**Sim**	**Exp**	
0.15	9.9	7.9	2.5	3.1	4.6	9.5	3.5	14.8	6.0	4.9	10.0	9.8	3.0	7.7	11.1	9.7	5.2	8.1	
0.3	18.4	14.9	4.3	6.9	9.3	18.3	7.0	28.9	10.5	8.8	18.8	17.3	5.8	13.8	19.9	16.9	9.4	15.8	
0.5	29.4	26.2	6.7	15.0	25.7	32.8	12.0	41.7	15.5	14.1	30.5	29.2	9.9	23.7	30.4	28.6	14.7	26.9	
R (A+B)	0.03		0.27		0.10		0.53		0.02		0.00		0.34		0.01		0.19		ΣR=1.49
R (A)	0.40		0.52		0.07		0.35		0.46		0.21		0.35		0.04		0.01		ΣR=2.41
R (B)	0.11		0.13		0.53		0.76		0.17		0.13		0.33		0.01		0.58		ΣR=2.75

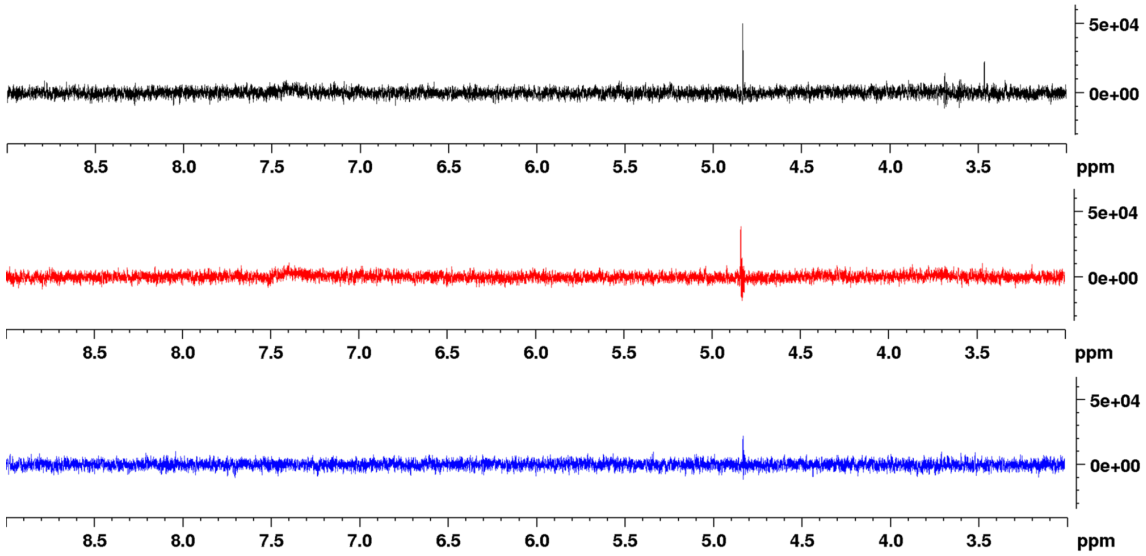

Figure A3.1.4. STD NMR control spectra of hexasaccharide (1) (black line), pentasaccharide (2) (red line) and nonasaccharide (3) (blue line) acquired in the same condition as STD NMR spectra reported in Figures 3.1.4, 3.1.6 and 3.1.7.

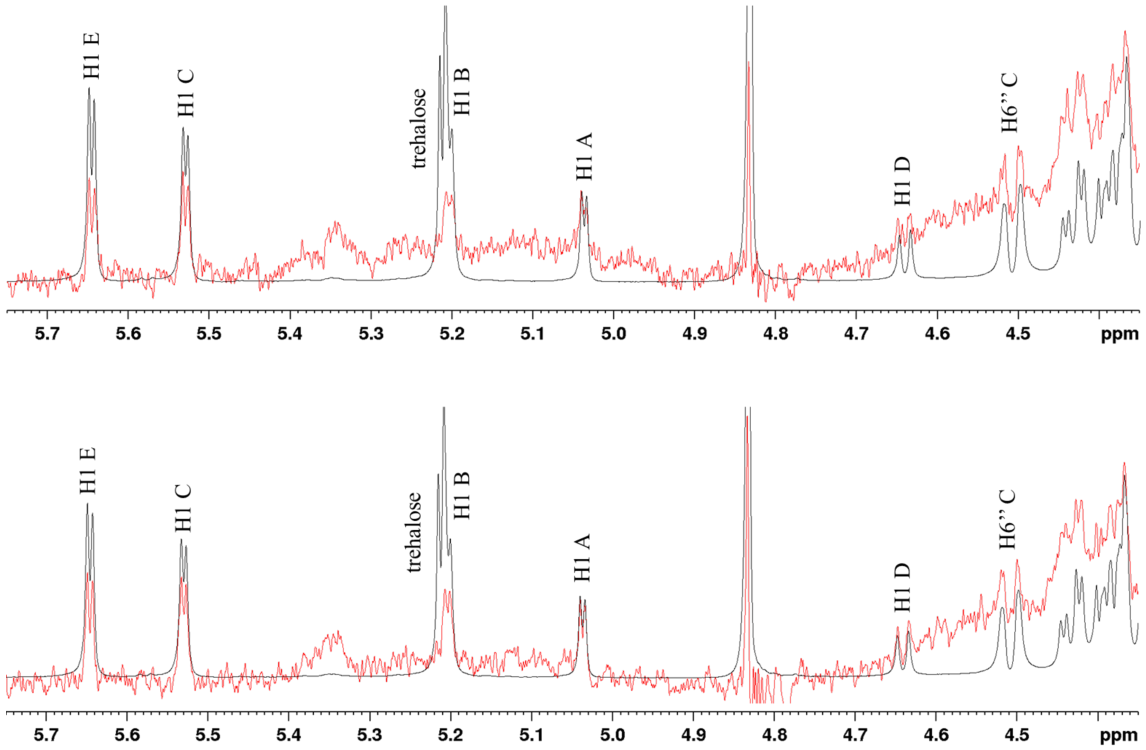

Figure A3.1.5. STD NMR analysis of pentasaccharide (2) in interaction with S1-RBD. Comparison of the STD NMR spectra (red lines) superimposed on the corresponding reference spectra (black lines) acquired at 2 s saturation time (upper panel) and 3 s saturation time (bottom panel).

3.2. Inhibition of SARS-CoV-2 entry by enoxaparin

3.2.1. Introduction

SARS-CoV-2 infection is associated with hypercoagulability caused by direct invasion of endothelial cells and/or release of proinflammatory cytokines.[89] Heparin and low molecular weight heparins (LMWHs), widely used as anticoagulant drugs, have shown potential benefits in managing the high susceptibility to blood clotting in critically ill COVID-19 patients.[100,101] These patients have an increased risk of developing life-threatening thrombotic complications, such as pulmonary embolism and deep vein thrombosis. The administration of LMWHs aims to prevent the formation of clots and minimize the likelihood of these severe events.[102] Recent studies have reported a reduction in mortality rates and improved patient outcomes when therapeutic doses of LMWHs were utilized in the management of clotting disorders associated with severe COVID-19 conditions. In addition to their anticoagulant effects, LMWHs have shown promise as anti-inflammatory agents in controlling COVID-19 complications. The immune response triggered by SARS-CoV-2 can lead to an excessive release of proinflammatory cytokines, known as cytokine storm. This immune system overreaction can cause extensive damage to vital organs, particularly the lungs. The anti-inflammatory properties of LMWHs have been studied to potentially mitigate the cytokine storm and prevent its destructive effects on the body. Researchers have suggested that LMWHs may act as regulators of the immune response, modulating the release of inflammatory mediators and reducing the severity of lung injury in severe COVID-19 cases. Beyond their well-known anticoagulant and anti-inflammatory actions, LMWHs possess antiviral properties against viruses belonging to the Coronaviridae family, including SARS-CoV-2. Preliminary studies have revealed that heparin and LMWHs could inhibit SARS-CoV-2 entry into host cells, potentially blocking the initial stages of infection.[40] Although the antiviral activity of heparin and LMWHs is intriguing, further research is required to fully understand their mechanisms of action and determine their clinical significance in treating COVID-19.

3.2.2. Aim of the work

The aim of this study was to explore the capability of enoxaparin, a type of LMWH, to hinder SARS-CoV-2 attachment to host cells. To achieve this goal, a batch of enoxaparin was fractionated by size exclusion chromatography (SEC) to isolate components characterized

by specific molecular size. The isolated fractions were fully characterized via NMR and high-performance size exclusion chromatography combined with a triple detector array (HP-SEC/TDA). Then, they were subjected to isothermal titration calorimetry (ITC) analysis to determine their binding to S1-RBD and biological assays to evaluate their antiviral activity. As described in Chapter 3.1, S1-RBD is the portion of the SARS-CoV-2 S protein that recognizes and binds the host cell ACE2 receptor. By assessing the interaction between enoxaparin fractions and S1-RBD, our purpose was to determine their capacity to disrupt viral attachment, thereby preventing the initial stages of viral invasion. The biological assays, which involved in vitro studies using cell cultures infected with SARS-CoV-2, was carried out to examine the impact of enoxaparin fractions on viral replication and infectivity. This study represents a critical step in understanding the multifaceted pharmacological properties of enoxaparin and its potential role in counteracting COVID-19.

3.2.3. Results

3.2.3.1. Isolation, purification and characterization of enoxaparin fractions

A batch of enoxaparin (D41) was fractionated as outlined in Figures 3.2.1. Details of the fractionation process are described in the experimental section. All the isolated fractions were structurally characterized via NMR and HP-SEC/TDA. The most relevant compositional data determined via NMR and the molecular weights defined by HP-SEC/TDA are reported in Tables 3.2.1, 3.2.2 and 3.2.3. The monosaccharide and disaccharide composition of the fractions is quite similar to each other (Table 3.2.1). Except for the non-reducing and reducing residues (1,6-anhydro GlcNS and GlcNS6S-αRed), whose percentage contents are inversely proportional to the molecular weight, other units do not show significant variations. The average degree of polymerization (dp) was calculated through the ratio between the total volume of the anomeric signals (uronic acids, glucosamines and linkage region residues) and that of the anomeric signals of the residues at the reducing end (Table 3.2.2). The fractions with lower molecular weights (G15205_B to G15205_G) exhibit a slightly higher degree of sulfation (as for the non-reducing and reducing residues, the degree of sulfation is also inversely proportional to the molecular weight). Size homogeneous fractions with a very low degree of polydispersity (~1) were obtained (Table 3.2.3).

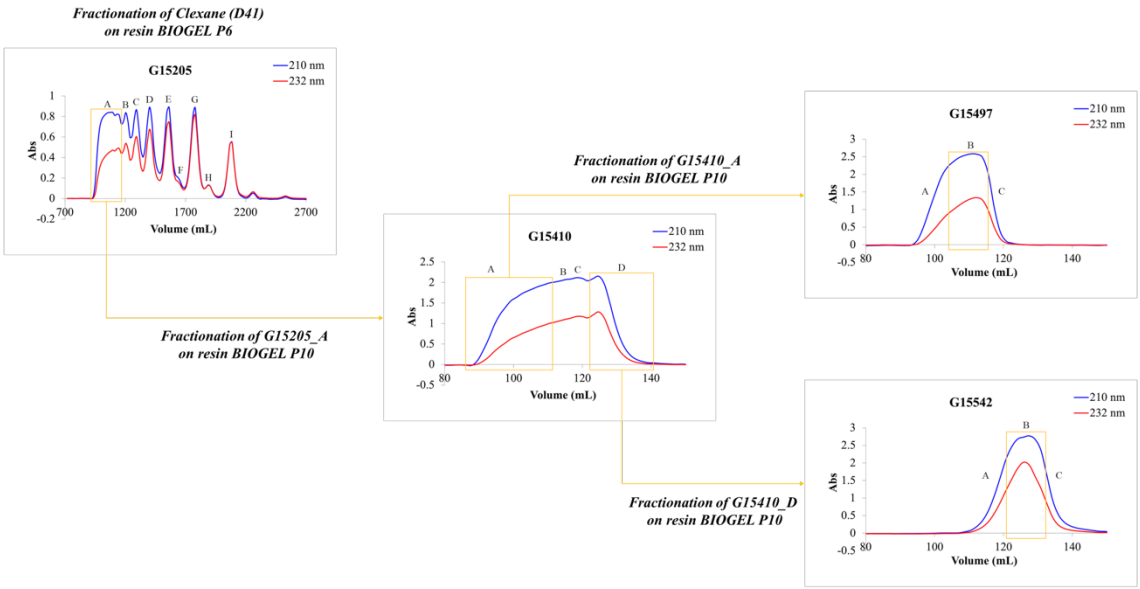

Figure 3.2.1. Scheme of enoxaparin fractionation by size exclusion chromatography (SEC).

Table 3.2.1. Percentage contents of IdoA2S, GlcA-GlcNS3S6S, GlcNAc6X, 1,6-anhydro GlcNS, GlcNS6S-αRed, 6-O-sulfation of GlcN (A6S), degree of sulfation (DS) of the isolated enoxaparin fractions determined via NMR.

Fraction	IdoA2S	GlcA-GlcNS3S6S	GlcNAc6X	1,6-anhydro GlcNS	GlcNS6S-αRed	A6S	DS
G15205_A	56.7	3.3	14.9	1.0	3.4	80.6	2.37
G15205_B	54.8	3.2	13.3	1.5	4.1	81.5	2.42
G15205_C	55.2	2.8	11.9	1.8	5.0	82.0	2.44
G15205_D	56.2	2.9	9.8	2.1	6.5	82.8	2.48
G15205_E	53.0	2.8	7.2	3.1	9.8	83.9	2.59
G15205_G	50.8	2.9	2.9	4.2	14.4	85.1	2.69
G15497_B	57.9	3.8	15.8	0.5	2.0	81.1	2.36
G15410_B	55.3	3.5	14.8	0.9	2.7	81.1	2.36
G15410_C	56.0	3.3	13.9	1.0	3.1	81.3	2.38
G15542_B	53.9	3.8	14.4	1.2	3.6	81.4	2.35

Table 3.2.2. Degree of polymerization (dp) of the isolated enoxaparin fractions determined via NMR.

Fraction	dp
G15205_A	20
G15205_B	14
G15205_C	12
G15205_D	10
G15205_E	8
G15205_G	6
G15497_B	26
G15410_B	20
G15410_C	18
G15542_B	16

Table 3.2.3. Molecular weight parameters [number-average mean molecular weight (Mn), weight-average mean molecular weight (Mw) and polydispersity (Mw/Mn)] of the isolated enoxaparin fractions defined via HP-SEC/TDA.

Fraction	Mw (Da)	Mn (Da)	Mw/Mn
G15205_A	6772	6333	1.07
G15205_B	4473	4357	1.03
G15205_C	3615	3609	1.00
G15205_D	3112	3107	1.00
G15205_E	2554	2534	1.01
G15205_G	1888	1882	1.00
G15497_B	8046	8003	1.01
G15410_B	6186	6169	1.00
G15410_C	5441	5417	1.00
G15542_B	4807	4785	1.01

3.2.3.2. Evaluation of the affinity and activity of isolated enoxaparin fractions

The microcalorimetric technique ITC can be exploited to study protein-ligand interactions. It measures the heat absorbed or released upon binding and provides thermodynamic data, including the stoichiometry (N), the binding constant (K_d) and the enthalpy and entropy contributions to the binding (ΔH and ΔS, respectively) to deduce the interaction mechanism. ITC experiments were performed to study the interaction between selected enoxaparin fractions with different degree of polymerization (dp) and S1-RBD, evaluating the possible correlation between the binding affinity and the chain length. ITC thermograms, obtained adding an exact volume of ligand at regular intervals to a solution of S1-RBD, were fitted using the one set of sites model. Table 3.2.4 reports a summary of the collected data. Firstly, heparin (A5370) and enoxaparin (D41) were analysed. The same affinity ($K_d = 3\times10^{-7}$), a different stoichiometry (0.19 and 0.43 for heparin and enoxaparin, respectively) and a

favorable enthalpy term (–2.00 to –16.2 Kcal mol^{-1}) were observed for the aforementioned samples. The entropy term is favorable for heparin (–6.80 Kcal mol^{-1}) and unfavorable for enoxaparin (7.34 Kcal mol^{-1}). The titration of S1-RBD with the enoxaparin fractions showed that the binding affinity does not increase with the increasing of the chain length: the binding constants are $1\text{-}4\times10^{-7}$ M for all the isolated fractions (as the binding constant of the unfractionated enoxaparin). The N values are around 1 for the shortest enoxaparin fractions (fractions with dp 6 or dp 8 bind a single molecule of S1-RBD) and in the range 0.47-0.67 (as the N value of the unfractionated enoxaparin) for the enoxaparin fractions from dp 10 to dp 26 (these fractions bind two molecules of S1-RBD). The binding free energy turned out to be negative for all fractions. The enthalpic contribution is negative for the shortest and longest enoxaparin fractions (dp 6, dp 8, dp 20 and dp 26) and positive for the enoxaparin fractions from dp 10 to dp 18. The negative enthalpic contribution is presumably due to the formation of ionic interactions, hydrogen bonds and van der Waals interactions between the side chains of basic residues on the protein surface and the sulfate groups of the ligand. All enoxaparin fractions have a favorable entropy contribution (–3.58 to –12.5 Kcal mol^{-1}), probably due to a release of water molecules to bulk upon protein-ligand interaction.

Table 3.2.4. Summary of the ITC data.

Samples	dp	Mw (Da)	N_1 (sites)	K_{d1} (M)	ΔH_1 (Kcal mol^{-1})	ΔG_1 (Kcal mol^{-1})	$-T\Delta S_1$ (Kcal mol^{-1})
A5370	-	15766	0.19 ± 0.08	3.58×10^{-7} ± 2.89×10^{-7}	–2.00 ± 1.26	–8.80	–6.80
D41	-	4729	0.43 ± 0.04	3.04×10^{-7} ± 1.38×10^{-7}	–16.2 ± 2.73	–8.89	7.34
G15205_G	6	1888	1.32 ± 0.12	4.47×10^{-7} ± 1.75×10^{-7}	–5.08 ± 0.72	–8.67	–3.58
G15205_E	8	2554	1.13 ± 0.06	3.84×10^{-7} ± 1.19×10^{-7}	–19.4 ± 1.86	–8.75	–10.6
G15205_D	10	3112	0.67 ± 0.02	2.59×10^{-7} ± 0.72×10^{-7}	1.38 ± 0.10	–8.99	–10.4
G15205_B	14	4473	0.51 ± 0.03	2.91×10^{-7} ± 1.28×10^{-7}	2.27 ± 0.31	–8.92	–11.2
G15410_C	18	5441	0.53 ± 0.02	1.22×10^{-7} ± 0.45×10^{-7}	3.03 ± 0.22	–9.43	–12.5
G15410_B	20	6186	0.57 ± 0.03	2.01×10^{-7} ± 0.86×10^{-7}	–1.16 ± 0.12	–9.14	–7.98
G15497_B	26	8046	0.47 ± 0.02	2.82×10^{-7} ± 0.91×10^{-7}	–1.65 ± 0.17	–8.94	–7.28

The capability of enoxaparin and the longest enoxaparin fractions to inhibit SARS-CoV-2 cells invasion was evaluated using Vero cells model and compared to that of heparin. The analysis confirmed that heparin inhibits SARS-CoV-2 replication.[40] The same dose of

enoxaparin (100 µg/mL) was tested and a significant lower inhibition of plaque formation was observed (65% and 30% for heparin and enoxaparin, respectively). Surprisingly, all the enoxaparin fractions showed no activity, regardless of their molecular weight/chain length (Figure 3.2.2).

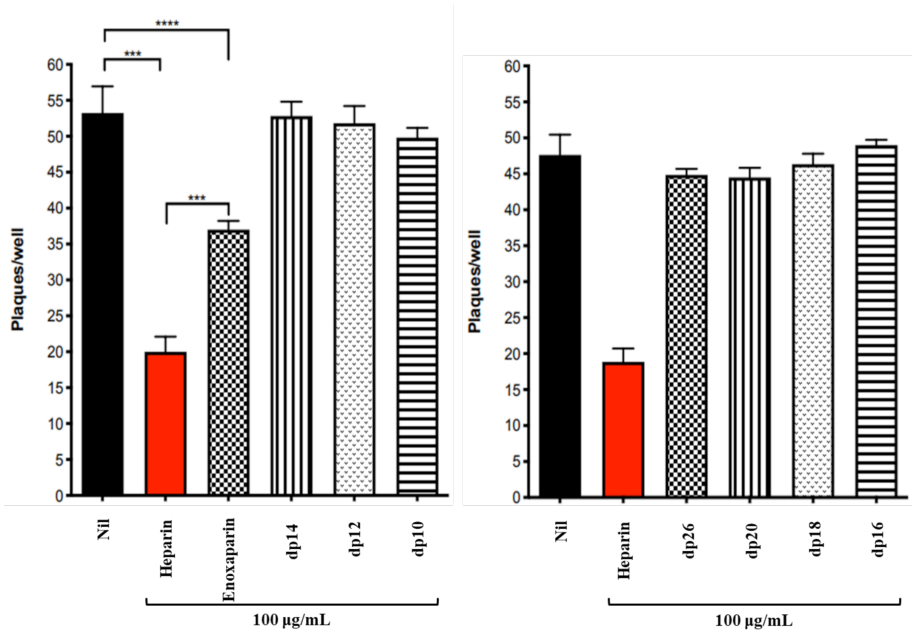

Figure 3.2.2. Effect of heparin, enoxaparin and the longest enoxaparin fractions (100 µg/mL) added 1h before infecting Vero cells with 50 plaque forming units (PFUs) of SARS-CoV-2. Nil is no treatment. The results are expressed as number of plaques/well and represent the mean ± the standard deviation of one experiment in quadruplicate cultures. The *p* value was calculated by the Mann-Whitney U test.

3.2.4. Conclusions

The ability of enoxaparin to inhibit SARS-CoV-2 cell invasion was tested and compared to that of unfractionated heparin. Although enoxaparin has shown less potency than heparin, our interest focused on assessing the possible correlation of the antiviral properties and the affinity to S1-RBD of enoxaparin with the degree of polymerization (dp) of its chains. To this end, a batch of enoxaparin was fractionated by SEC, obtaining fractions with different chain length (from dp 6 to dp 26). The inhibition of SARS-CoV-2 cell invasion was evaluated for all fractions using the plaque assay. Surprisingly, none of the fractions has shown an appreciable inhibitory activity. Further studies, to support the hypothesis that a cooperative effect between short and long chains is necessary to inhibit cell invasion, could be necessary. The affinity of the fractions to S1-RBD was investigated by ITC; preliminary results showed that the chain length does not affect the affinity to S1-RBD ($K_d = 10^{-7}$ M for

a wide range of chain lengths), while different protein-to-ligand ratio was observed for shorter (dp 6 and dp 8) and longer (> dp 8) chains, equal to 1:1 and 2:1, respectively.

3.2.5. Materials and methods

3.2.5.1. Ligands and protein

Enoxaparin (Clexane) was supplied by Sanofi. Recombinant SARS-CoV-2 spike protein (S1-RBD) was purchased from Sino Biological.

3.2.5.2. Fractionation by size exclusion chromatography

Fractionation by SEC to isolate enoxaparin fractions with different chain length was performed on Biogel P6 and Biogel P10 columns. On Biogel P6 column (5 × 170 cm), 300-350 mg sample dissolved in 5 mL purified water were loaded and eluted with 0.25 M NH_4Cl at a flow rate of 1.8 mL/min. The flow-through was collected in about 8 mL fractions and their UV absorbance was detected at 210-232 nm. The fractions of interest were collected, pooled and lyophilized. On Biogel P10 column (1.6 × 200 cm), 55-60 mg sample dissolved in 1 mL purified water were loaded and eluted with 0.25 M NH_4Cl at a flow rate of 0.16 mL/min. The flow-through was collected in about 2 mL fractions and their UV absorbance was detected at 210-232 nm. Also in this case, the fractions of interest were collected, pooled and lyophilized.

3.2.5.3. Desalting

Desalting was performed using HW40S TSK 5 × 85 cm and HW40S TSK 2.6 × 60 cm. Samples were loaded and eluted with 10% ethanol at a flow rate of 5 mL/min (HW40S TSK 5 × 85) or 1.4 mL/min (HW40S TSK 2.6 × 60 cm). Absorbance at 210-232 nm was evaluated for each fraction. The fractions of interest were collected, pooled and lyophilized.

3.2.5.4. NMR measurements

NMR spectra were measured on a Bruker Avance NEO 500 MHz spectrometer and Bruker Avance III 600 MHz spectrometer, both equipped with a 5 mm TCI cryoprobe. HSQC experiments were performed according to the method published by Ronzoni Institute.[103] Briefly, spectra were recorded at 303 K using the following acquisition parameters: number of scans 24, dummy scans 32, relaxation delay 2 s, spectral width 8 ppm (F2) and 80 ppm

(F1), transmitter off-set 4.7 ppm (F2) and 80 ppm (F1), $^{1}J_{C\text{-}H}$ = 150 Hz; 1024 points were recorded for each of 320 increments. The FIDs were processed as follows: spectrum size 4096 (F2) and 1024 (F1) (zero-filling in both dimensions), QSINE window multiplication in both dimensions and Fourier transformation. HSQC spectra were integrated using the standard Topspin routine according to the procedure published by Ronzoni Institute.[103]

3.2.5.5. Molecular weight determination by high performance-size exclusion chromatography combined with a triple detector array (HP-SEC/TDA)

A HP-SEC system with three detectors (refractive index, viscometer and light-scattering) [Viscotek-TDA-302 (Malvern Panalytical Instruments, UK)], equipped with the pump "SmartLine pump 1000" (Knauer) and the degasser "Degasi® GPC" (Biotech), was used for determining the molecular weights. The HP-SEC system was fitted out with G2500 and G3000 7.8 mm × 30 cm TSK GMPWXL Tosoh columns, based on hydroxylated polymethacrylate (Tosoh Bioscience GmbH, Germany). For each analysis, a concentration of about 5 mg/mL (heparin sample) and 10 mg/mL (enoxaparin samples), a flow rate of 0.6 mL/min, a temperature of 40 °C and an injection volume of 100 µL were set. As mobile phase, 0.1 M $NaNO_3$ was used. The chromatogram analysis was performed with the OmniSEC 4.6.2 software. For each sample, all molecular weight parameters [number-average mean molecular weight (Mn), weight-average mean molecular weight (Mw) and polydispersity (Mw/Mn)] were calculated.

3.2.5.6. Isothermal titration calorimetry (ITC) measurements

Lyophilized S1-RBD aliquots (20 mM phosphate buffer pH 7.4, 300 mM NaCl, 10% glycerol, 5% trehalose, 5% mannitol and 0.01% Tween80) were reconstituted with 0.4 mL of water. To remove the excess of preservatives, the protein solution was exchanged with 20 mM HEPES pH 7.2 containing 200 mM NaCl (HEPES buffer), using Amicon® Ultra centrifugal filters (10 kDa filter, 0.5 mL). The aliquots were preserved at 0-4 °C at a protein concentration of 10 mM. Lyophilized heparin and enoxaparin samples were initially resuspended with 0.5 mL of HEPES buffer. The stock solutions were then diluted at different concentrations for the ITC analysis. The latter were carried out at 25 °C using a MicroCal PEAQ-ITC microcalorimeter (Malvern). For all the experiments, S1-RBD was taken in the cell (2-10 mM) and enoxaparin fractions in the syringe (30-300 mM). The ITC measurements consisted of twenty injections of 2 µL with an initial delay of 180 s and a

delay of 180 s between two subsequent injections. The stirring rate was set to 500 rpm. To estimate the dissociation constants (K_d) and the enthalpy variations (ΔH), the data were fitted using the MicroCal PEAQ-ITC analysis software.

3.2.5.7. Viral plaque-forming assays

For cellular invasion assays, Vero cells (ECACC) were plated (2.5×10^5 cells per well in 24-well plates) and cultured in EMEM supplemented with 10% (v/v) foetal calf serum (complete medium). After 24 hours, the cells were exposed to heparin or enoxaparin samples (100 μg/mL) in 300 μL of complete medium; 1 hour later, they were incubated with the virus solution containing 50 plaque-forming units (PFUs) of Italy/UniSR1/2020 strain (GISAID accession ID: EPI_ISL_413489). After incubation for 1 hour at 37 °C, the supernatants were discarded and 500 μL of 1% (w/v) methylcellulose overlay (in complete medium) were added to each well. Three days later, the cells were fixed using a 6% (v/v) formaldehyde in phosphate-buffered saline solution and stained with 1% (w/v) crystal violet in 70% (v/v) methanol. The plaques were counted using a stereoscopic microscope (SMZ-1500, Nikon).

Chapter 4

4. Insights into the binding of glycol-split heparin oligosaccharides to heparanase

4.1. Introduction

4.1.1. Heparanase

Heparanase (HPSE) is an endo-β-D-glucuronidase that cleaves heparan sulfate (HS) chains on cell surfaces and in the extracellular matrix (ECM).[104-109] Specifically, it catalyzes the hydrolysis of the β (1-4) glycosidic bond between a D-glucuronic acid (GlcA) and the subsequent D-glucosamine (GlcN) within the HS chains.[110-113] The cleavage of HS by HPSE is fundamental in both physiological processes and pathological conditions. The enzymatic activity of HPSE is dysregulated in various diseases, such as cancer, inflammation, cardiovascular diseases and neurological disorders. In cancer, HPSE is often upregulated, promoting changes in the ECM that facilitate tumor growth and spread. Therefore, the inhibition of HPSE represents a therapeutic strategy for limiting cancer progression, angiogenesis and metastasis. The crystal structure of human HPSE in both apo form and bound to HS oligosaccharides was determined through X-ray crystallography (Figure 4.1).[114] It revealed the overall structure of the protein and the nature of the residues belonging to the active site. Moreover, it uncovered insights into the substrate recognition and the hydrolysis mechanism, providing a starting point for the design of novel HPSE inhibitors. Nevertheless, some key features, that allow HPSE to recognize HS, remain poorly understood. HPSE is produced as an inactive precursor called pro-HPSE. The latter is characterized by a linker peptide (6 kDa) which forms a α-helical domain sitting above part of the active site, preventing the binding with HS. The proteolytic cleavage of this peptide by lysosomal cathepsins gives rise to the active form. The protein consists of two major domains: a larger $(\beta/\alpha)_8$-TIM barrel domain (50 kDa) flanked by a smaller β-sandwich domain (8 kDa). The former, characterized by eight alternating β-strands and α-helices, presents a cleft containing the catalytic pair formed by the glutamic acids E225 and E343. These residues represent the proton donor (E225) and the nucleophile (E343) required for the hydrolysis mechanism. Moreover, the aforementioned cleft is lined with a number of arginine and lysine residues forming two heparin binding domains (HBDs), which play a key role in the HS recognition and binding.[115,116] The β-sandwich domain comprising eight β-strands is crucial for both

enzymatic and non-enzymatic activities of HPSE. The minimum HS-like sequence recognized by HPSE is a trisaccharide unit (GlcNX-GlcA-GlcNX).[117] The enzyme preferentially binds high sulfated domains within the HS chains. The development of HPSE inhibitors is an encouraging strategy to counteract the upregulation of the HPSE activity in cancer. Several HPSE inhibitors were tested, including sulfated polysaccharides (HS mimetics) and small molecules.[118-122]

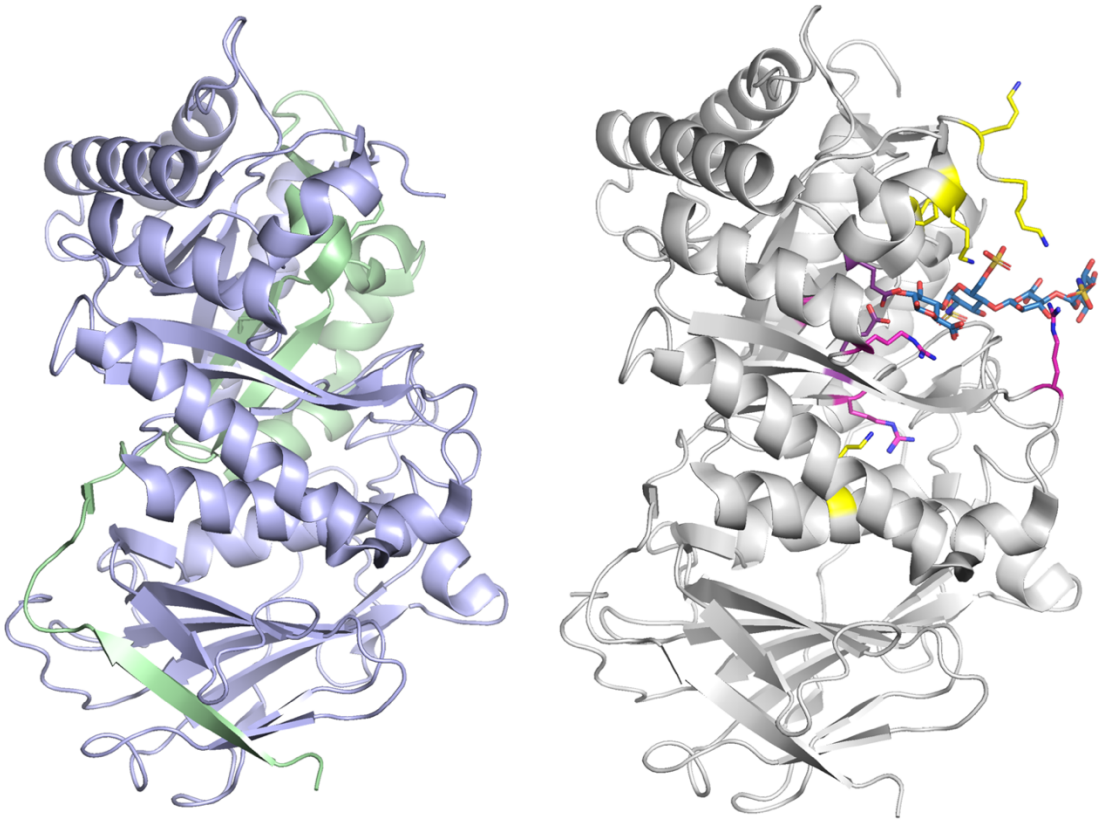

Figure 4.1. Crystal structure of human HPSE in its apo form is shown on the left of the panel. Two domains can be discerned: the $(\beta/\alpha)_8$-TIM barrel domain containing the HS-binding cleft (blue ribbon) and the β-sandwich domain (green ribbon). Crystal structure of human HPSE bound to a tetrasaccharide structure is shown on the right of the panel. HPSE is reported as gray ribbon, the glutamic acid residues of the catalytic pair are drawn as purple tubes, the arginine and lysine residues forming the two heparin-binding domains are displayed as fuchsia and yellow tubes.

4.1.2. Glycol-split heparins

Heparin has proven to inhibit the activity of HPSE as structurally similar to its natural substrate HS. Specifically, it binds to HPSE, preventing its interaction with the HS chains. This interference decreases the enzymatic activity of HPSE, reducing its ability to cleave HS on cell surfaces and in the ECM. The use of heparin in anticancer therapy is limited due to

its potent anticoagulant activity. Additionally, some of the glycosidic linkages between GlcA and GlcN within the heparin chains are susceptible to the cleavage of HPSE, resulting in a partial depolymerisation and loss of the inhibitory activity. To enable the administration of the drug without risking excessive bleeding, a number of heparin derivatives with low anticoagulant potency were generated and evaluated for their anti-HPSE activity. Glycol-split heparins, characterized by reduced anticoagulant activity and increased potency of inhibition against HPSE, are promising inhibitors of HPSE.[123,124] They are obtained from heparins through a chemical process that involves periodate oxidation followed by borohydride reduction and determines the C2-C3 ring opening of non-sulfated uronic acids. The medical uses of glycol-split heparins are favored by the loss of the anticoagulant activity associated with the splitting of the essential GlcA within the active site responsible for binding AT. As structural analogs of heparin and HS, they maintain affinity toward HPSE, but do not represent substrates for the enzyme due to the induced modification of GlcA. The mechanism of HPSE inhibition by glycol-split heparins has not been elucidated yet. A current hypothesis suggests that the conformational freedom of the glycol-split units favours the accommodation of the sugar chain within the active site of HPSE.

4.2. Aim of the work

The aim of this work was to shed light on the molecular basis underlying the interaction between glycol-split heparins and HPSE. Two distinct trisaccharides, here namely trisaccharide (1) and trisaccharide (2), were synthesized. The structure of trisaccharide (1) is D-GlcNS6S α (1-4) L-gs-IdoA α (1-4) D-GlcNS6S α (1-4) OMe, while that of trisaccharide (2) is D-GlcNS6S α (1-4) D-gs-GlcA β (1-4) D-GlcNS6S α (1-4) OMe (Figure 4.2). The only difference between the two glycans is the configuration of C5 in the glycol-split uronate moiety: trisaccharide (1) is characterized by glycol-split iduronic acid, while trisaccharide (2) presents a glycol-split glucuronic acid.

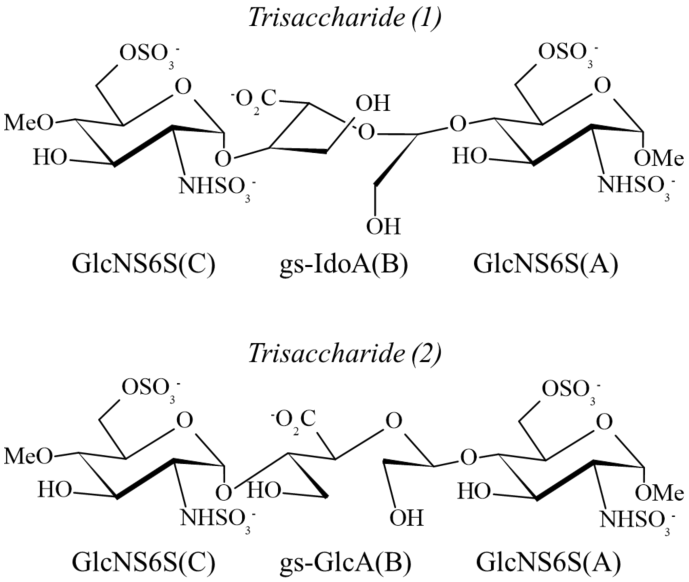

Figure 4.2. Chemical structures of trisaccharide (1) with glycol-split iduronic acid and trisaccharide (2) with glycol-split glucuronic acid.

The inhibitory activity of the two trisaccharides was determined by the colorimetric assay described in Hammond et al. (2010), reveling that only trisaccharide (1) is moderately able to inhibit HPSE (IC_{50}=50µg/mL) (Figure 4.3).[125] Briefly, the assay uses the synthetic pentasaccharide Fondaparinux as enzyme substrate and detects one of the products of the catalysis, which contains a newly formed reducing terminus, with the tetrazolium salt WST-1.[125] In the assay, Roneparstat, a non-anticoagulant 100% N-acetylated and glycol-split heparin acting as a potent heparanase inhibitor, was tested as reference.

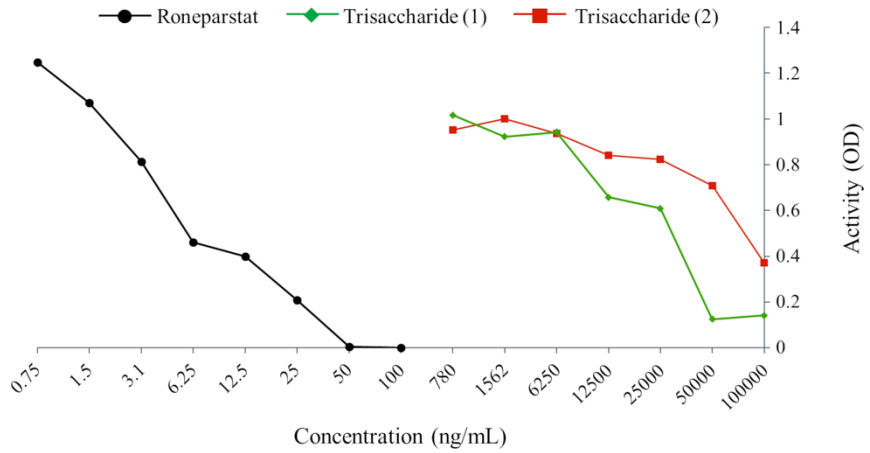

Figure 4.3. Dose-dependent inhibition of HPSE activity by the reference Roneparstat (black, IC_{50}=0.025µg/mL), trisaccharide (1) (green, IC_{50}=50µg/mL) and trisaccharide (2) (red, not active).

To correlate the inhibitory activity with specific binding properties, the interaction of trisaccharides (1) and (2) to HPSE was investigated by a combined saturation transfer difference (STD) NMR and molecular modelling approach.

4.3. Results

4.3.1. Analysis of the interaction between trisaccharide (1) and heparanase

The binding of trisaccharide (1) to HPSE was analysed by means of STD NMR. The STD NMR analysis was carried out to identify the ligand moieties deeply involved in the substrate recognition. The STD NMR spectra do not display a good signal-to-noise ratio (Figure 4.4). This is due to the slow protein-ligand exchange, namely the high affinity of the ligand for the protein. The kinetics of high affinity complexes are not favourable for the STD NMR experiment.

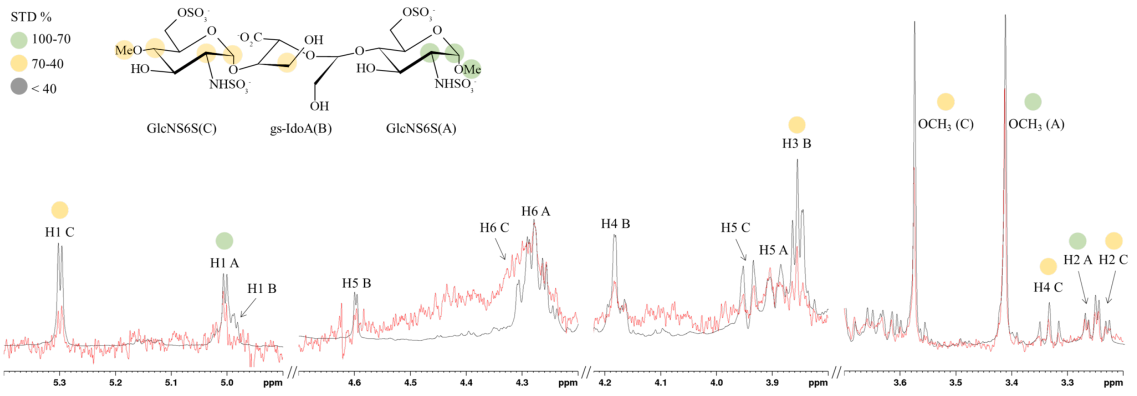

Figure 4.4. STD NMR analysis of trisaccharide (1) in interaction with HPSE. Binding epitope map of trisaccharide (1) as derived by STD NMR data. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of trisaccharide (1)-HPSE mixture with a molecular ratio of 200:1, at 295 K.

However, some information was obtained by comparing the off-resonance and STD NMR spectra. The STD NMR data revealed that the glucosamine at the reducing end [GlcNS6S(A)] is closer to the binding pocket than the glucosamine at the non-reducing end [GlcNS6S(C)] (Table 4.1). Specifically, the highest STD effects were observed for H1A (79%), H2A (100%) and OCH_3(A) (91%) protons, suggesting that GlcNS6S(A) is the sugar unit mostly involved in the binding. Appreciable STD enhancements were also detected for H1C (43%), H2C (85%), H4C (70%) and OCH_3(C) (65%) protons, unveiling that also GlcNS6S(C) participates to the interaction. Additionally, H3B belonging to the glycol-split iduronic acid unit [gs-IdoA(B)] exhibits a considerable involvement in the molecular recognition (its relative STD percentage is 63%).

Table 4.1. Absolute and relative STD percentages (Abs and Rel STD %) of selected ligand protons.

Trisaccharide (1)								
Monosaccharide	**GlcNS6S(C)**				**gs-IdoA(B)**	**GlcNS6S(A)**		
Proton	**H1C**	**H2C**	**H4C**	**OCH_3(C)**	**H3B**	**H1A**	**H2A**	**OCH_3(A)**
Abs STD %	0.39	0.77	0.64	0.59	0.57	0.72	0.91	0.83
Rel STD %	43	85	70	65	63	79	100	91

A docking study was performed to understand how trisaccharides (1) is recognized by HPSE from a 3D point of view. Selected poses were extracted considering the lowest GlideScore values (Figure 4.5).

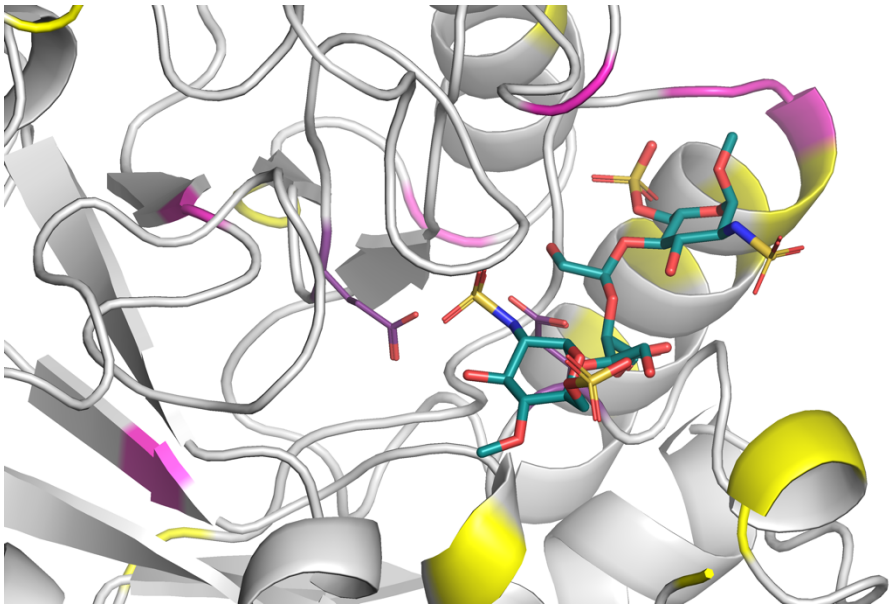

Figure 4.5. Docking pose of trisaccharide (1) in complex with HPSE fitting the STD NMR data. Trisaccharide (1) is drawn as dark green tubes; HPSE is shown as grey ribbon; the catalytic pair, consisting of E225 and E343, is represented by thin purple tubes; the arginine and lysine residues close to the catalytic site are displayed as fuchsia and yellow ribbons, respectively.

4.3.2. Analysis of the interaction between trisaccharide (2) and heparanase

The molecular recognition of trisaccharide (2) by HPSE was firstly investigated by STD NMR. Differently from the trisaccharide (1)-HPSE complex, the STD NMR spectra of trisaccharide (2) interacting with HPSE are characterized by a good signal-to-noise ratio (Figure 4.6). This is due to the most favourable dissociation constant (K_d) of the ligand, making it appropriate for being analysed by ligand-based NMR techniques, in particular STD NMR.

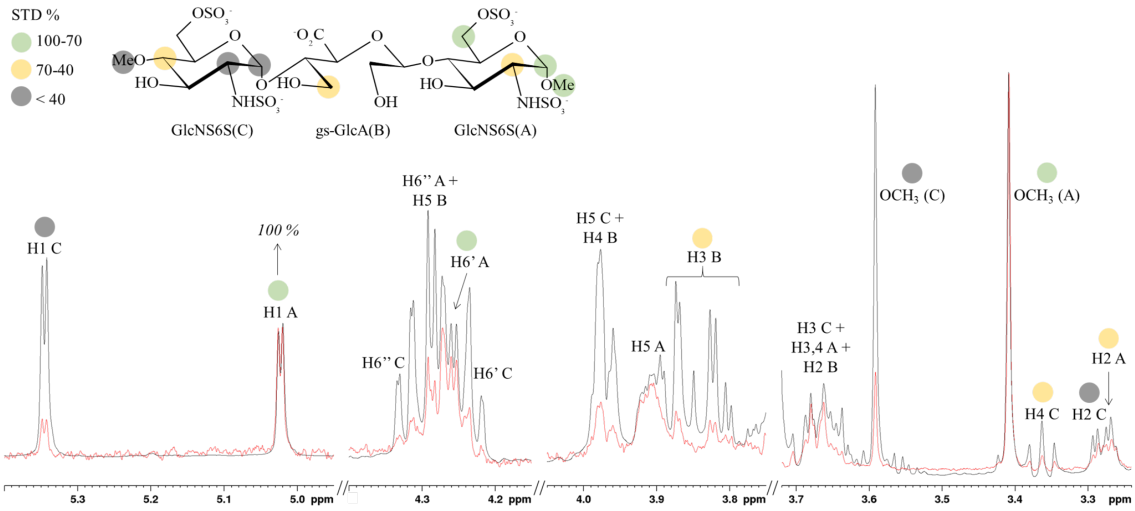

Figure 4.6. STD NMR analysis of trisaccharide (2) in interaction with HPSE. Binding epitope map of trisaccharide (1) determined from STD_0 values. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of trisaccharide (2)-HPSE mixture with a molecular ratio of 200:1, at 295 K.

The epitope map of trisaccharide (2) bound to HPSE was defined by the initial growth rate approach described in the paragraph 2.1.3.1. This method guarantees an accurate mapping that avoids possible artifacts due to re-binding phenomena and differences in the longitudinal relaxation time (T1) of the ligand protons. After measuring the STD intensities at different saturation times (Table 4.2), the STD build-up curves were fitted using Equation 2.7 (Figure 4.7) and the initial slopes (STD_0) were first calculated using Equation 2.8 and then normalized (Table 4.3). The data revealed the different involvement of the glucosamine at the non-reducing end [GlcNS6S(C)] and the glucosamine at the reducing end [GlcNS6S(A)] in the trisaccharide (2)-HPSE interaction. Also in this case, the strongest STD effects were observed for H1A (61%), H2A (43%), H6'A (100%) and OCH_3(A) (77%), revealing that GlcNS6S(A) is mainly involved in the binding. The lowest STD enhancements were detected for H1C (13%), H2C (24%), H4C (41%) and OCH_3(C) (19%), providing evidence that GlcNS6S(C) points further away from the binding pocket of HPSE than GlcNS6S(A). The magnetization transfer characterized H3B (44%) suggests a minor involvement of the glycol-split glucuronic acid unit [gs-GlcA(B)] in the substate recognition.

Table 4.2. Absolute STD percentages (Abs STD %) of selected ligand protons defined at different saturation times (0.75, 1, 1.5, 2 and 3 s).

Trisaccharide (2)									
Mono-saccharide	**GlcNS6S(C)**				**gs-GlcA(B)**	**GlcNS6S(A)**			
Proton	**H1C**	**H2C**	**H4C**	**OCH_3(C)**	**H3B**	**H1A**	**H2A**	**H6'A**	**OCH_3(A)**
Abs STD % (0.75 s)	-	0.69	0.69	0.44	0.84	1.22	0.98	1.78	1.82
Abs STD % (1 s)	0.43	0.79	0.79	0.64	1.04	2.13	1.67	2.49	2.37
Abs STD % (1.5 s)	0.56	0.9	0.9	0.69	1.05	2.94	1.81	2.54	3.15
Abs STD % (2 s)	0.72	1.39	0.91	0.89	1.14	3.51	2.15	2.7	3.53
Abs STD % (3 s)	0.85	1.57	0.93	1.02	1.32	4.05	2.83	3.0	4.22

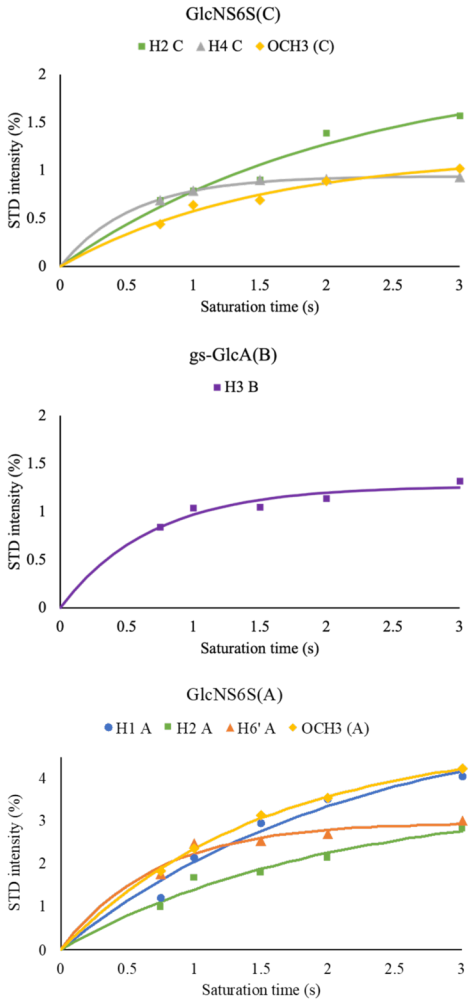

Figure 4.7. STD build-up curves show STD intensity (%) as a function of saturation time (s).

Table 4.3. STD_0 percentages (STD_0 %) and relative STD_0 percentages (Rel STD_0 %) of selected ligand protons.

Trisaccharide (2)									
Mono-saccharide	**GlcNS6S(C)**				**gs-GlcA(B)**	**GlcNS6S(A)**			
Proton	**H1C**	**H2C**	**H4C**	**OCH_3(C)**	**H3B**	**H1A**	**H2A**	**H6'A**	**OCH_3(A)**
STD_0 %	0.54	0.98	1.71	0.79	1.84	2.55	1.78	4.15	3.19
Rel STD_0 %	13	24	41	19	44	61	43	100	77

A docking study was performed to build 3D models of the trisaccharide (2)-HPSE complex. The reduced relaxation matrix approach, called RedMat, was carried out to select the docking pose that fits the experimentally determined binding epitope (Pose 1 in Table 4.4 and Figure 4.8).

Table 4.4. Comparison between simulated and experimental STD_0 factors for selected protons of tetrasaccharide (2) in complex with HPSE.

		Pose 1	*Pose 2*	*Pose 3*	*Pose 4*	*Pose 5*	*Pose 6*
Proton	**Exp STD_0**	**Sim STD_0**	**Sim STD_0**	**Sim STD_0**	**Sim STD_0**	**Sim STD_0**	**Sim STD_0**
H1C	13	35	37	29	54	54	8
H2C	24	16	64	54	91	63	0
H4C	41	14	78	71	77	68	0
OCH_3(C)	19	29	100	98	97	87	4
OCH_3(C)	19	18	87	86	86	100	0
OCH_3(C)	19	30	97	100	100	80	0
H3B	44	79	31	23	43	82	24
H3B	44	100	25	37	29	97	17
H1A	61	53	36	5	32	39	93
H2A	43	57	42	13	72	52	76
H6A	100	82	11	0	69	24	68
H6'A	100	56	18	4	67	11	54
OCH_3(A)	77	86	22	0	13	14	88
OCH_3(A)	77	73	33	0	16	19	100
OCH_3(A)	77	89	25	0	16	14	92
		R-factor	**R-factor**	**R-factor**	**R-factor**	**R-factor**	**R-factor**
		0.41	0.94	1.10	0.89	0.98	0.45

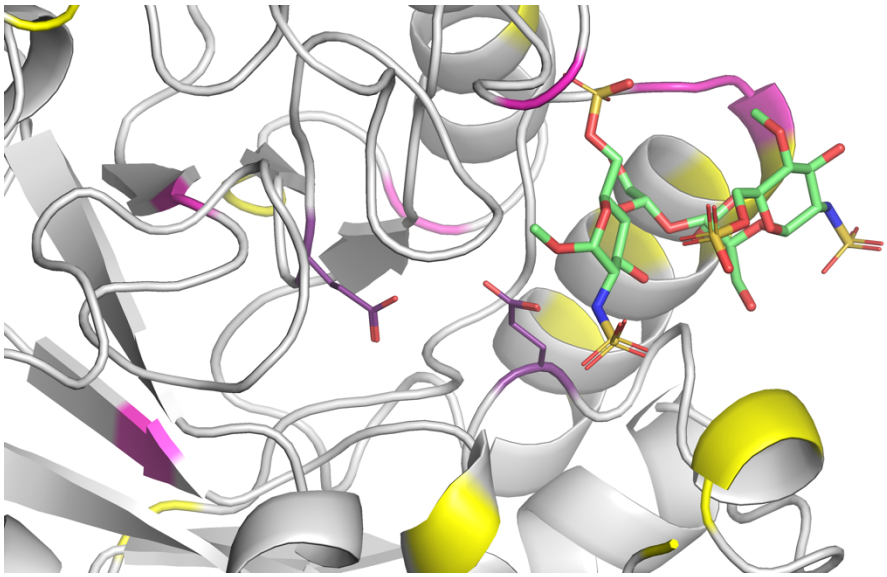

Figure 4.8. Docking pose of trisaccharide (2) in complex with HPSE fitting the STD NMR data. Trisaccharide (2) is drawn as light green tubes; HPSE is shown as grey ribbon; the catalytic pair, consisting of E225 and E343, is represented by thin purple tubes; the arginine and lysine residues close to the catalytic site are displayed as fuchsia and yellow ribbons, respectively.

Theoretical binding epitopes outlined by RedMat can be properly compared to the experimental ones if the latter are mapped using the initial growth rate approach. For the trisaccharide (1)-HPSE system, the STD percentages were measured only at a single and relatively long saturation time. Therefore, their comparison with the theoretical STD values was not feasible under these circumstances.

4.4. Conclusions

Heparanase (HPSE) plays a critical role in cleaving HS chains on cell surfaces and within the ECM. A dysregulation of its activity is associated with various pathological processes, including cancer, inflammation, cardiovascular diseases and neurological disorders. Glycol-split heparins possess structural motifs that enable their binding to HPSE, inhibiting its activity. The understanding of the molecular details characterizing the interaction between glycol-split heparins and HPSE may lead the design of optimized inhibitors with a greater specificity, higher affinity, enhanced efficacy and reduced side effects. Two glycol-split compounds, a trisaccharide (1) with a glycol-split iduronic acid and a trisaccharide (2) with a glycol-split glucuronic acid, were synthesized. The binding of trisaccharides (1) and (2) to HPSE was investigated by a combined STD NMR and molecular docking approach. The STD NMR analysis revealed that the two glycol-split trisaccharides present a distinct affinity for HPSE. Specifically, it unveiled that the binding of trisaccharide (1) is favoured over that

of trisaccharide (2). Moreover, the STD NMR results showed that both trisaccharides (1) and (2) are characterized by a comparable epitope map for their interaction with HPSE. Both mappings displayed that the sugar units composing trisaccharides (1) and (2) are differently involved in the molecular recognition, underlining that the reducing end moiety mediates the binding, while the non-reducing end and uronate residues enhance the interaction. The docking pose of trisaccharide (1) bound to HPSE was compared with that of trisaccharide (2) interacting with HPSE, revealing molecular insight into the binding mode (Figure 4.9). The inhibitory activity of the two glycol-split trisaccharides against HPSE was tested, pointing out that trisaccharide (1) is moderately active, while trisaccharide (2) is an inactive compound. This study highlighted the significance of specific structural motifs within glycol-split heparins that contribute to their affinity and activity. Further research using longer glycol-split oligosaccharides is necessary to fully characterize the precise mechanisms underlying the glycol-split heparin-HPSE interaction.

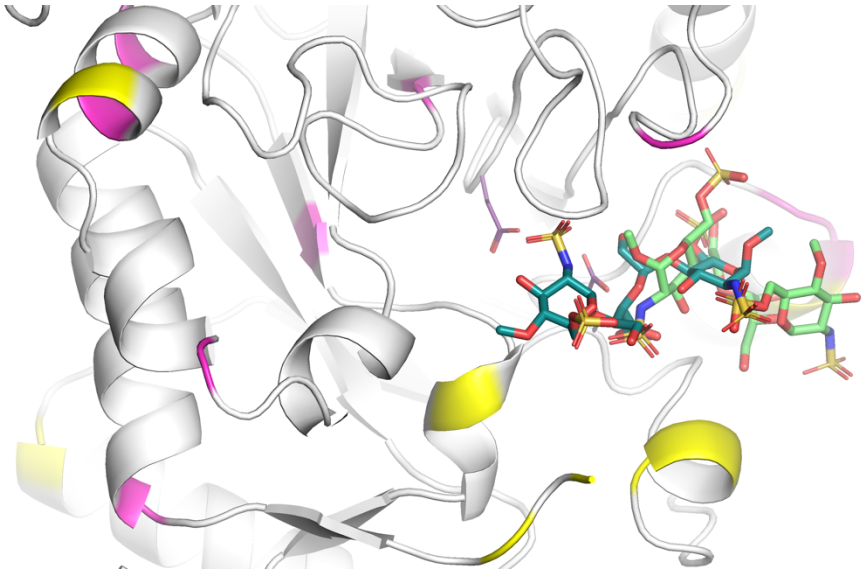

Figure 4.9. Trisaccharide (1)-HPSE and trisaccharide (2)-HPSE complexes obtained by docking calculations. Trisaccharide (1) and trisaccharide (2) are drawn as dark green and light green tubes, respectively; HPSE is shown as grey ribbon; the catalytic pair, consisting of E225 and E343, is represented by thin purple tubes; the arginine and lysine residues close to the catalytic site are displayed as fuchsia and yellow ribbon, respectively.

4.5. Materials and methods

4.5.1. Ligands and protein

Trisaccharides (1) and (2) were synthetized by Dr. Minghong Ni (Ronzoni Institute). HPSE protein was provided by Prof. Dr. Vito Ferro (University of Queensland). HPSE aliquots

(0.15 mg/mL in 25 mM Tris pH 7.0 with 400 mM NaCl) were washed with 15 mM phosphate buffer pH 5.8 with 150 mM NaCl and 0.3 mM EDTA (D_2O) using VWR® centrifugal filters (10 kDa membrane, 0.5 mL).

4.5.2. NMR experiments

NMR experiments were performed with a Bruker Avance III spectrometer equipped with a high-sensitivity 5 mm TCI cryoprobe operating at 600.13 MHz.

For the NMR characterization of trisaccharides (1) and (2), ^{1}H, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC spectra were acquired. Tables A5.1 and A5.2 report the ^{1}H and ^{13}C chemical shift assignments of the two trisaccharides; Figures A5.1 and A5.2 show their ^{1}H-^{13}C HSQC spectra.

For the STD NMR measurements, each oligosaccharide was dissolved in 0.2 mL of a purified HPSE solution, yielding a molar ratio of ligand:protein 200:1. The final concentrations were approximately 1.5 mM for the ligand and 7 μM for the protein. All spectra were recorded using the pulse sequence stddiffesgp.3 at 295 K. For both trisaccharides, the number of scans were 2560, the number of dummy scans 16, the recycle delay 6 s. For trisaccharide (1), the saturation time was set to 3 s; for trisaccharide (2), different saturation times were utilized (0.75 s, 1 s, 1.5 s, 2 s and 3 s). A 10 ms spin-lock pulse was used to remove the broad resonances of the protein. The on-resonance irradiation was performed at 580 Hz, whereas the off-resonance control irradiation at 20000 Hz. The STD NMR spectra and the STD NMR intensities were obtained as described in Chapter 3.

4.5.3. Docking calculations

Docking calculations were performed in Maestro using Glide (Schrodinger, LLC, New York, NY).[126,127] The crystal structure of HPSE (PDB ID: 5E9B) was downloaded from the Protein Data Bank. The target structure was refined using the Protein Preparation Wizard tool. The bond orders were assigned and the hydrogens were added to all atoms in the structure. The protonation state for each residue was calculated with Epik at pH 7.4. The structure was refined to optimize the hydrogen bond network using the OPLS3e force field. The minimization was terminated when the energy converged or the RMSD value reached a cut-off of 0.30 Å. The receptor grid was generated with the Receptor Grid Generation tool, setting a square box centred on the co-crystallized ligand. No constraints and no flexible side chains were included in the docking protocol. Both ligands were optimized to lower energy

conformers using Ligand Preparation tool. The standard precision (SP) docking mode was performed on the generated grid of the protein structure. The final evaluation of the protein-ligand binding was done using the GlideScore values.

4.5.4. Reduced relaxation matrix analysis of the STD initial slopes (RedMat)

The reduced relaxation matrix analysis, implemented in the form of a web application called RedMat, was employed to calculate the theoretical STD initial slopes (STD_0) from the 3D models of the trisaccharide (2)-HPSE complex. For the RedMat calculation, the following parameters were set: the NMR spectrometer frequency at 600 MHz, the complex rotational correlation time at 37.869 ns, the concentrations of ligand and protein at 2000 μM and 10 μM, respectively, and the cut-off distance at 15 Å. The agreement between the theoretical and experimental STD_0 factors was evaluated using the R-factor. The latter was calculated as reported in Equation 3.6, where $I^{exp,k}$ is the experimental STD_0 value for a proton k ($STD_0^{exp,k}$) and $I^{calc,k}$ is the STD_0 value simulated using the reduced relaxation matrix algorithm ($STD_0^{calc,k}$).

Appendix

Table A4.1. $^{1}H/^{13}C$ chemical shift assignment of trisaccharide (1).

Proton	GlcNS6S(C)	gs-IdoA(B)	GlcNS6S(A)
1	5.30/98.4	4.99/106.0	5.01/100.6
2	3.24/60.1	3.65-3.74/ 64.3	3.26/60.0
3	3.70/73.1	3.85/62.5	3.76/74.2
4	3.34/81.1	4.18/79.7	3.62/77.2
5	3.95/70.8	4.60/80.0	3.90/70.2
6	4.17-4.30/ 68.6	-	4.25-4.27/ 68.9
OCH_3	3.58/62.3	-	3.41/57.7

Table A4.2. $^{1}H/^{13}C$ chemical shift assignment of trisaccharide (2).

Proton	GlcNS6S(C)	gs-GlcA(B)	GlcNS6S(A)
1	5.35/98.2	4.88/106.6	5.03/100.6
2	3.29/60.1	3.65/64.0	3.27/60.2
3	3.71/73.3	3.82-3.89/ 61.6	3.68/73.2
4	3.36/81.1	3.98/80.3	3.67/77.8
5	3.97/71.1	4.29/81.0	3.91/70.6
6	4.23-4.33/ 68.7	-	4.26-4.28/ 68.9
OCH_3	3.59/62.3	-	3.41/57.7

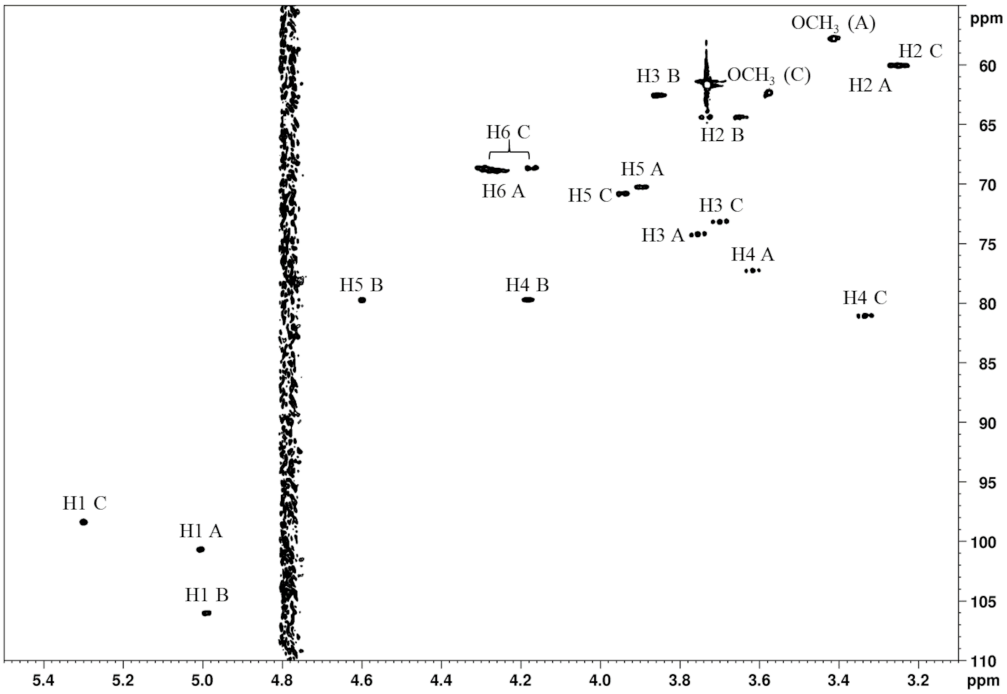

Figure A4.1. ^{1}H-^{13}C HSQC NMR spectrum of trisaccharide (1).

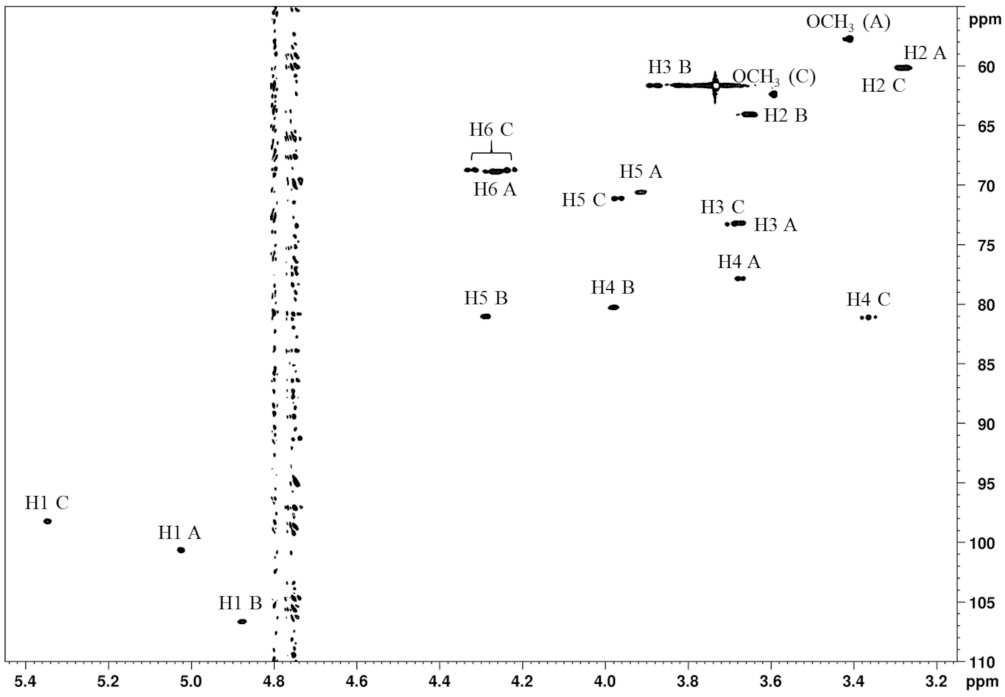

Figure A4.2. ^{1}H-^{13}C HSQC NMR spectrum of trisaccharide (2).

Chapter 5

5. Multisolvent STD NMR and MD simulations to explore the interaction between heparin oligosaccharides and AT

5.1. Introduction

As already mentioned in Chapter 1, the anticoagulant action of heparin derives from its ability to bind and activate antithrombin (AT).[128,129] This binding determines an enhancement of the inhibitory activity of AT against several proteases of the coagulation system, such as factors IIa (thrombin) and Xa. Antithrombin (AT) is a single-chain glycoprotein belonging to the serpin family. Structurally, AT consists of three β-sheets (A to C), nine α-helices (A to I) and a reactive centre loop (RCL) that connects the larger β-sheet A to the smaller β-sheet C (Figure 5.1). The heparin binding site comprises three domains, including the N-terminal region (K11 and R13), the N-terminal end of the helix A (N45, R46 and R47), the helix D (E113, K114, K125 and R129) with its N-terminal loop.[130,131] Antithrombin (AT) circulates in a latent form until it binds a specific pentasaccharide sequence [D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS3S6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4)] present in the heparin chains. The binding to heparin induces the elongation of the helix D through a conformational change that ejects its RCL from the β-sheet A (Figure 5.1). This release reorients the RCL, allowing it to capture the target protease (factors IIa and Xa) (active form).

As described in Chapter 1, the study of GAG-protein complexes is fundamental for understanding the molecular mechanisms underlying GAG-protein interactions and for designing innovative glycomimetic drugs. Ligand-based NMR experiments, including STD NMR, water-LOGSY and tr-NOESY, allow to characterize the epitope map and the bioactive conformation of the ligand. Information about the nature of the protein residues (aliphatic, aromatic or polar) involved in the binding process can be obtained by the novel multifrequency and multisolvent STD NMR approach called differential epitope mapping (DEEP) by STD NMR.[47] The possibility to detect polar patches is particularly interesting in the investigation of GAG-protein interactions, since the binding sites of the GAG-binding proteins are often lined with arginine and lysine residues.

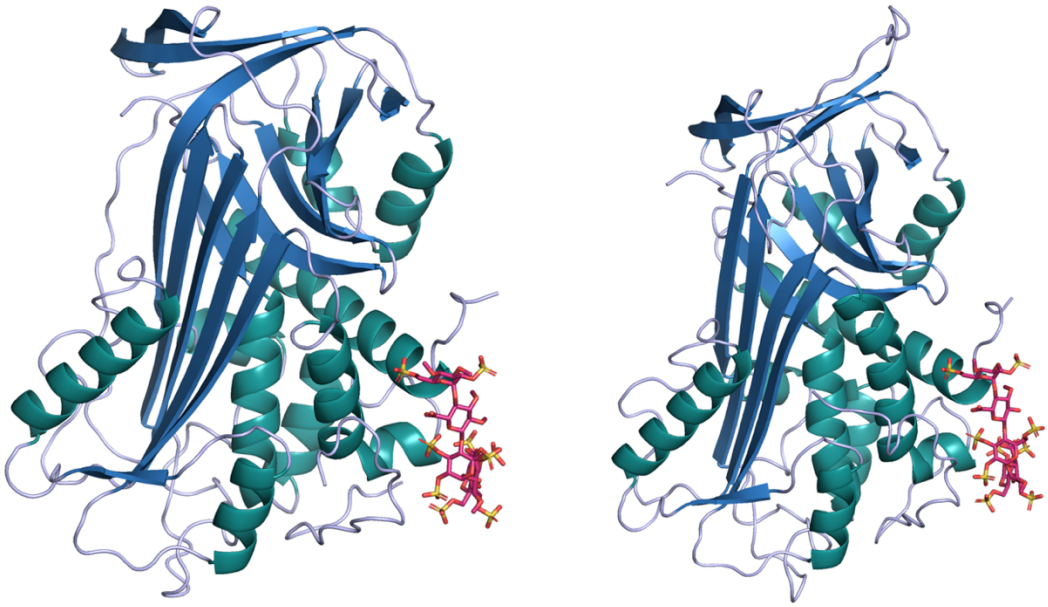

Figure 5.1. Crystal structure of AT bound to a pentasaccharide sequence in its latent (left) and active (right) forms (PDB ID: 1AZX). AT is shown as light blue, dark blue and green ribbon; pentasaccharide is displayed as fuchsia tubes.

5.2. Aim of the work

The first aim of the study was to evaluate the effectiveness of the DEEP-STD NMR approach in different solvents (D_2O and H_2O) in determining the presence of arginine residues in the binding sites of GAG-binding proteins, which are mainly constituted by positively charged amino acids, such as arginine (R) and lysine (K). Antithrombin (AT), a widely studied GAG-binding protein, was chosen as a model protein because its heparin binding site is characterized by a significant number of positively charged residues (K11, R13, R46, R47, K114, K125 and R129).[25] It is known that AT binds to heparin recognizing a specific pentasaccharide sequence present in about 30% of its chains.[26,132-135] Therefore, two synthetic heparin pentasaccharides, here namely pentasaccharide (1) and pentasaccharide (2), were selected as ligand probes. The structure of pentasaccharide (1) is D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS3S6S α (1-4) L-IdoA2S α (1-4) D-GlcNS6S α (1-4) OMe (it corresponds to the heparin sequence that is specifically recognized by AT), while that of pentasaccharide (2) is D-GlcNS6S α (1-4) D-GlcA β (1-4) D-GlcNS3S6S α (1-4) L-IdoA α (1-4) D-GlcNS6S α (1-4) OMe (Figure 5.2). The sugar sequence of these oligosaccharides is closely related; the only difference is in the sulfation in position 2 of the iduronic acid unit: pentasaccharide (1) presents a sulfate group (IdoA2S), while pentasaccharide (2) is characterized by a hydroxyl group (IdoA). Interestingly, the absence of the 2-O-sulfate group

in the iduronate moiety determines a significant reduction of the affinity of pentasaccharide (2) for AT.

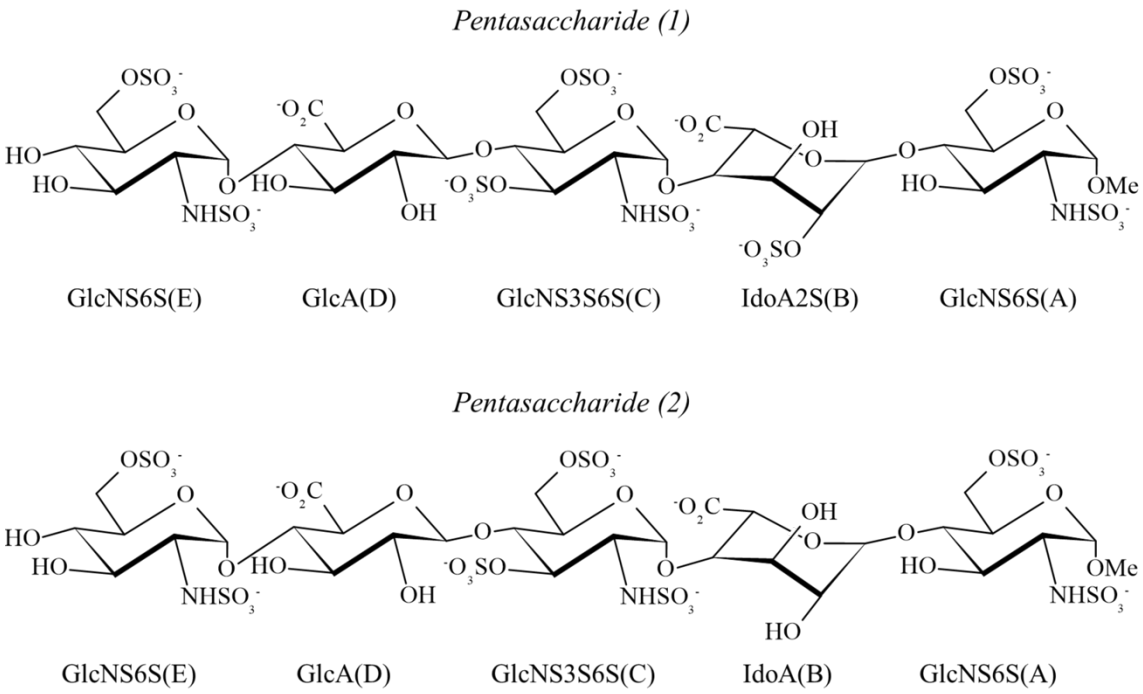

Figure 5.2. Chemical structure of the heparin pentasaccharides selected as AT binders. The sugar residues are depicted in their prevalent conformation in the unbound state (L-IdoA2S and L-IdoA residues can be either in $^{1}C_{4}$ chair conformation or $^{2}S_{0}$ skew-boat conformation, depending on their sulfation degree and that of the neighbouring glucosamines).

The binding of pentasaccharide (1) to AT is well described: the stiff non-reducing trisaccharide of pentasaccharide (1), attracted by the positively charged surface of the AT binding site, induces a protein conformational change that allows the interaction with the flexible reducing disaccharide.[25] The latter is also assisted by the conformational flexibility of the iduronic acid unit, which adopts a pure $^{2}S_{0}$ skew-boat conformation upon binding, enhancing the contacts. Figure 5.3 shows a pentasaccharide-AT complex obtained by X-ray crystallography (PDB ID: 1AZX).

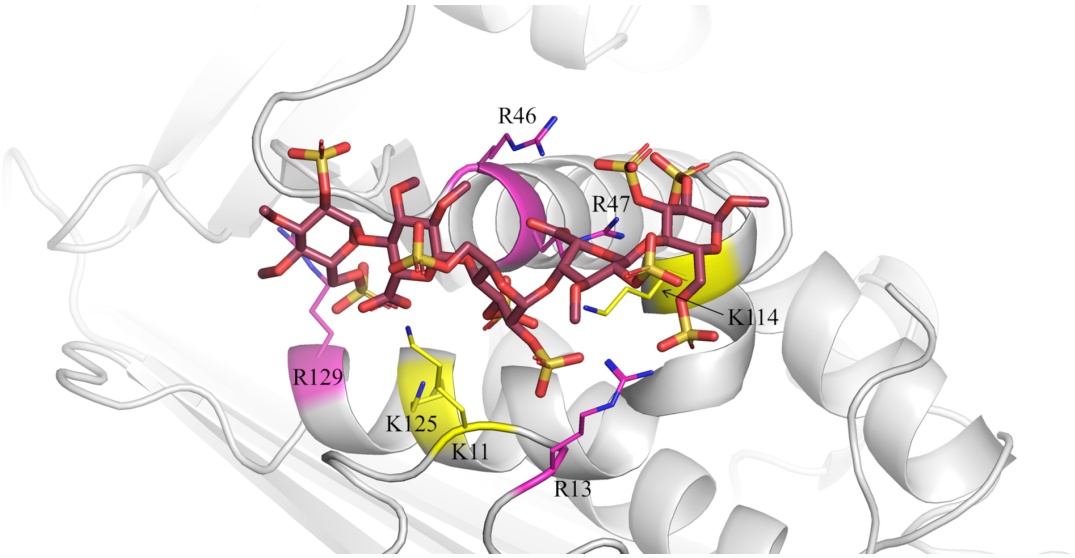

Figure 5.3. Pentasaccharide in complex with AT (PDB ID: 1AZX). Pentasaccharide is represented as purple tubes; AT is depicted as grey ribbon; selected arginine (R) and lysine (K) residues are indicated as thin fuchsia and yellow tubes, respectively.

Since the reducing end disaccharide of pentasaccharide (1) interacts with AT in an area of the binding site with the highest density of arginine residues, our assumption was to observe a relative increase in the STD intensities of the ligand protons belonging to the aforementioned moiety of the glycan in the STD NMR experiment conducted in H_2O. The same behaviour was also expected from pentasaccharide (2), whose binding mode should be similar to that of pentasaccharide (1).

Furthermore, the second objective of the work was to underline possible differences in the interaction of pentasaccharides (1) and (2) with AT by a combined use of STD NMR, DEEP-STD NMR and MD simulations. Firstly, the binding of pentasaccharides (1) and (2) to AT was investigated by STD NMR to delineate the ligand moieties involved in the molecular recognition process and differential solvent DEEP-STD NMR to define the polar nature of the protein residues characterizing the binding pocket. Additionally, NOESY and tr-NOESY experiments were run to describe the ligand conformation before and upon AT-binding. Docking and MD simulations were performed to build static and dynamic 3D models of the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes. Finally, RedMat calculations were carried out to estimate the theoretical binding epitope for each system. Their comparison with the corresponding experimental data allowed to validate the 3D models of the complexes.

5.3. Results

5.3.1. STD NMR and differential solvent DEEP-STD NMR analysis

5.3.1.1. Study of the interaction between pentasaccharide (1) and AT

For performing differential solvent DEEP-STD NMR analysis of the pentasaccharide (1)-AT complex, two samples were prepared: the first one in deuterated buffer (D_2O) and the second one in aqueous buffer (H_2O). Moreover, the sample in H_2O was prepared in different buffer condition: with and without adding deuterated EDTA. Traces of bivalent ions, such as Ca^{2+}, were present in the water used to prepare the sample, determining a line-broadening of some peaks in the proton spectrum (data not shown). The presence of the Ca^{2+} ions mainly affects the signal belonging to the anomeric proton of the iduronic acid unit [IdoA2S(B)], because its 2-O-sulfate group, together with the N-sulfate group of the neighbouring trisulfated glucosamine [GlcNS3S6S(C)], forms a coordination complex with these ions. The addition of EDTA allowed to remove these traces, improving the spectral resolution. The preparation of the sample without EDTA was necessary to observe the signal of the amine proton of GlcNS3S6S(C). This proton is characterized by a low exchange rate with the bulk water as blocked by an intramolecular hydrogen bond. An increase in this rate is observed when EDTA is added to the sample, due to the involvement of the groups participating in the above-named bond in the interaction with EDTA. The STD NMR experiments were run setting 0.5 s as saturation time. Figures A5.1 and A5.2 display the STD NMR spectra recorded for the pentasaccharide (1)-AT complex. Table 5.1 summarizes the absolute and relative STD intensities calculated as described in the experimental section. Absolute STD values of about 1% for the deuterated sample and in the range of 1-2% for the aqueous sample were measured. Unexpectedly, the total magnetization transferred from the protein to the ligand is lower in D_2O than in H_2O. After evaluating the relative STD percentages, the binding epitope map of the pentasaccharide (1)-AT complex was defined either in D_2O or in H_2O. The interacting epitope, mapped in D_2O, showed that H2C, H6''C, H2B, H2A, H3A and OCH_3(A) receive the strongest saturation (80-100%), while H1E, H2E, H5E, H1D, H1C, H1B and H1A are less affected by the magnetization transfer (60-80%). The binding epitope map, defined in H_2O, displayed that all the protons present a STD intensity in the range of 60-80%, except for H2C, which STD value is of 100%. The lowest STD intensity was measured for the amine proton of GlcNS3S6S(C) (<40%). A rough comparison was carried out to analyse the difference between the binding epitopes obtained

in D_2O and H_2O; it indicated that H2A receives less saturation when the STD NMR experiment is conducted in H_2O.

Table 5.1. Absolute and relative STD percentages (STD %), calculated as described in the experimental section, of selected ligand protons.

	Pentasaccharide (1)			
	Absolute STD %		**Relative STD %**	
Proton	**D_2O**	**H_2O**	**D_2O**	**H_2O**
H1E	1.04	1.46	70	69
H2E	1.16	1.52	78	72
H5E	1.03	1.27	69	60
H1D	1.07	1.29	72	61
H1C	1.04	1.49	70	70
H2C	1.49	2.11	100	100
H6''C	1.22	1.58	82	75
H1B	0.95	1.37	64	65
H2B	1.21	1.42	81	67
H1A	0.98	1.38	66	66
H2A	1.42	1.65	95	78
H3A	1.31	1.57	88	74
OCH_3(A)	1.22	1.63	82	77

After mapping the interacting epitopes, the DEEP-STD factors (ΔSTDs) were defined as explained in the experimental section. The raw and processed data are reported in Table 5.2. Negligible ΔSTDs were collected for the protons of pentasaccharide (1) bound to AT. Indeed, as shown in Figure 5.4, the strongest negative and positive values were around |0.1|.

Table 5.2. DEEP-STD NMR analysis using different solvents (D_2O/H_2O) of the pentasaccharide (1)-AT complex.

Pentasaccharide (1)				
Proton	**STD % D_2O**	**STD % H_2O**	**STD H_2O/D_2O Ratio**	**ΔSTD**
H1E	1.04	1.46	1.40	0.09
H2E	1.16	1.52	1.31	0.00
H5E	1.03	1.27	1.23	–0.08
H1D	1.07	1.29	1.21	–0.10
H1C	1.04	1.49	1.42	0.12
H2C	1.49	2.11	1.41	0.11
H6’’C	1.22	1.58	1.29	–0.02
H1B	0.95	1.37	1.44	0.13
H2B	1.21	1.42	1.17	–0.13
H1A	0.98	1.38	1.41	0.10
H2A	1.42	1.65	1.16	–0.15
H3A	1.31	1.57	1.20	–0.11
OCH_3(A)	1.22	1.63	1.33	0.03
	Sum	***Sum***	***Average***	
	15.16	19.73	1.31	
			St. Dev.	
			0.10	

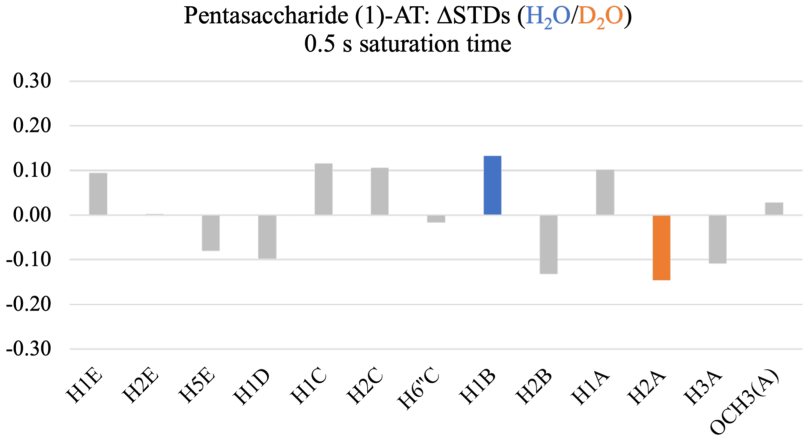

Figure 5.4. Differential epitope mapping (H_2O/D_2O) of pentasaccharide (1) in complex with AT. ΔSTD histogram: proton with the strongest positive ΔSTD is in blue; proton with the highest negative ΔSTD is in orange.

5.3.1.2. Study of the interaction between pentasaccharide (2) and AT

The analysis described in the previous section was also carried out for the pentasaccharide (2)-AT complex. Two samples, in D_2O and in H_2O, were prepared. The STD NMR spectra were recorded using a saturation time of 0.5 s (Figures A5.3 and A5.4). For both deuterated and aqueous samples, absolute STD values between 1% and 2% were measured (Table 5.3). Differently from the pentasaccharide (1)-AT system, the total saturation transferred from the protein to the ligand is lower in D_2O than in H_2O. The interacting epitope of the pentasaccharide (2)-AT complex was mapped either in D_2O or H_2O. Because of the superimposition between the signals of the anomeric protons of both IdoA(B) and GlcNS6S(A), they were not integrated, even though they are involved in the interaction. The epitope mapping in D_2O showed that the STD intensity is strong (80-100%) for H2E, H2C and H2A, medium (60-80%) for H1E, H5E, H1D, H1C and H6''C and low (40-60%) for H4B and OCH_3(A). The same mapping in H_2O displayed the higher STD enhancements for H1E, H5E, H2C and H2A and the lower STD effects for H1D, H1C, H6''C, H4B and OCH_3(A). The amine proton of GlcNS3S6S(C) presents a STD intensity in the range 40-60%. A preliminary comparison between the maps defined in D_2O and H_2O showed a difference in H1E, which acquires an extra saturation in H_2O, and H1D, which loses some magnetization in H_2O.

Table 5.3. Absolute and relative STD percentages (STD %), calculated as described in the experimental section, of selected ligand protons.

	Pentasaccharide (2)			
	Absolute STD %		**Relative STD %**	
Proton	**D_2O**	**H_2O**	**D_2O**	**H_2O**
H1E	2.01	2.10	76	90
H2E	2.14	1.59	81	68
H5E	1.93	1.59	73	68
H1D	1.89	1.27	72	54
H1C	1.72	1.35	65	58
H2C	2.44	2.15	93	92
H6''C	1.74	1.32	66	57
H4B	1.45	1.00	55	43
H2A	2.63	2.34	100	100
OCH_3(A)	1.56	1.22	59	52

To properly analyse the differences in the interacting epitopes mapped in D_2O and H_2O, the DEEP-STD NMR protocol was applied. The raw and processed data are reported in Table

5.4. As shown in Figure 5.5, a negative ΔSTD for H1E and positive ΔSTDs for H1D and H4B were obtained.

Table 5.4. DEEP-STD NMR analysis using different solvents (D_2O/H_2O) of the pentasaccharide (2)-AT complex.

Pentasaccharide (2)				
Proton	**STD % D_2O**	**STD % H_2O**	**STD D_2O/H_2O Ratio**	**ΔSTD**
H1E	2.01	2.10	0.96	–0.30
H2E	2.14	1.59	1.34	0.09
H5E	1.93	1.59	1.21	–0.04
H1D	1.89	1.27	1.48	0.23
H1C	1.72	1.35	1.28	0.02
H2C	2.44	2.15	1.14	–0.12
H6''C	1.74	1.32	1.31	0.05
H4B	1.45	1.00	1.45	0.19
H2A	2.63	2.34	1.12	–0.13
OCH_3(A)	1.56	1.22	1.28	0.02
	Sum	***Sum***	***Average***	
	19.50	15.94	1.26	
			St. Dev.	
			0.16	

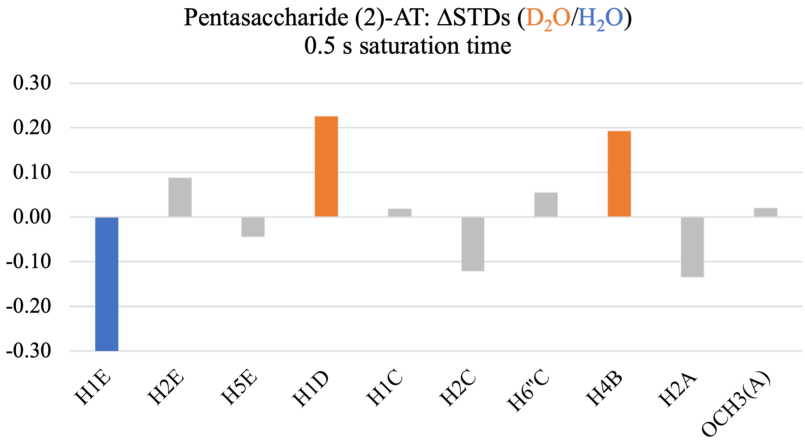

Figure 5.5. Differential epitope mapping (D_2O/ H_2O) of pentasaccharide (2) in complex with AT. ΔSTD histogram: protons with the strongest positive ΔSTDs are in orange; proton with the highest negative ΔSTD is in blue.

5.3.1.3. STD and DEEP-STD NMR data comparison of pentasaccharides (1) and (2) bound to AT

The binding epitopes of the two systems were compared (Table 5.5 and Figure 5.6). The comparison in D_2O revealed that OCH_3(A) is closer to the protein surface in the pentasaccharide (1)-AT complex. The minor involvement of this group in the binding of pentasaccharide (2) to AT is correlated to the lower affinity of the ligand for the protein. Specifically, the absence of the 2-O-sulfate group in the iduronate unit of pentasaccharide (2) determines a higher mobility of the IdoA(B)-GlcNS6S(A) disaccharide at the reducing end. Moreover, an additional difference was observed for H6''C: this group is closer to the AT surface in the pentasaccharide (1)-AT system due to a key electrostatic interaction between the 6-O-sulfate group of GlcNS3S6S(C) and R46. Due to the loss of the GlcNS3S6S(C)(6S)-R46 contact in the pentasaccharide (2)-AT complex, H6''C is farther from the binding area. The same comparison in H_2O also unveiled that H1E and H2A are closer to the binding pocket in the pentasaccharide (2)-AT system. These results are in agreement with the MD simulation data described in the following paragraph 5.3.3.2 (Figures 5.15 and A5.9).

Table 5.5. Relative STD percentages (STD %) of pentasaccharides (1) and (2) bound to AT calculated in D_2O and H_2O at 0.5 s saturation time.

	0.5 s saturation time			
	D_2O		**H_2O**	
Proton	**Pentasaccharide (1)-AT**	**Pentasaccharide (2)-AT**	**Pentasaccharide (1)-AT**	**Pentasaccharide (2)-AT**
H1E	70	76	69	90
H2E	78	81	72	68
H5E	69	73	60	68
H1D	72	72	61	54
H1C	70	65	70	58
H2C	100	93	100	92
H6''C	82	66	75	57
H2A	95	100	78	100
OCH_3(A)	82	59	77	52

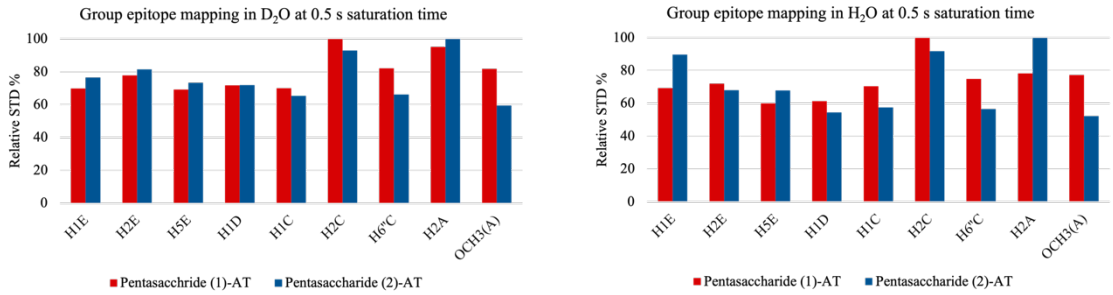

Figure 5.6. STD histograms: relative STD percentages (STD %) of pentasaccharides (1) and (2) bound to AT calculated in D_2O and H_2O at 0.5 s saturation time.

For each system, the level of saturation achieved in D_2O was compared with that acquired in H_2O. This comparison revealed that the saturation transferred to pentasaccharide (1) is lower in D_2O than H_2O (15.16% < 19.73%); on the contrary, the magnetization that receives pentasaccharide (2) is higher in D_2O than H_2O (19.50% > 15.94%). This difference could be related to the different affinity of the two pentasaccharides for AT. A high affinity determines a long residence time of the ligand in the protein pocket. Notably, a long residence time implies a greater protection of the polar protons of the protein residues interacting with the ligand from their exchange with the bulk water, enhancing the possibility for these protons to contribute to the magnetization transfer. The forenamed protection is more relevant when pentasaccharide (1) bound to AT because it is characterized by a higher affinity for the protein and a longer residence time in the binding site. This evaluation allowed to underline the difference in the binding affinity of the two ligands.

To compare equally the DEEP-STD NMR data collected from the analysed systems, only the STD NMR signals in common between the two complexes were selected. Interestingly, the two complexes did not show the same pattern in the DEEP-STD NMR data (Table 5.6 and Figure 5.7). The differential epitope map of the pentasaccharide (1)-AT complex showed a negative ΔSTD for H1C and a positive ΔSTD for H2A (Table 5.6 and Figure 5.7). Nevertheless, the DEEP-STD factors of this system are quite small and probably not significant. On the contrary, the differential epitope map of the pentasaccharide (2)-AT complex displayed a stronger negative ΔSTD for H1E and a stronger positive ΔSTD for H1D (Table 5.6 and Figure 5.7). The relative decrease of the STD effect characterizing H1E in D_2O suggests its proximity to slow exchanging polar residues (*i.e.,* arginine); the relative increase of the STD enhancement belonging to H1D in D_2O indicates the presence of fast exchanging polar residues (*i.e.,* lysine) close to non-exchangeable residues interacting with the ligand.

Table 5.6. DEEP-STD NMR analysis using different solvents (D_2O/H_2O) of the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

	Pentasaccharide (1)-AT				*Pentasaccharide (2)-AT*			
Proton	**STD % D_2O**	**STD % H_2O**	**STD D_2O/H_2O Ratio**	**ΔSTD**	**STD % D_2O**	**STD % H_2O**	**STD D_2O/H_2O Ratio**	**ΔSTD**
H1E	1.04	1.46	0.71	−0.05	2.01	2.10	0.96	−0.28
H2E	1.16	1.52	0.76	0.00	2.14	1.59	1.34	0.11
H5E	1.03	1.27	0.82	0.05	1.93	1.59	1.21	−0.02
H1D	1.07	1.29	0.83	0.06	1.89	1.27	1.48	0.25
H1C	1.04	1.49	0.70	−0.07	1.72	1.35	1.28	0.04
H2C	1.49	2.11	0.71	−0.06	2.44	2.15	1.14	−0.10
H6''C	1.22	1.58	0.78	0.01	1.74	1.32	1.31	0.08
H2A	1.42	1.65	0.86	0.09	2.63	2.34	1.12	−0.11
OCH_3(A)	1.22	1.63	0.75	−0.02	1.56	1.22	1.28	0.04
	Sum	**Sum**	**Average**		**Sum**	**Sum**	**Average**	
	10.71	13.99	0.77		18.05	14.94	1.24	
			St. Dev.				**St. Dev.**	
			0.06				0.15	

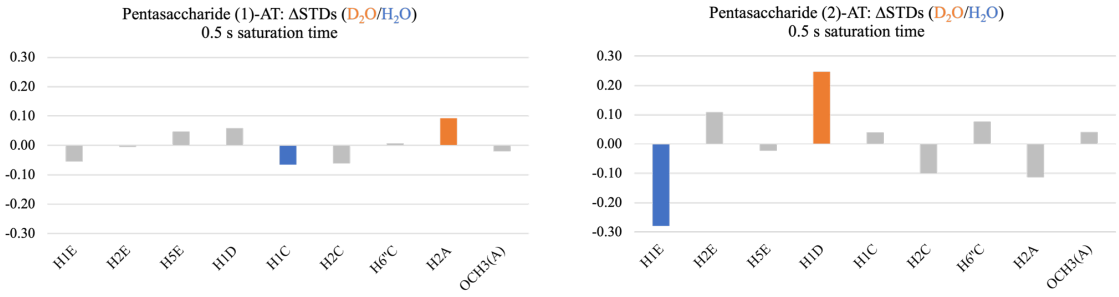

Figure 5.7. Differential epitope mapping (D_2O/ H_2O) of both pentasaccharides (1) and (2) in complex with AT. ΔSTD histogram: protons with the strongest positive ΔSTDs are in orange; protons with the highest negative ΔSTDs are in blue.

5.3.2. NOE and tr-NOE data analysis

5.3.2.1. Conformational analysis of pentasaccharides (1) and (2) in the unbound and AT-bound states

Three-bond proton-proton coupling constants ($^3J_{H-H}$) of pentasaccharides (1) and (2) in the free state were measured to define the conformation of the sugar rings (Tables 5.7 and 5.8). The analysis of the $^3J_{H-H}$ values showed that the glucosamines [GlcNS6S(E, C, A)] and the glucuronic acid [GlcA(D)] in both pentasaccharides adopt a $^4C_1$ chair conformation.[84] The

evaluation of the $^3J_{H-H}$ magnitudes allowed to establish the conformer populations of IdoA2S(B) in pentasaccharide (1). The iduronate moiety exists in equilibrium between the $^1C_4$ chair (36%) and $^2S_0$ skew-boat (64%) conformations.[13] Unfortunately, the overlapping of the signals belonging to IdoA(B) did not permit to perform the same analysis in pentasaccharide (2).

Table 5.7. $^3J_{H-H}$ coupling constants (Hz) of the sugar units of pentasaccharide (1) in unbound state.

$^3J_{H-H}$	GlcNS6S(E)	GlcA(D)	GlcNS3S6S(C)	IdoA2S(B)	GlcNS6S(A)
$^3J_{1-2}$	3.7	7.9	3.5	3.7	3.6
$^3J_{2-3}$	10.0	9.9	10.7	7.3	10.5
$^3J_{3-4}$	-	-	-	-	-
$^3J_{4-5}$	-	-	-	3.1	-

Table 5.8. $^3J_{H-H}$ coupling constants (Hz) of the sugar units of pentasaccharide (2) in unbound state.

$^3J_{H-H}$	GlcNS6S(E)	GlcA(D)	GlcNS3S6S(C)	IdoA(B)	GlcNS6S(A)
$^3J_{1-2}$	3.7	7.9	3.5	-	3.5
$^3J_{2-3}$	9.9	9.3	10.6	-	10.6
$^3J_{3-4}$	-	-	-	-	-
$^3J_{4-5}$	-	-	-	2.5	-

NOESY and tr-NOESY experiments were carried out to describe the intra-residue and inter-glycosidic conformational changes upon binding (Figures A5.5 and A5.6). The NOE results confirmed the conformational data obtained from the $^3J_{H-H}$ analysis for both pentasaccharides in the free state. The evaluation of the NOE values also allowed the conformation of IdoA(B) in pentasaccharide (2) to be defined. A significant variation in the intensity of the cross-peaks from the NOESY spectra acquired for the free ligands to the tr-NOESY spectra recorded in presence of the protein was observed, providing proof of the binding (Tables 5.9 and 5.10).

Table 5.9. H1-H2 NOEs and tr-NOEs of GlcNS6S (E, C, A) of pentasaccharide (1) in unbound and AT-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage.

	GlcNS6S(E)		*GlcNS3S6S(C)*		*GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2
0.15	4.1	9.4	4.4	9.6	4.4	8.4
0.3	8.7	20.6	8.6	19.2	6.8	18.1
0.5	14.3	35.2	14.0	31.3	12.1	31.6

Table 5.10. H1-H2 NOEs and tr-NOEs of GlcNS6S(E, C, A) of pentasaccharide (2) in unbound and AT-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage.

	GlcNS6S(E)		*GlcNS3S6S(C)*		*GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2	H1-H2
0.15	4.1	9.4	5.3	8.8	4.2	8.4
0.3	8.7	20.2	8.0	17.8	7.3	18.2
0.5	14.3	34.9	13.1	27.4	11.6	33.1

The conformation of the iduronate residue in pentasaccharides (1) and (2) before and upon AT-binding was investigated by analysing the NOESY and tr-NOESY spectra. As extensively described in Chapter 3.1, the ratio between the H5-H2 and H5-H4 NOEs provides the conformer populations of the iduronic acid unit. Unfortunately, the partial overlapping between the H5-H3 and H5-H4 cross-peaks in the NOESY and tr-NOESY spectra of pentasaccharide (1) and pentasaccharide (1)-AT complex, respectively, did not allow to confirm the conformer populations of IdoA2S in the unbound and bound states. Because of the H5-H3/H5-H4 superimposition, the obtained ratios between the H5-H2 and H5-H4 NOEs/tr-NOEs are underestimated. Nevertheless, it is well described in literature that the iduronic acid unit belonging to the pentasaccharide (1) structure exists in equilibrium between the $^1C_4$ chair and $^2S_0$ skew-boat conformations and approaches a pure $^2S_0$ form upon AT-binding.[136] The interaction between pentasaccharide (2) and AT changes the H5-H2/H5-H4 NOE ratio of IdoA(B) from 0.3 to 0.7, confirming that the conformation of IdoA, which adopts an almost pure $^1C_4$ chair conformation, is shifted toward the $^2S_0$ skew-boat form in the AT-bound state (Table 5.11).

Table 5.11. H5-H2 and H5-H4 NOEs and tr-NOEs of IdoA(B) of pentasaccharide (2) in unbound and AT-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage. The ratio H5-H2/H5-H4 is shown in brackets.

	IdoA(B)	
T_{mix}	**NOE**	**tr-NOE**
	H5-H2/ H5-H4	H5-H2/ H5-H4
0.15	0.8/3.2 (0.3)	5.1/7.1 (0.7)
0.3	1.7/6.5 (0.3)	10.5/15.4 (0.7)
0.5	3.2/11.0 (0.3)	16.9/25.9 (0.7)

The inter-glycosidic NOEs and tr-NOEs were defined for both glycans in the unbound and bound states, respectively. 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 in Tables 5.12 and 5.13. Their comparison underlined that the GlcA(D)-GlcNS3S6S(C) linkage in both pentasaccharides

undergoes to a conformational change upon AT-binding. Additionally, ROESY and tr-ROESY spectra were recorded and no differences between the ROE/tr-ROE and NOE/tr-NOE results were detected (data not reported).

Table 5.12. H1-H6 and H1-H4 inter-glycosidic NOEs and tr-NOEs characterizing the backbone conformation of pentasaccharide (1) in unbound and AT-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage. The ratios H1-H6/H1-H4 are shown in brackets.

	GlcNS6S(E)-GlcA(D)		*GlcA(D)-GlcNS3S6S(C)*		*IdoA2S(B)-GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H4	H1-H4	H1-H6''/ H1-H4	H1-H6''/ H1-H4	H1-H6/ H1-H4	H1-H6/ H1-H4
0.15	3.0	8.2	4.6/2.9 (1.6)	9.4/8.5 (1.1)	2.3/3.1 (0.7)	5.1/10.2 (0.5)
0.3	6.3	18.2	7.5/5.5 (1.4)	12.6/19.4 (0.6)	4.3/6.3 (0.7)	9.6/18.2 (0.5)
0.5	10.8	30.9	9.9/8.6 (1.2)	14.8/28.4 (0.5)	6.5/10.2 (0.6)	15.5/27.8 (0.6)

Table 5.13. H1-H6 (or H1-H3) and H1-H4 inter-glycosidic NOEs and tr-NOEs characterizing the backbone conformation of pentasaccharide (2) in unbound and AT-bound states. The mixing time (T_{mix}) is reported in seconds. The NOE and tr-NOE intensities are in percentage. The ratios H1-H6/H1-H4 and H1-H3/H1-H4 are shown in brackets.

	GlcNS6S(E)-GlcA(D)		*GlcA(D)-GlcNS3S6S(C)*		*GlcNS3S6S(C)-IdoA(B)*		*IdoA(B)-GlcNS6S(A)*	
T_{mix}	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**	**NOE**	**tr-NOE**
	H1-H4	H1-H4	H1-H6''/ H1-H4	H1-H6''/ H1-H4	H1-H3/ H1-H4	H1-H3/ H1-H4	H1-H6/ H1-H4	H1-H6/ H1-H4
0.15	3.1	8.6	4.1/2.8 (1.4)	9.1/8.1 (1.1)	5.1/2.4 (2.2)	10.8/5.2 (2.1)	1.8/5.2 (0.3)	4.5/13.1 (0.3)
0.3	6.4	18.7	7.3/5.4 (1.3)	11.8/18.4 (0.6)	10.1/4.9 (2.1)	20.7/10.5 (2.0)	3.6/10.5 (0.3)	11.5/27.5 (0.4)
0.5	10.9	33.0	8.9/8.1 (1.1)	14.5/25.8 (0.6)	16.5/8.2 (2.0)	32.7/17.7 (1.8)	6.1/17.8 (0.3)	20.8/45.4 (0.5)

5.3.3. In silico studies

5.3.3.1. Characterizing the interaction of pentasaccharides (1) and (2) with AT by molecular docking

The orientation of the two ligands on the protein surface was confirmed by performing docking calculations. The evaluation of the obtained poses was carried out considering the binding energy (GlideScore value) and the distance from the crystal structure of the pentasaccharide-AT complex (PDB ID: 1AZX) [root-mean-square deviation (RMSD) value]. For the pentasaccharide (1)-AT complex, few poses were identified. In particular, the best pose, characterized by GlideScore of –8.996 and RMSD of 1.987 Å, is quite superimposable to that obtained from X-ray crystallography (PDB ID: 1AZX). For the pentasaccharide (2)-AT complex, the best pose is defined by GlideScore of –9.470 and

RMSD of 2.984 Å. Differently from the other system, several poses were collected, including many characterized by shifts in the glucosamine at the reducing end [GlcNS6S(A)]. In general, the docking results suggested what is already known about the binding mode of the heparin oligosaccharides to AT: the pentasaccharide sequence, here represented by the structures (1) and (2), specifically binds AT in the active site formed by the helix D of AT. Moreover, they underlined a difference in the binding affinity between the two pentasaccharides. The few poses obtained for pentasaccharide (1) imply a tight binding of this ligand to the protein, while the higher number of structures collected for pentasaccharide (2) indicates a greater flexibility of this compound in the binding pocket.

5.3.3.2. Investigating the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes by MD simulations

Superimposable geometries for the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes, where the ligands fit the active site of the protein as reported in the crystal structure of the pentasaccharide-AT complex (PDB ID: 1AZX), were submitted to MD simulations. Figures 5.8 and 5.9 report the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes selected from the production stage of the MD simulation trajectories.

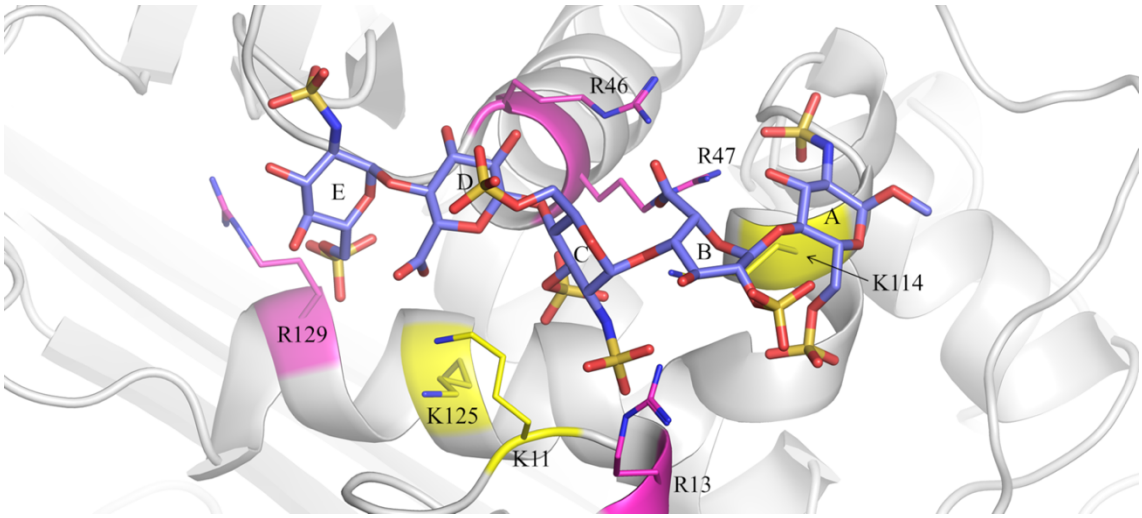

Figure 5.8. Pentasaccharide (1)-AT complex selected from the MD simulation trajectory (600 ns simulation time). Pentasaccharide (1) is represented by purple tubes; AT is reported as grey ribbon; selected arginine (R) and lysine (K) residues are indicated by thin fuchsia and yellow tubes, respectively.

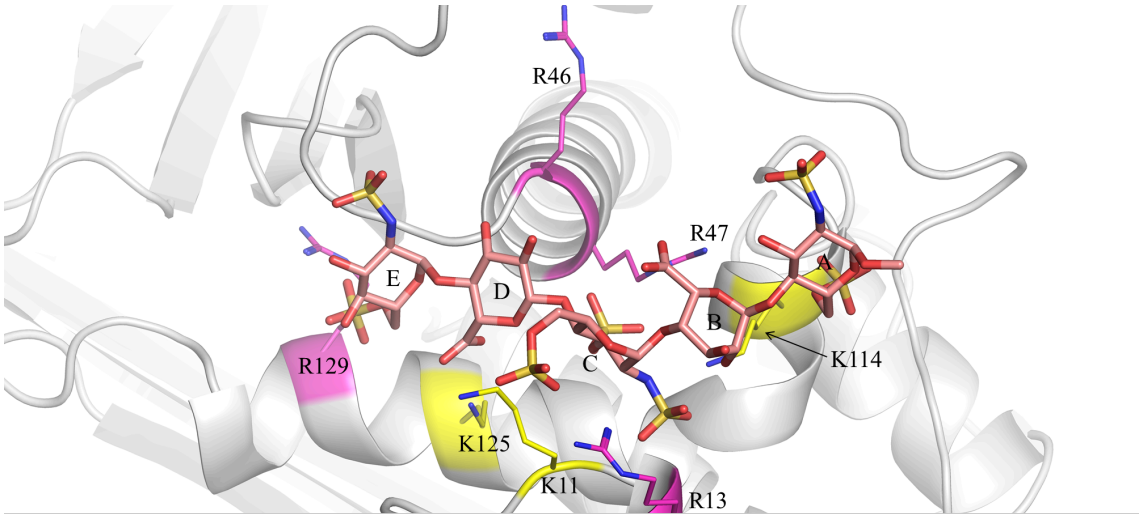

Figure 5.9. Pentasaccharide (2)-AT complex selected from the MD simulation trajectory (600 ns simulation time). Pentasaccharide (2) is represented by pink tubes; AT is reported as grey ribbon; selected arginine (R) and lysine (K) are indicated by thin fuchsia and yellow tubes, respectively.

The analysis of the RMSD distance showed that pentasaccharide (1) remains stable in the binding pocket of AT for the entire simulation time. It also displayed that pentasaccharide (2) in the AT-bound state presents a higher dynamic behaviour that is correlated to its lower affinity for the protein (Figure 5.10).

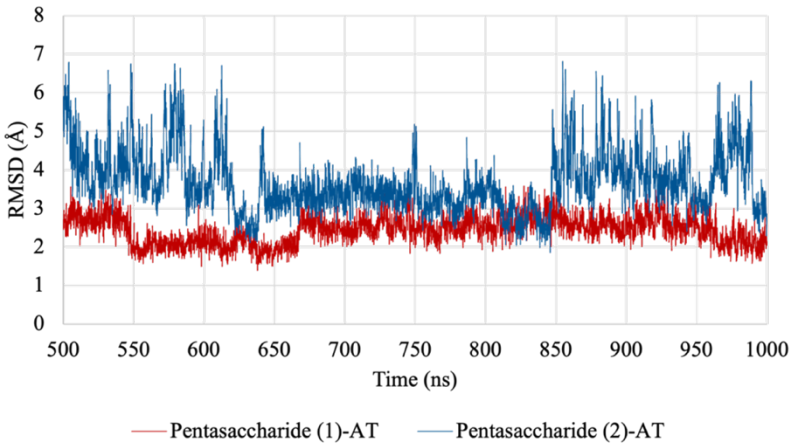

Figure 5.10. The red and blue lines show the RMSD distance calculated for pentasaccharide (1) and (2) in AT-bound state during the last 500 ns MD simulations. All the distances are expressed in Angstrom (Å).

The iduronic acid unit in the $^{1}C_{4}$ chair and $^{2}S_{0}$ skew-boat conformations exhibits a distinct H5B-H2B distance (4.0 Å and 2.4 Å, respectively). The distance between the H5 and H2 protons in the IdoA2S(B) and IdoA(B) residues was monitored during MD simulations to get information about the conformation of the iduronate moiety in the bound state (Figure 5.11). For both IdoA2S(B) and IdoA(B), the distribution of the H5B-H2B distance revealed

two distinct states corresponding to the $^{1}C_4$ and $^{2}S_0$ forms. Their relative population percentage was estimated, indicating that IdoA2S(B) is in equilibrium between the $^{1}C_4$ (50%) and $^{2}S_0$ (50%) conformations and IdoA(B) exists in the $^{1}C_4$ (90%) and $^{2}S_0$ (10%) forms. These findings are partially in alignment with literature data and our experimental results.

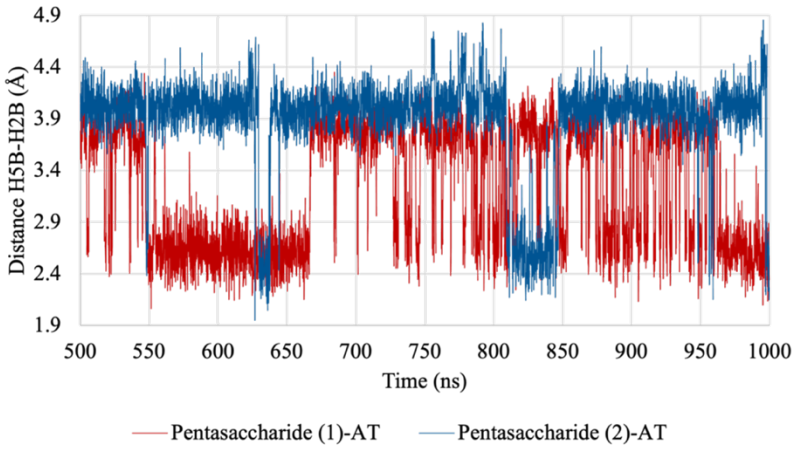

Figure 5.11. Plot of the H5B-H2B distance in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes during the last 500 ns MD simulations.

The glycosidic conformation of pentasaccharides (1) and (2) bound to AT was described by Ramachandran plots (Figure 5.12). The conformational analysis revealed that the glycosidic linkages in pentasaccharide (1) are characterized by one conformation for each dihedral angle, while the rotation around the glycosidic bonds in the pentasaccharide (2) structure is quite relevant (Figure 5.13). Specifically, it unveiled that more than one conformational state is allowed for the glycosidic linkages between the GlcNS3S6S(C) and IdoA(B) residues and between the IdoA(B) and GlcNS6S(A) residues in pentasaccharide (2) upon AT-binding, suggesting the higher flexibility at the reducing end moiety of this glycan in the AT-binding site (Figures 5.14 and 5.15). The most populated states for each dihedral angle of the two systems and the relative population percentage of the dihedral angle between IdoA(B) and GlcNS6S(A) are reported in Table 5.14.

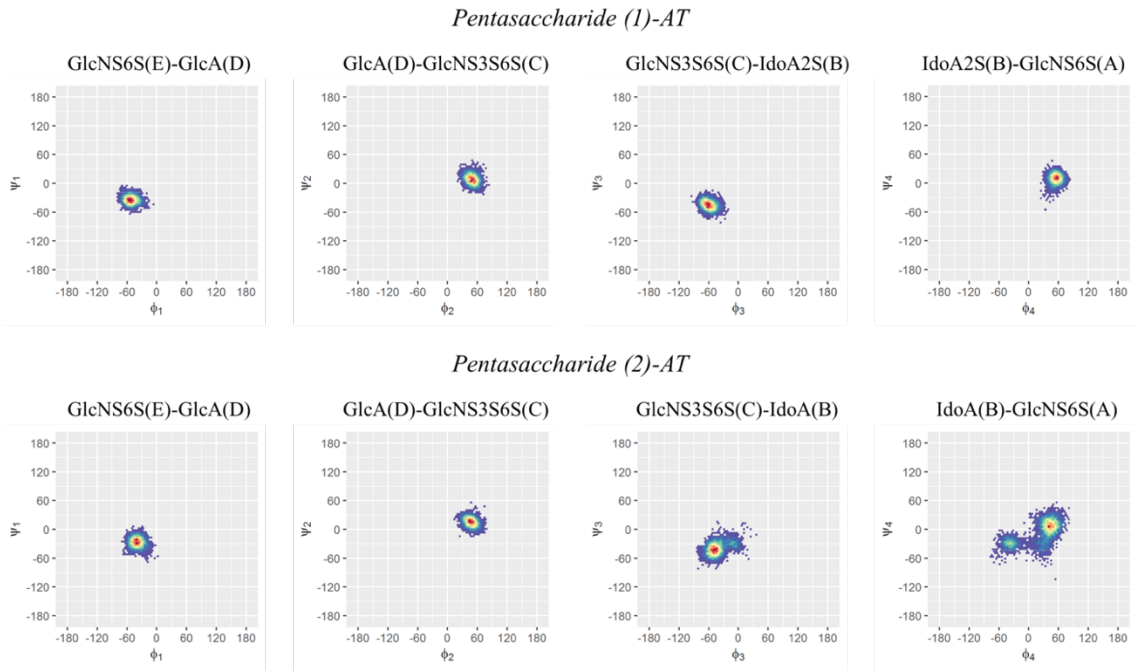

Figure 5.12. Ramachandran plots of the glycosidic dihedral angles φ_i/ψ_i of pentasaccharides (1) and (2) bound to AT. The glycosidic dihedral states are sampled by MD simulation in the production stage. A density colour map is superposed on each Ramachandran plot; the colour gradient (blue to red) is proportional to the density of the sampled φ_i/ψ_i states and qualitatively predicts the preferred conformation of each glycosidic linkage.

Table 5.14. The most populated states of each glycosidic dihedral angle φ_i/ψ_i of pentasaccharides (1) and (2) bound to AT. The relative population percentage (%) for the inter-glycosidic linkages characterized by more than one conformational state is reported in brackets.

	φ_1/ψ_1	**φ_2/ψ_2**	**φ_3/ψ_3**	**φ_4/ψ_4**
Pentasaccharide (1)-AT	–54°/–32°	50°/9°	–61°/–46°	59°/12°
Pentasaccharide (2)-AT	–41°/–29°	43°/17°	–45°/–42°	44°/6° (78%); –37°/–28° (22%)

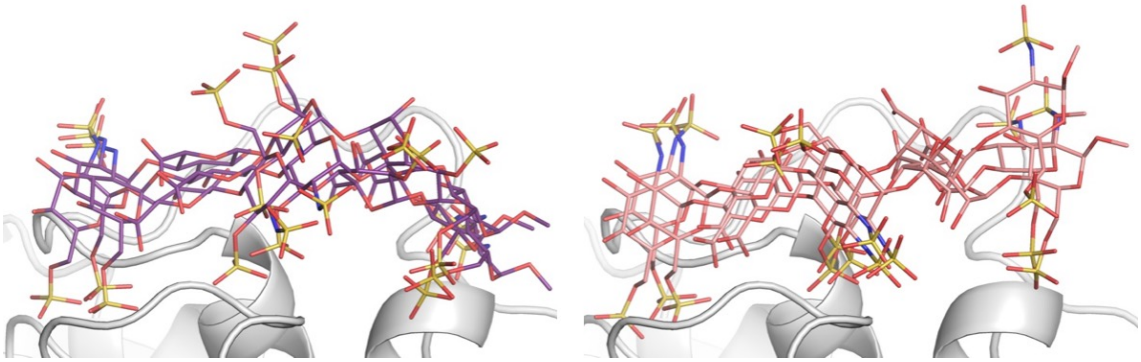

Figure 5.13. Pentasaccharide (1) and pentasaccharide (2) interacting with AT. For each pentasaccharide, three structures were selected from the corresponding MD simulation. These snapshots were sampled considering the most populated states of each glycosidic dihedral angle φ_i/ψ_i. AT is shown as gray ribbon, pentasaccharide (1) structures are displayed as purple tubes, pentasaccharide (2) structures are drawn as pink tubes.

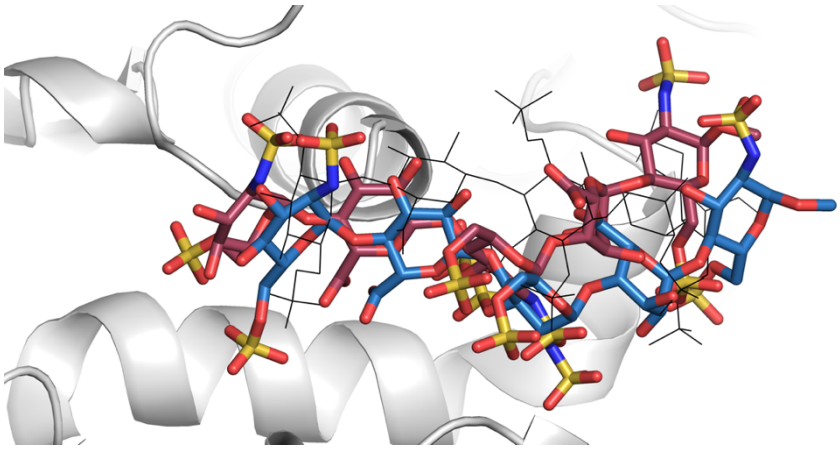

Figure 5.14. Pentasaccharide (1) and pentasaccharide (2) bound to AT. Pentasaccharide (1) presents a single geometry for each dihedral angle φ_i/ψ_i; while pentasaccharide (2) exhibits two conformational states for the dihedral angle φ_4/ψ_4. AT is shown as gray ribbon, pentasaccharide (1) is displayed as black lines, pentasaccharide (2) with the dihedral angle φ_4/ψ_4 of 44°/6° is drawn as fuchsia tubes, pentasaccharide (2) with the dihedral angle φ_4/ψ_4 of –37°/–28° is reported as blue tubes.

The interaction between pentasaccharide (1) and AT is characterized by manifold contacts between the negatively charged groups of the oligosaccharide, such as sulfate and carboxylic groups, and the positively charged side chains of the arginine (R) and lysine (K) residues that characterize the AT binding site. The main electrostatic interactions are the following: [GlcNS6S(E)(6S)]-K125, [GlcNS6S(E)(6S)]-R129, [GlcA(D)(COO)]-K125, [GlcNS3S6S(C)(NS)]-R13, [GlcNS3S6S(C)(NS,3S)]-K114, [GlcNS3S6S(C)(6S)]-R46, [IdoA2S(B)(2S)]-R13, [IdoA2S(B)(COO)]-R46, R47 and K114, [GlcNS6S(A)(NS)]-R46 and R47, [GlcNS6S(A)(6S)]-R13 and K114. Monitoring these interactions in the pentasaccharide (2)-AT complex showed that pentasaccharide (2) loses some contacts with key amino acids of the AT binding site (Figures A5.7-A5.11). In particular, the interactions [GlcNS6S(E)(6S)]-K125, [GlcNS3S6S(C)(6S)]-R46, [IdoA2S(B)(COO)]-R46, R47 and K114, [GlcNS6S(A)(NS)]-R46 and R47 and [GlcNS6S(A)(6S)]-R13 are lost in the pentasaccharide (2)-AT complex. The electrostatic interaction between the carboxylic group of GlcA(D) and K125 becomes particularly relevant in the binding of pentasaccharide (2) to AT. Indeed, the average distance for [GlcA(D)(COO)]-K125 is 7.3 Å and 4.4 Å in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes, respectively. Moreover, the contact between the N-sulfate group of GlcNS3S6S(C) and R13 is characterized by a higher stability in the pentasaccharide (2)-AT system.

The influence of the water molecules on the binding of pentasaccharides (1) and (2) to AT was evaluated using the radial distribution function [g(r)]. The application of this algorithm allowed to determine the number density of water molecules as a function of a distance (r)

from a reference atoms (or set of atoms), including the protein and sugar units. The solvation analysis of the protein residues belonging to the heparin binding site showed that the hydration networks around R13, K114 and R129 and characterizing K11, R47 and K125 determine a significant shielding effect in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes, respectively; it also displayed that the cluster of water molecules surrounding R46 is similar in both complexes (Figure A5.12). The same evaluation, carried out for the sugar units, revealed that GlcNS3S6S(C) and IdoA2S(B)/IdoA(B) are more solvated in the pentasaccharide (2)-AT complex; it also unveiled comparable shells of water molecules around the other glycan moieties (Figure A5.13). To explain the difference in the DEEP-STD NMR data collected for the analysed systems, this kind of analysis was also performed for the ligand protons of which a significant ΔSTD factor was identified (Figure A5.14). As previously mentioned, a negative ΔSTD was obtained for H1E only in the pentasaccharide (2)-AT complex (Table 5.6 and Figure 5.7). A difference in the solvation of this group may explain our experimental result: H1E loses its solvation shell upon binding of pentasaccharide (2) to AT (Figure A5.14). Therefore, it may receive an extra saturation transfer from the nearby arginine residue R129 when the STD NMR experiment is conducted in H_2O (Figures 5.9 and A5.7). A significant positive ΔSTD was obtained for H1D only in the pentasaccharide (2)-AT complex (Table 5.6 and Figure 5.7). This is probably correlated to the solvation shell of the nearby lysine residue K125 that is more relevant in the pentasaccharide (2)-AT complex (Figures A5.8 and A5.12). The exchange of this residue with the bulk water may determine a leakage of magnetization in H1D when the STD NMR experiment is performed in H_2O.

5.3.3.3. Model validation by RedMat

The methodology based on the reduced relaxation matrix (RedMat) was carried out for validating the 3D models built by MD simulations. The theoretical binding epitope was estimated using the MD trajectories of the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes (Tables A5.1-A5.4). Figures A5.15 and A5.16 show the evolution of the R-factor over the last 500 ns of MD simulation. An average R-factor of 0.26 (with a standard deviation of 0.03) for the pentasaccharide (1)-AT system and of 0.29 (with a standard deviation of 0.03) for the pentasaccharide (2)-AT system revealed a good agreement between the theoretical and experimental binding epitopes. This analysis involved an approximation. The experimental binding epitope was determined not through the initial growth rate

approach, but only employing a single saturation time. Since the latter was set close to zero, the experimental STD values were treated as equivalent to the STD intensities at the saturation time of 0 (STD_0).

5.4. Conclusions

This study suggests the possibility to apply differential solvent DEEP-STD NMR approach for identifying slow exchanging polar residues (*i.e.*, arginine) in GAG-binding proteins and for orienting GAG-based structures when the 3D structure of the receptor is known. It revealed that the number of residues with slow exchanging polar protons and their distribution in the binding pocket are critical insights to be considered. Moreover, the collected data indicated that it is fundamental to study how the hydration network affects the molecular recognition process, including the solvation of both ligand and protein residues that are involved in the binding and the presence of structural water molecules. A preliminary investigation by MD simulations can help in evaluating if the features of both ligand and protein are adequate to get key information from a potential differential solvent DEEP-STD NMR analysis.

Ligand-based NMR techniques and computational methods were applied to analyse the interaction between two heparin pentasaccharides and AT, underlining that the selected ligands, despite their remarkable structural similarity, exhibit their own binding features. The binding of pentasaccharides (1) and (2) to AT was investigated by STD NMR experiments conducted in D_2O and H_2O. In the pentasaccharide (1)-AT complex, the largest level of saturation is observed in H_2O; while in the pentasaccharide (2)-AT complex, it is achieved in D_2O. This difference suggested a higher affinity of pentasaccharide (1) for AT. The STD binding epitopes of pentasaccharides (1) and (2) were compared, revealing only minor differences. The differential solvent DEEP-STD NMR protocol was applied to identify the polar residues (*i.e.,* arginine) that characterize the heparin-binding site on the AT surface. Interestingly, the analysed complexes did not show the same DEEP-STD NMR data. The slightly different orientation of pentasaccharides (1) and (2) within the AT binding site determines a distinct distribution of the water molecules in proximity of certain sugar and protein residues. This is supposed to correlate with the observed differences in the DEEP-STD NMR results of the two complexes. The 3D models of the complexes were built by MD simulations (Figure 5.15). The study of the conformational properties revealed that the iduronate moiety exists in equilibrium between the $^1C_4$ and $^2S_0$ forms in the

pentasaccharide (1)-AT complex and predominantly adopts the $^{1}C_4$ conformation in the pentasaccharide (2)-AT complex. Moreover, it showed that the glycosidic linkages in pentasaccharide (1) are characterized by a single φ_i/ψ_i geometry, while the reducing end in pentasaccharide (2) presents a higher conformational flexibility. The evaluation of the electrostatic interactions in the pentasaccharide (1)-AT complex indicated that some key contacts are lost in the pentasaccharide (2)-AT complex. Finally, the 3D models of the two complexes were validated by comparing simulated and experimental STD NMR data. The findings of this work were turned out to be in agreement with the published results regarding the study of the interaction between two heparin hexasaccharides, differing in the sulfation degree of their iduronate unit, and AT.[136]

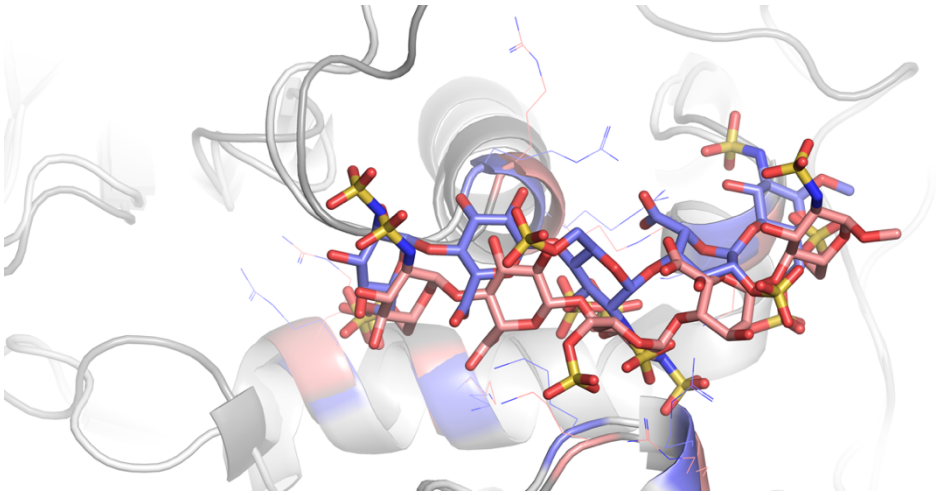

Figure 5.15. Pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes selected from the corresponding MD simulation trajectories (600 ns simulation time). Pentasaccharide (1) and pentasaccharide (2) are represented by purple and pink tubes, respectively. AT is reported as grey ribbon in both complexes; selected arginine and lysine residues are indicated by thin purple and pink tubes in pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes, respectively.

5.5. Materials and methods

5.5.1. Ligands and protein

Antithrombin (AT) protein (Kybernin P) was purchased from CSL Behring Italia. The product, containing AT and some excipients (amino acetic acid, sodium citrate, sodium chloride and hydrochloric acid/sodium hydroxide), was purified by affinity chromatography using a HiTrap™ Heparin HP Column (Cytiva). After equilibrating the column with 10 column volumes (CV) of binding buffer (10 mM phosphate buffer pH 7.4 with 150 mM NaCl), the sample was applied using a syringe fitted to the luer connector. Subsequently to

the column washing with 5 to 10 CV of binding buffer, the protein was eluted with 5 to 10 CV of elution buffer (10 mM phosphate buffer pH 7.4 with 3 M NaCl). The purified product was loaded onto a Amicon® Ultra-4 Centrifugal Filter (10 kDa membrane, 4 mL) for desalting and buffer exchanging. The final concentration was determined using Thermo Scientific™ NanoDrop™ Microvolume UV-Vis Spectrophotometer. The purity was assessed by the matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometer.

5.5.2. NMR experiments

NMR experiments were performed using a 600 MHz NMR spectrometer featuring a cryoprobe.

For characterizing pentasaccharides (1) and (2), ^{1}H, ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY and ^{1}H-^{13}C HSQC spectra were acquired. Tables A5.5 and A5.6 report the ^{1}H and ^{13}C chemical shift assignments of the two pentasaccharides; Figures A5.17 and A5.18 show their ^{1}H-^{13}C HSQC spectra.

For acquiring STD NMR spectra, the samples were prepared dissolving both ligand and protein in 10 mM phosphate buffer pH 7.4 with 150 mM NaCl (D_2O or H_2O). Additionally, 0.3 mM deuterated-EDTA was added in the H_2O buffer. The final concentrations were 2 mM for the ligand and 50 μM for the protein. The STD NMR spectra were recorded using the pulse sequences stddiff.3 and stddiffesgp.3 for the sample in D_2O and H_2O, respectively. They were acquired at 298 K by setting 1024 and 2048 scans for the pentasaccharide (1)-AT complex and the pentasaccharide (2)-AT complex, respectively, 4 dummy scans, 1.5 s recycle delay and 0.5 s saturation time. A 10 ms spin-lock pulse was used to remove the broad resonances of the protein. On-resonance and off-resonance frequencies were set at 480 Hz and 24000 Hz, respectively. The STD NMR spectrum and the STD NMR intensities were obtained as described in the experimental section of Chapter 3.1.

The DEEP-STD NMR protocol relies on running two STD NMR experiments under two different conditions, such as two irradiation frequencies or two solvents, and quantifying the differences between the two STD NMR data sets.[40] The result provides a differential epitope map, which is determined by calculating the DEEP-STD factor for each proton (ΔSTD_i):

$$\Delta STD_i = \frac{STD_{exp1,i}}{STD_{exp2,i}} - \frac{1}{n}\sum_{i}^{n}\left(\frac{STD_{exp1,i}}{STD_{exp2,i}}\right) \qquad \text{(Equation 5.1)}$$

where $STD_{exp1,i}$ and $STD_{exp2,i}$ are the STD intensities for each proton *i* in the experiment 1 (exp1) and 2 (exp2), respectively. Firstly, the experiment with the largest sum of STD values is defined as exp1 to obtain a consistent scale of ΔSTD_i; then, the ratio of the STD intensities in exp1 and exp2 for each proton and the average of these ratios are calculated; finally, the ratio average is subtracted from each of the individual STD ratios to remove the contribution from the intrinsic differences in the protein saturation level. Indeed, the STD ratios report not only the differences in the epitopes but also reflect the different level of the protein saturation achieved in the experiments. If the condition change affects the residues lining the binding pocket, we will observe positive values for protons experiencing a relative increase in STD values in exp1 and negative values for protons experiencing a relative increase in STD values in exp2. The increase/decrease of STD values is relative because it derives from a comparison of binding epitopes and not from a comparison of absolute STD values. The DEEP-STD factors to consider significant depend on the size of the STD factors for the protein-ligand system under study and can be determined from the standard deviation of the data. In this study, only the strongest STD factors were considered significant.

For the NOESY experiment, 0.7 mg oligosaccharide was solubilized in 0.2 mL of 10 mM phosphate buffer pH 7.4 with 150 mM NaCl (D_2O); for the tr-NOESY experiment, the sample was prepared by dissolving 0.7 mg oligosaccharide and 3.9 mg AT in 0.2 mL of the aforementioned buffer, reaching a molar ratio of ligand:protein 6:1. The final concentrations were about 2 mM for the ligand and 350 μM for the protein. All 2D NOESY and tr-NOESY experiments were performed at 285 K. The lower temperature was used to avoid the overlap of the H_2O signal with H5 IdoA2S(B)/IdoA(B) signals and, thus, to measure the NOE and tr-NOE values for the aforementioned protons. A total of 16 scans were collected for each free induction decay (2048×256 points) and the data were zero-filled to 2048×1024 points before the Fourier transformation. The NOESY/tr-NOESY spectra were acquired at three different mixing times (0.15 s, 0.3 s and 0.5 s). The NOE and tr-NOE values are expressed as a percentage of the mean value of the diagonal peaks of H1 GlcNS6S(E), H1 GlcNS3S6S(C), H1 IdoA2S(B)/IdoA(B) and H1 GlcNS6S(A).

5.5.3. Docking calculations

Docking calculations were performed in Maestro using Glide (Schrodinger, LLC, New York, NY) as described in the paragraph 4.5.3. The crystal structure of AT (PDB ID: 1AZX)

was downloaded from the Protein Data Bank. The final evaluation of the protein-ligand binding was done using the GlideScore and RMSD values.

5.5.4. MD simulations

The 3D models of the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes were built starting from the X-ray pentasaccharide-AT structure deposited in the Protein Data Bank (PDB ID: 1AZXY). The topology and coordinates files were created through the AmberTools package. Both systems were parametrised using the Amber (ff14SB) and Glycam06 forcefields for the protein and the ligand, respectively. They were solvated with the TIP3P water model to obtain a orthogonal bounding box of 15 Å and no counterions were added. The equilibration protocol and production dynamics were performed with the sander and pmemd modules of the Amber package, respectively. After the minimization of the solvent, the entire system was minimized. Before equilibrating at a constant pressure (1 atm), each system was heated to a temperature of 300 K at a constant volume. After the cell density equilibration, the MD simulations were run for 1 μs. The first 0.5 μs were considered as equilibration stage, while the last 0.5 μs were defined as production stage. The equilibration stage was established following the behaviour of the inter-glycosidic dihedral angles and the RMSD distance of the ligand in comparison to its initial position (time 0). In all cases, periodic boundary conditions and the particle mesh Ewald method were applied. A Langevin thermostat with a collision frequency of 3 ps^{-1} and a Berendsen barostat with a relaxation time of 2 ps were used. The SHAKE algorithm was employed to restrain all bonds involving hydrogen, allowing a timestep of 2 fs. A cut-off of 10 Å was used for all non-bonded interactions.

The profile of the local concentration of the water molecules surrounding selected atoms (or set of atoms) of protein residues, ligand units or protons was investigated using the radial distribution function [g(r)] as implemented in VMD 1.9.3.[137,138]

$$g(r) = \lim_{dr \to 0} \left[\frac{V}{4\pi N_{pair}} \cdot \frac{p(r)}{r^2 dr} \right] \qquad \text{(Equation 5.2)}$$

In Equation 5.2, the variable *r* represents the distance between the reference atom (or set of atoms) and the oxygen atom of each water molecule, *g(r) dr* reflects the concentration of water molecules in a spherical layer of infinitesimal thickness (between *r* and *r* + *dr*) centred

on the reference atom, V is the volume of the simulation cell, N_{pair} denotes the number of possible atom pairs.

5.5.5. RedMat calculations

The RedMat algorithm was used to simulate the STD values from the MD trajectories of both pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes. For the RedMat calculation, the following parameters were set: the NMR spectrometer frequency at 600 MHz, the complex rotational correlation time at 34.956 ns, the concentrations of ligand and protein at 2000 μM and 50 μM, respectively, and the cut-off distance at 10 Å.

Appendix

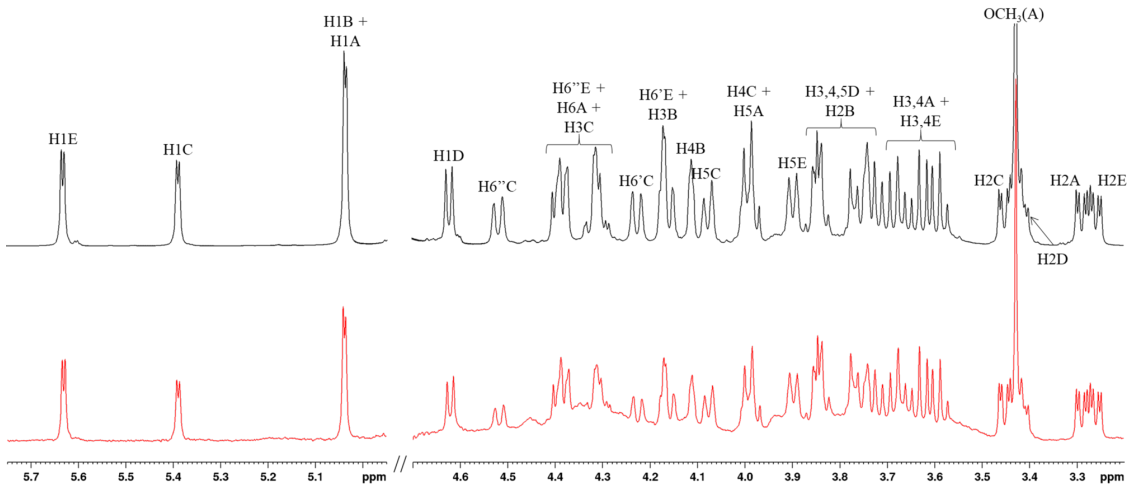

Figure A5.1. STD NMR analysis of pentasaccharide (1) in interaction with AT. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of pentasaccharide (1)-AT mixture with a molecular ratio of 50:1 in D_2O, at 298 K.

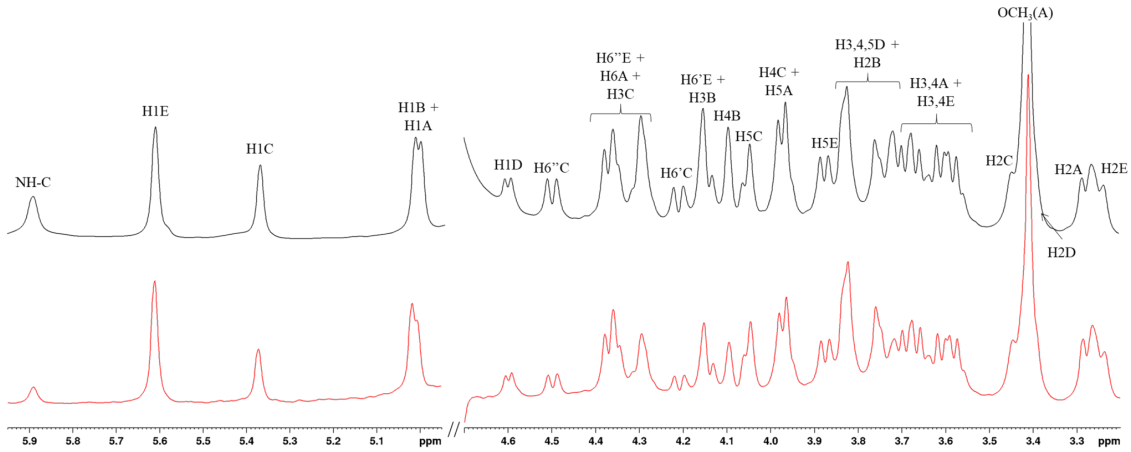

Figure A5.2. STD NMR analysis of pentasaccharide (1) in interaction with AT. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of pentasaccharide (1)-AT mixture with a molecular ratio of 50:1 in H_2O, at 298 K.

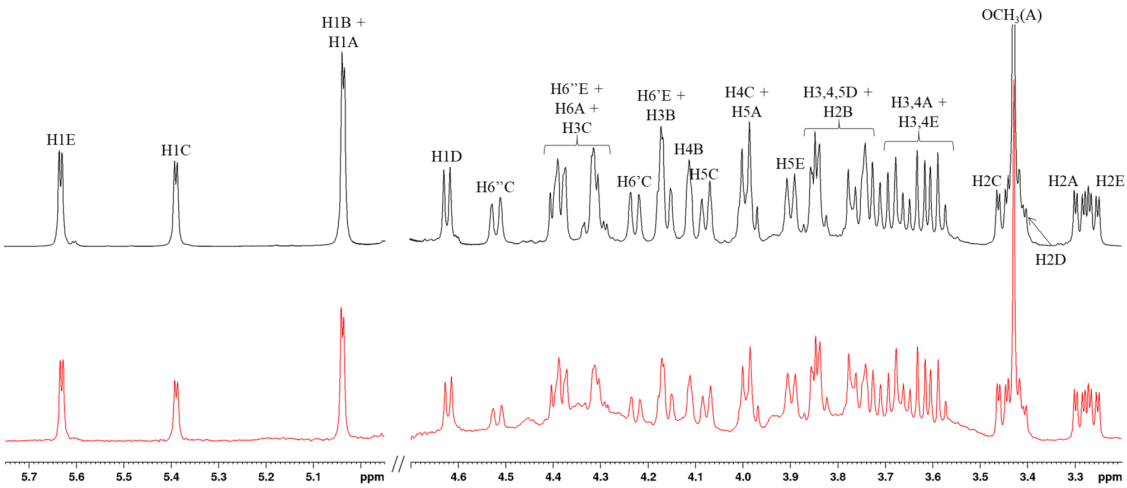

Figure A5.3. STD NMR analysis of pentasaccharide (2) in interaction with AT. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of pentasaccharide (2)-AT mixture with a molecular ratio of 50:1 in D_2O, at 298 K.

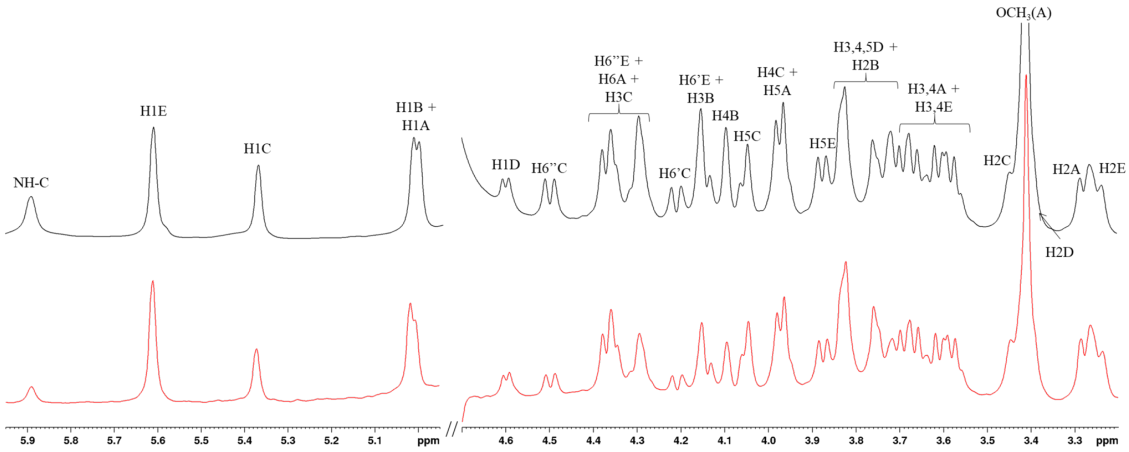

Figure A5.4. STD NMR analysis of pentasaccharide (2) in interaction with AT. Superimposition of the STD NMR spectrum (red line) and the reference spectrum (black line) of pentasaccharide (2)-AT mixture with a molecular ratio of 50:1 in H_2O, at 298 K.

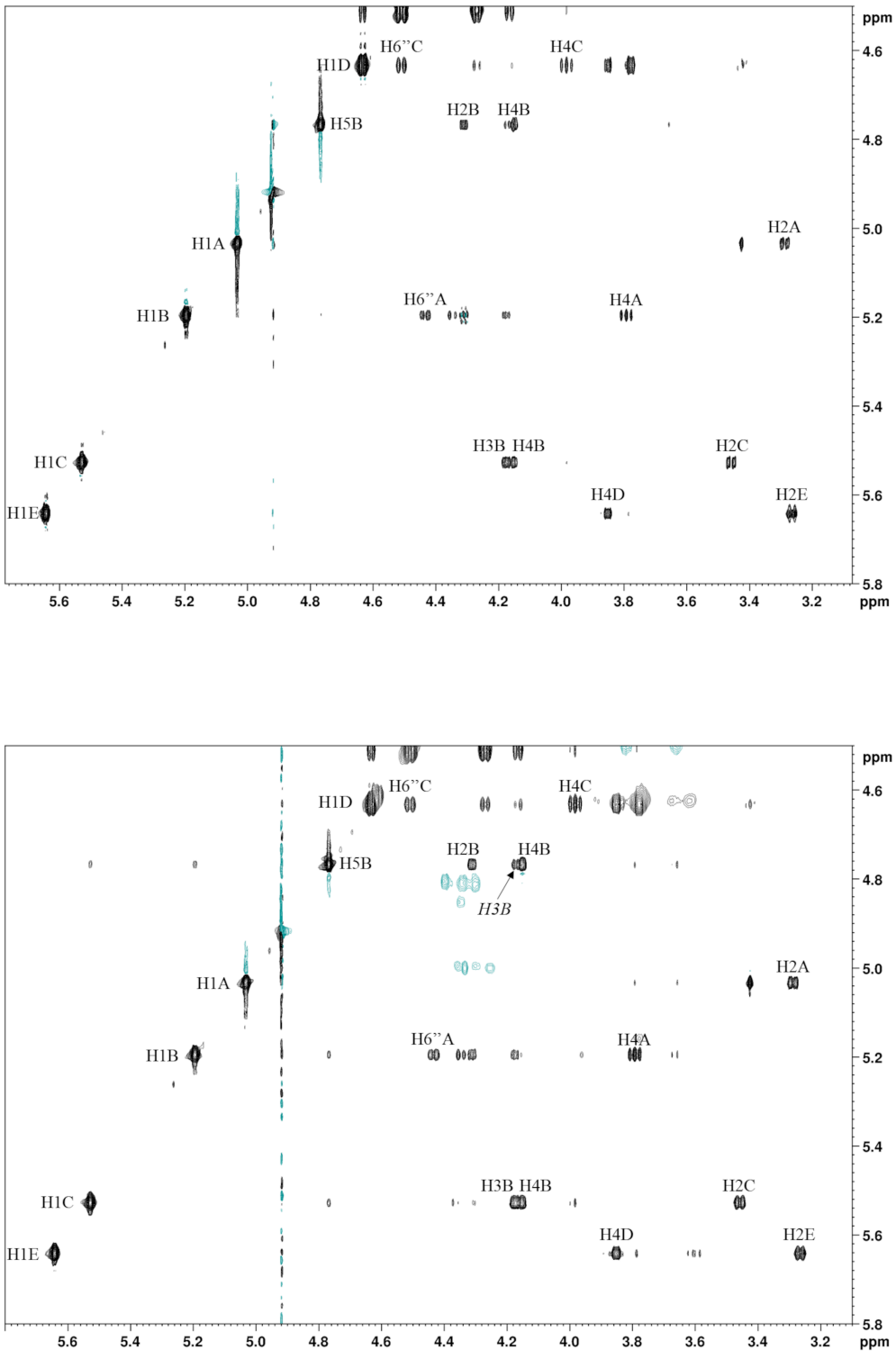

Figure A5.5. NOESY spectrum of pentasaccharide (1) (top panel). Tr-NOESY spectrum of pentasaccharide (1) interacting with AT (bottom panel).

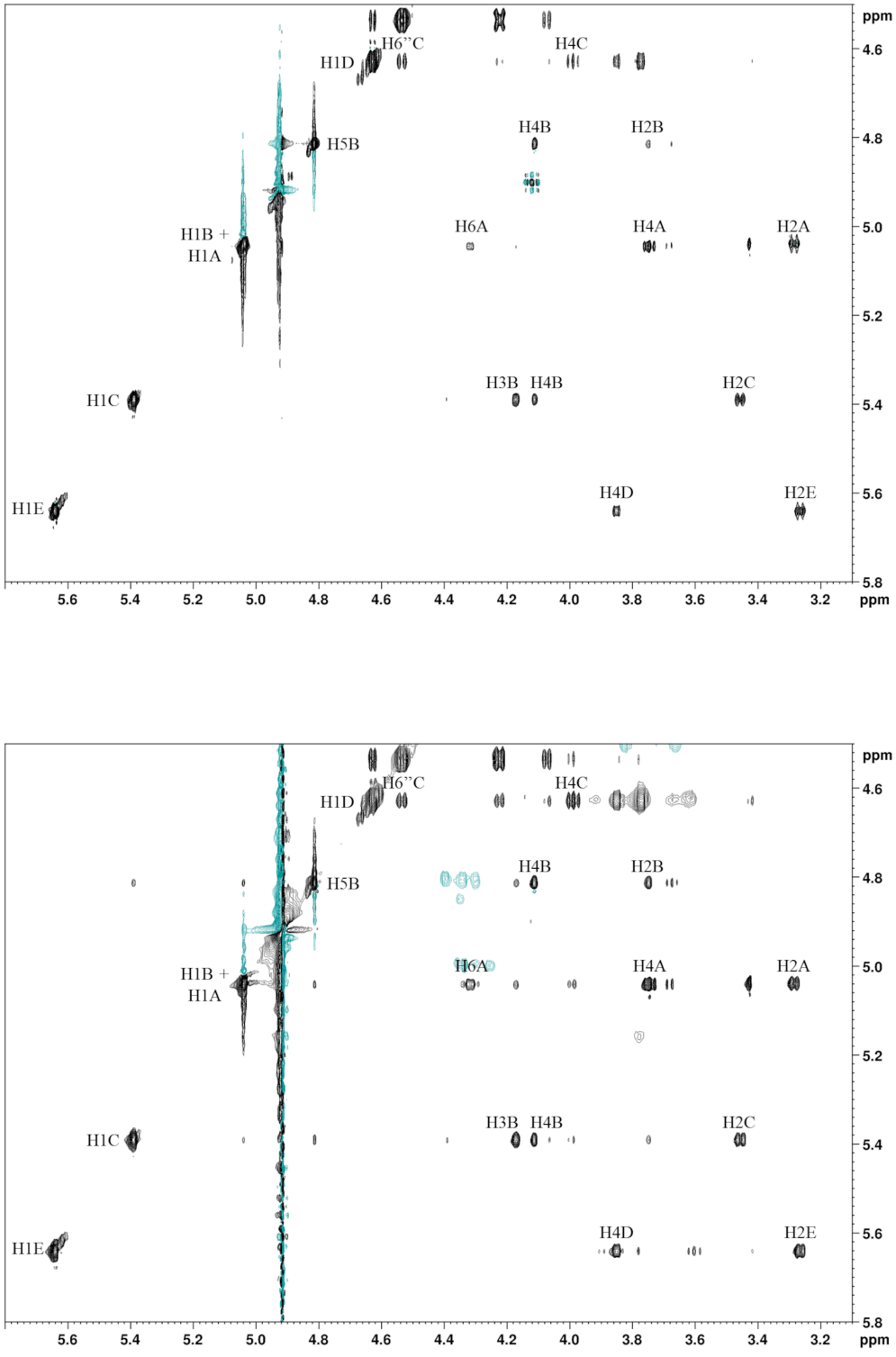

Figure A5.6. NOESY spectrum of pentasaccharide (2) (top panel). Tr-NOESY spectrum of pentasaccharide (2) interacting with AT (bottom panel).

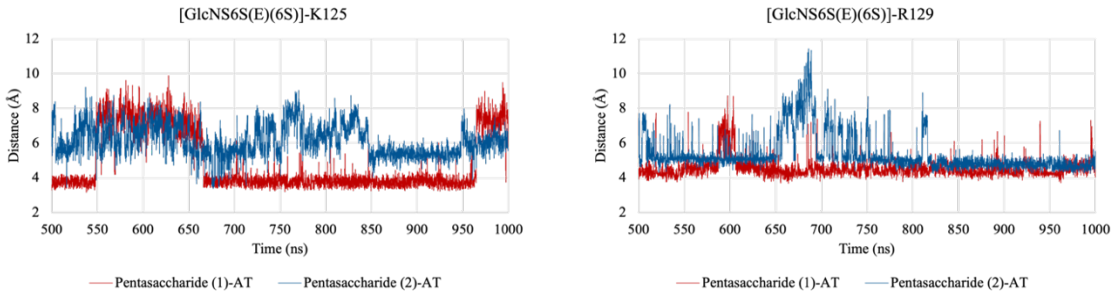

Figure A5.7. Distances (Å) of GlcNS6S(E)(6S) to R129 and K125 versus MD simulation time (ns) in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

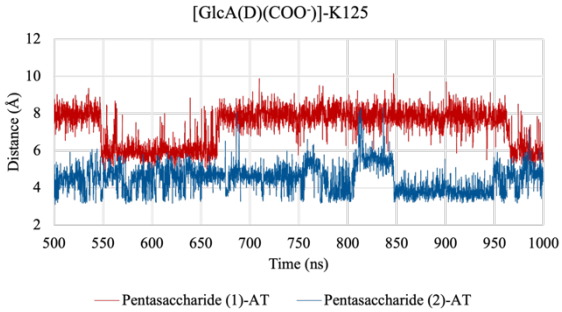

Figure A5.8. Distance (Å) between GlcA(D)(COO^-) and K125 versus MD simulation time (ns) in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

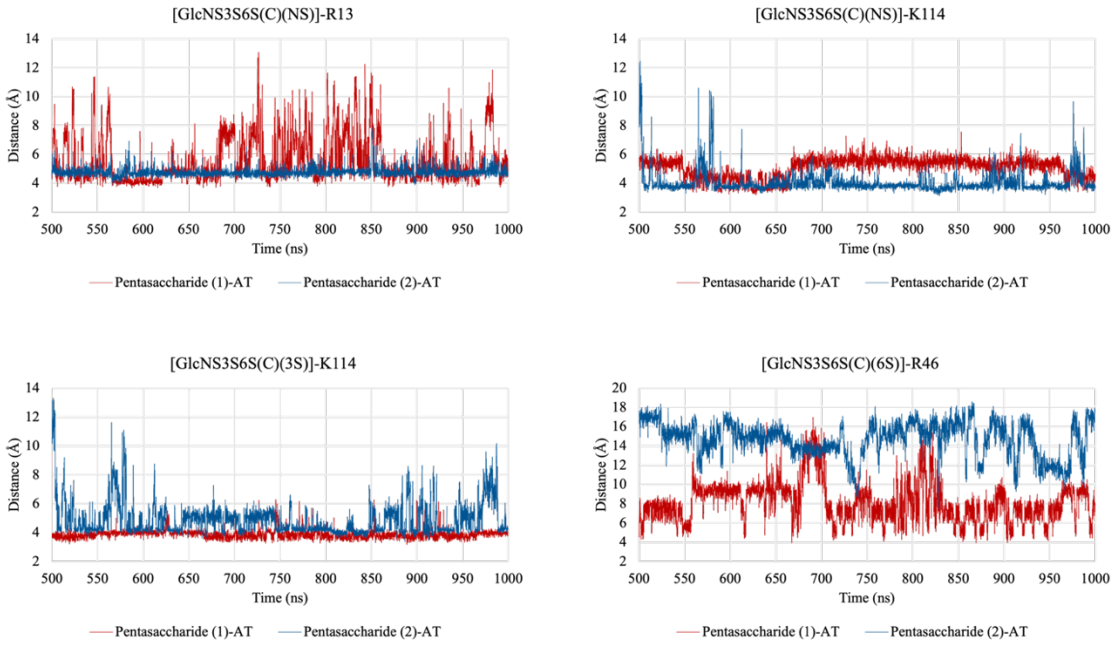

Figure A5.9. Distances (Å) of GlcNS3S6S(C)(6S) to R46, GlcNS3S6S(C)(6S) to K114 and GlcNS3S6S(C)(NS) to K114 and R13 versus MD simulation time (ns) in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

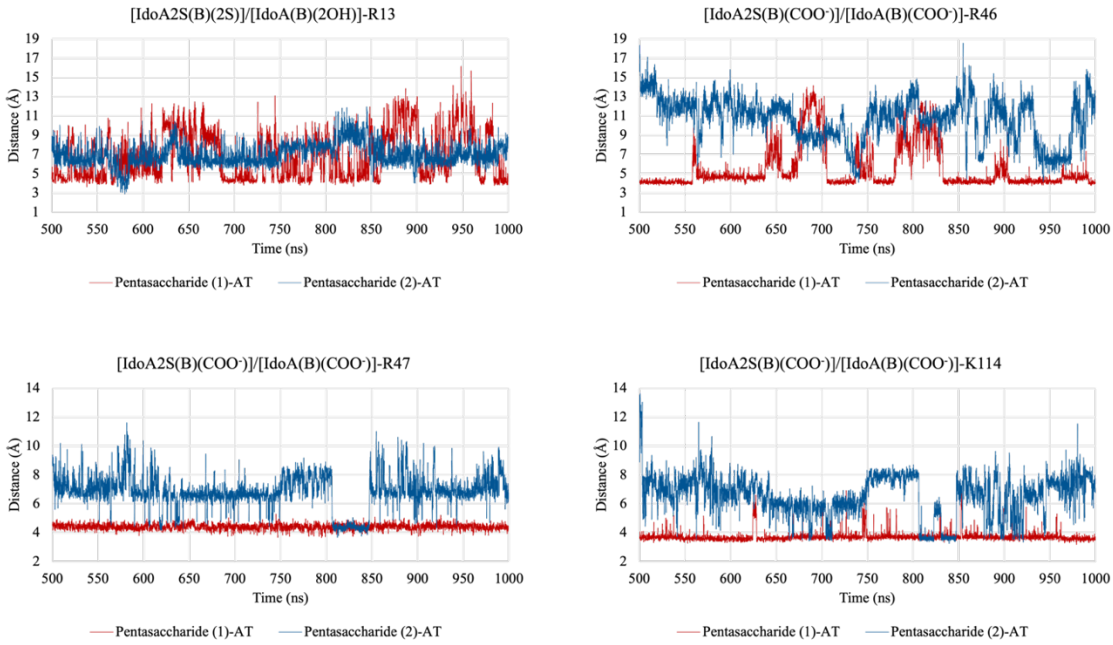

Figure A5.10. Distances (Å) of IdoA2S/IdoA(B)(COO^-) to K114, R47 and R46, IdoA2S(2S)/IdoA(B)(2OH) to R13 versus MD simulation time (ns) in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

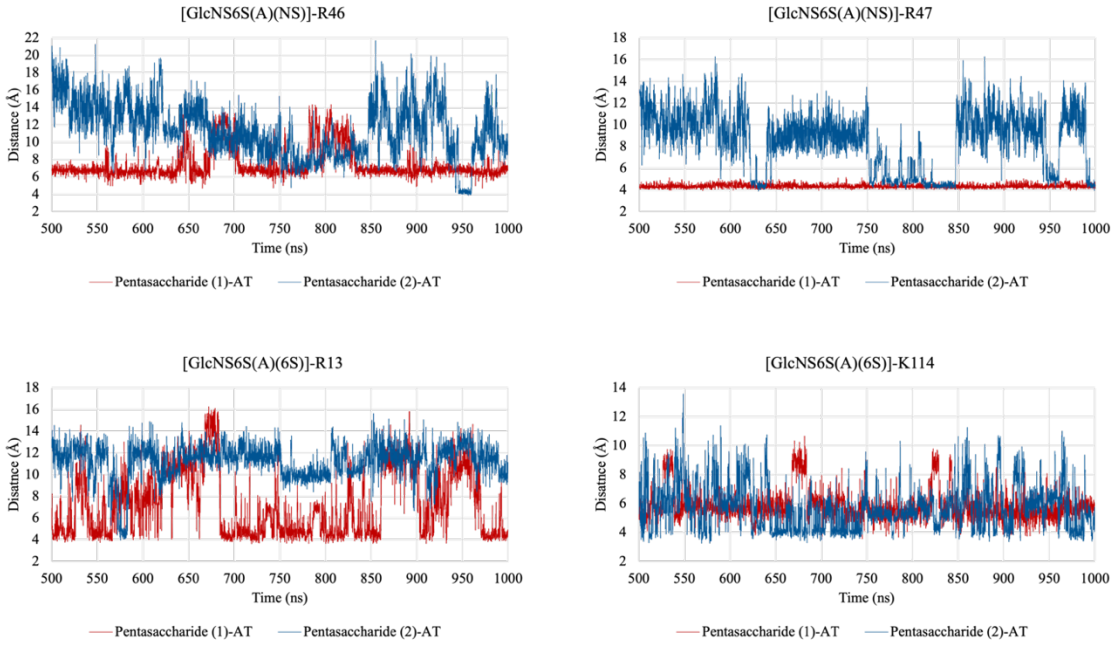

Figure A5.11. Distances (Å) of GlcNS6S(A)(6S) to K114 and R13, GlcNS6S(A)(NS) to R47 and R46 versus MD simulation time (ns) in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

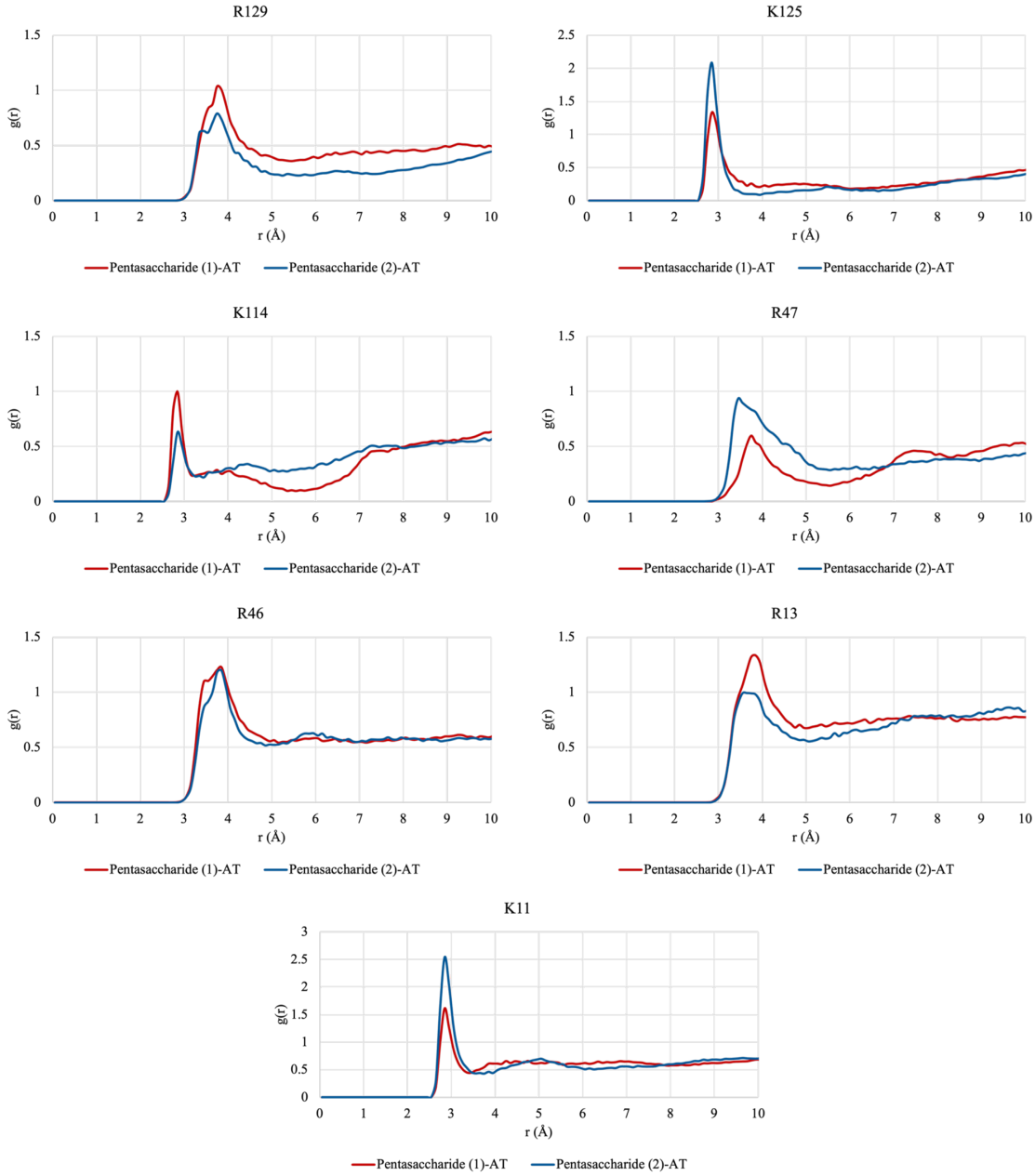

Figure A5.12. Radial distribution function [g(r)] curves between selected protein residues and water molecules in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

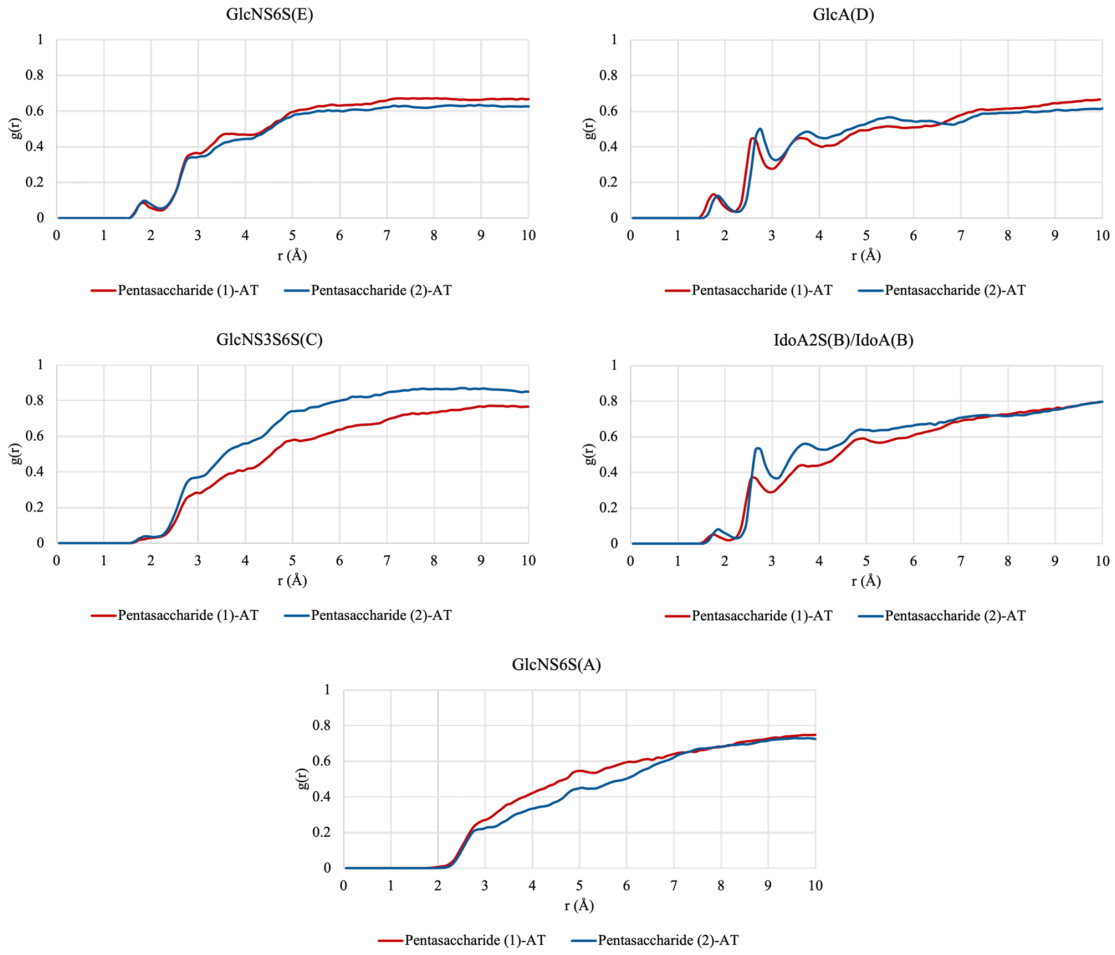

Figure A5.13. Radial distribution function [g(r)] curves between sugar units and water molecules in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

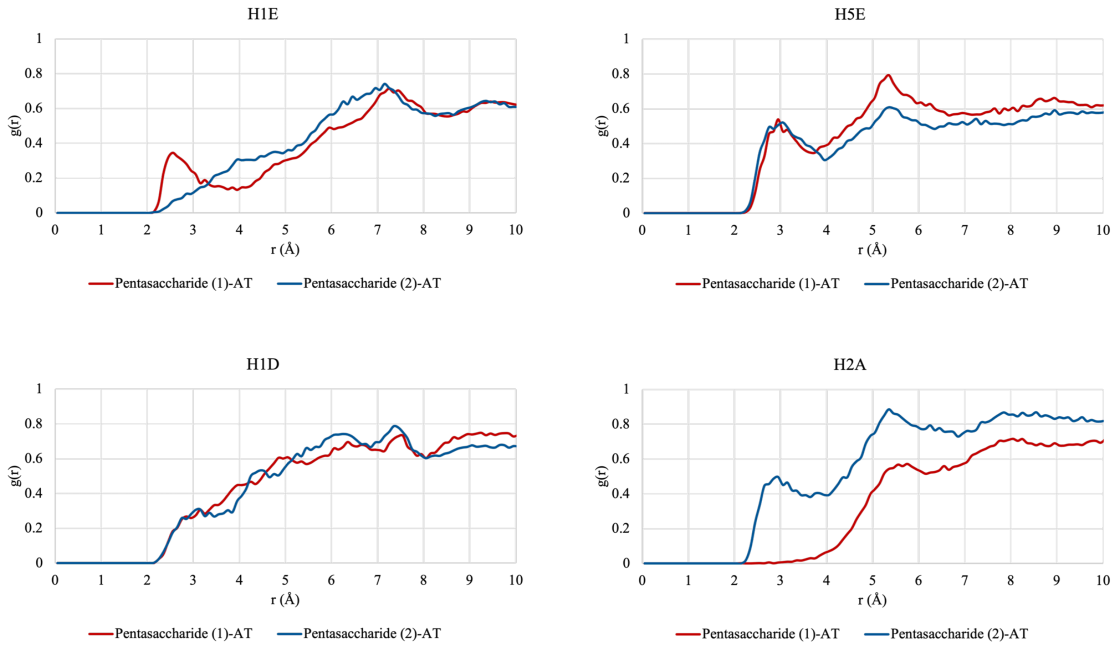

Figure A5.14. Radial distribution function [g(r)] curves between selected ligand protons and water molecules in the pentasaccharide (1)-AT and pentasaccharide (2)-AT complexes.

Table A5.1. Comparison between the binding epitopes obtained experimentally and calculated during MD simulation of pentasaccharide (1) in complex with AT. The 3rd column reports the average of STD values.

Proton	Exp STD	Sim STD
H1E	70	98
H2E	78	88
H5E	69	99
H1D	72	57
H1C	70	76
H2C	100	86
H6'C	82	82
H6''C	82	89
H1B	64	89
H2B	81	89
H1A	66	65
H2A	95	77
H3A	88	71
OCH_3(A)	82	56
OCH_3(A)	82	54
OCH_3(A)	82	52

Table A5.2. Comparison between simulated and experimental STD factors during MD simulation for selected ligand protons in the pentasaccharide (1)-AT complex. The R-factor was < 0.3.

		500-550 ns	550-600 ns	600-650 ns	650-700 ns	700-750 ns	750-800 ns	800-850 ns	850-900 ns	900-950 ns	950-1000 ns
Proton	**Exp STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**
H1E	70	100	96	97	96	98	97	100	100	100	100
H2E	78	87	85	87	87	86	90	91	88	88	86
H5E	69	100	100	100	100	100	100	100	98	99	98
H1D	72	55	56	55	54	55	62	58	56	57	62
H1C	70	80	76	73	75	76	81	73	74	77	80
H2C	100	87	81	84	84	83	89	86	86	92	87
H6'C	82	81	78	80	79	76	84	83	86	83	89
H6''C	82	88	89	88	88	86	90	90	89	89	89
H1B	64	87	92	89	94	89	91	88	86	87	82
H2B	81	89	93	90	94	89	92	88	87	87	85
H1A	66	62	65	63	64	62	64	65	68	69	68
H2A	95	79	75	73	75	75	73	76	80	79	82
H3A	88	73	72	71	71	67	68	70	72	72	74
OCH_3(A)	82	55	56	57	55	54	57	55	55	54	62
OCH_3(A)	82	52	54	55	53	50	53	54	53	54	59
OCH_3(A)	82	50	52	54	51	49	52	51	51	51	61

Table A5.3. Comparison between the binding epitopes obtained experimentally and calculated during MD simulation of pentasaccharide (2) in complex with AT. The 3rd column reports the average of STD values.

Proton	Exp STD	Sim STD
H1E	76	97
H2E	81	95
H5E	73	87
H1D	72	50
H1C	65	92
H2C	93	64
H6'C	66	53
H6''C	66	54
H4B	55	44
H2A	100	76
OCH_3(A)	59	70
OCH_3(A)	59	65
OCH_3(A)	59	62

Table A5.4. Comparison between simulated and experimental STD factors for selected ligand protons in the pentasaccharide (2)-AT complex during MD simulation. The R-factor was < 0.3.

		500-550 ns	550-600 ns	600-650 ns	650-700 ns	700-750 ns	750-800 ns	800-850 ns	850-900 ns	900-950 ns	950-1000 ns
Proton	**Exp STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**	**Sim STD**
H1E	76	100	100	96	100	100	100	100	100	100	100
H2E	81	93	94	100	94	97	99	95	97	98	98
H5E	73	85	90	89	88	91	86	87	87	88	92
H1D	72	47	49	47	48	53	49	48	46	47	47
H1C	65	91	97	99	92	96	94	94	90	91	96
H2C	93	62	63	65	63	61	60	57	58	57	61
H6'C	66	52	53	55	53	53	51	52	50	51	51
H6''C	66	54	55	61	54	52	54	49	50	48	53
H4B	55	45	45	43	42	45	45	42	41	40	42
H2A	100	71	76	76	75	76	75	72	69	71	73
OCH_3(A)	59	68	72	74	71	72	74	71	68	68	70
OCH_3(A)	59	62	65	69	67	67	66	62	63	63	66
OCH_3(A)	59	58	65	63	64	64	63	64	60	61	62

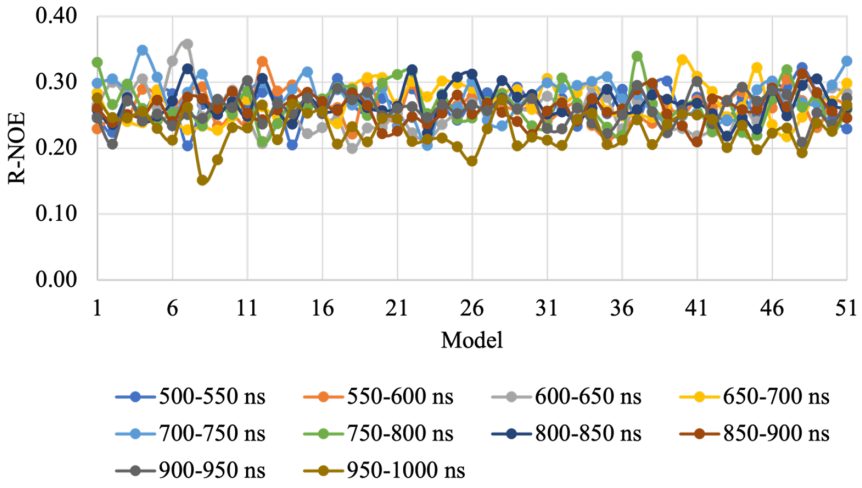

Figure A5.15. Evolution of the R-factor (R-NOE) of pentasaccharide (1) over the last 500 ns of MD simulation.

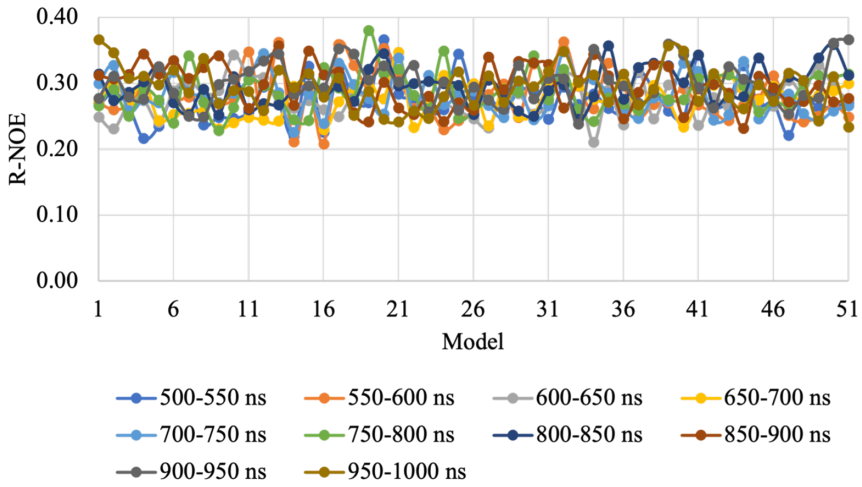

Figure A5.16. Evolution of the R-factor (R-NOE) of pentasaccharide (2) over the last 500 ns of MD simulation.

Table A5.5. $^{1}H/^{13}C$ chemical shift assignment of pentasaccharide (1).

Proton	GlcNS6S(E)	GlcA(D)	GlcNS6S(C)	IdoA2S(B)	GlcNS6S(A)
1	5.64/100.2	4.63/103.8	5.53/98.8	5.20/102.2	5.03/101.0
2	3.26/60.7	3.43/75.4	3.46/59.3	4.31/79.7	3.29/60.5
3	3.61/73.8	3.85/78.8	4.37/78.9	4.18/72.6	3.66/72.5
4	3.59/71.7	3.85/79.5	3.98/75.6	4.15/78.6	3.85/79.5
5	3.90/72.4	3.79/78.8	4.16/72.3	4.77/72.8	3.97/71.3
6	4.35-4.43/ 69.4	-	4.27-4.51/ 68.7	-	4.16-4.38/ 69.0
OCH_3	-	-	-	-	3.43/58.1

Table A5.6. $^{1}H/^{13}C$ chemical shift assignment of pentasaccharide (2).

Proton	GlcNS6S(E)	GlcA(D)	GlcNS6S(C)	IdoA(B)	GlcNS6S(A)
1	5.64/100.3	4.63/104.0	5.39/98.2	5.04/104.9	5.04/101.1
2	3.26/60.7	3.42/75.4	3.46/59.3	3.75/71.4	3.29/60.6
3	3.67/72.7	3.85/78.8	4.39/78.7	4.17/79.5	3.67/72.7
4	3.59/71.7	3.85/79.5	3.99/75.5	4.11/77.3	3.78/79.8
5	4.00/71.2	3.75/81.0	4.07/72.2	4.81/71.0	4.00/71.2
6	4.16-4.38/ 69.0	-	4.22-4.53/ 68.5	-	4.32/69.3
OCH_3	-	-	-	-	3.45/58.1

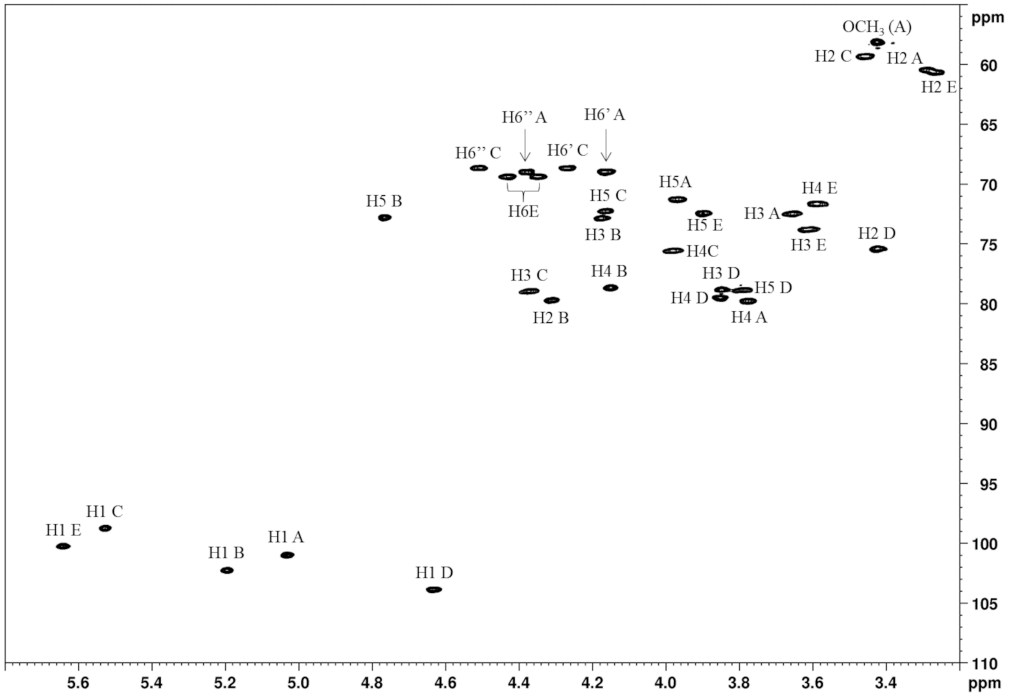

Figure A5.17. ^{1}H-^{13}C HSQC NMR spectrum of pentasaccharide (1).

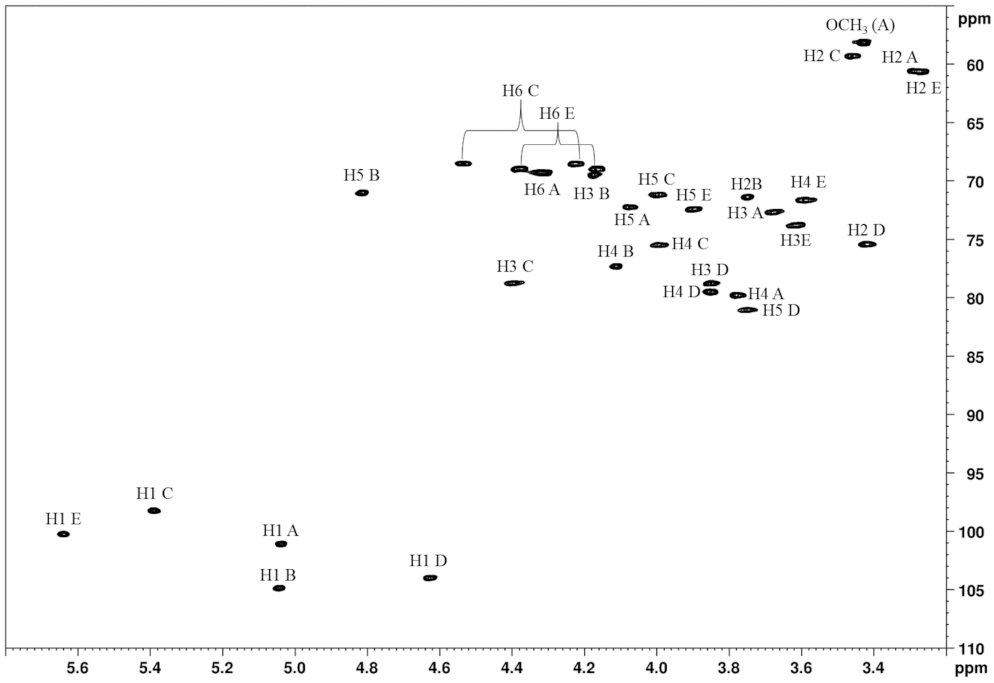

Figure A5.18. ^{1}H-^{13}C HSQC NMR spectrum of pentasaccharide (2).

Chapter 6

6. Concluding remarks

NMR spectroscopy is a versatile method for exploring protein-ligand interactions. It offers distinct advantages over orthogonal techniques, such as X-ray crystallography and cryo-electron microscopy (cryo-EM). NMR allows the analysis of protein-ligand interactions in solution, preserving native conditions and enabling the study of dynamic processes, such as conformational changes, fluctuations and transient interactions. NMR methods are particularly useful for investigating weak interactions, which can be difficult to detect with other techniques, providing information on ligands with low binding affinity. Additionally, NMR can determine binding constants, association/dissociation rates and thermodynamic parameters. Unlike X-ray crystallography and cryo-EM, NMR experiments require relatively small amounts of protein and ligand samples and do not involve sample destruction. The combined use of NMR spectroscopy and molecular modelling represents a valuable strategy for gaining insights into the molecular mechanisms underlying protein-ligand interactions.

This thesis focused on applying NMR techniques and computational methods to investigate the interactions between heparin/HS and three different proteins, contributing to a better understanding of their biological functions. The recognition process of HS by the S protein of SARS-CoV-2 was analysed, revealing the low specificity of the binding and supporting the co-receptor role of HS in the entry of the virus into host cells. Additionally, the structural features of binding and inhibitory mechanism of glycol-split heparins as potential and innovative inhibitors of HPSE were studied. Finally, a particular use of STD NMR and in silico approaches allowed to define the binding differences between two similar heparin pentasaccharides when interacting with AT.

Bibliography

1. Lindahl, U., Höök, M. *Annu. Rev. Biochem.* **1978**, *47*, 385–417.
2. Jackson, R. L., Busch, S. J., Cardin A. D. *Physiol. Rev.* **1991**, *71*, 481–539.
3. Kjellén, L., Lindahl, U. *Annu. Rev. Biochem.* **1991**, *60*, 443–475.
4. Yanagishita, M. *Experientia.* **1993**, *49*, 366–368.
5. Symbol Nomenclature for Graphical Representation of Glycans *Glycobiology* **2015**, *25*, 1323–1324.
6. Updates to the Symbol Nomenclature for Glycans guidelines *Glycobiology* **2019**, *29*, 620–624.
7. Rosenberg, R. D., Lam, L. *Proc. Natl. Acad. Sci. U. S. A.* **1979**, *76*, 1218–1222.
8. Gallagher, J. T., Lyon, M., Steward, W. P. *Biochem. J.* **1986**, *236*, 313–325.
9. Casu, B., Lindahl, U. *Adv. Carbohydr. Chem. Biochem.* **2001**, *57*, 159–206.
10. Mulloy, B., Forster, M. J., Jones, C., Davies, D. B. *Biochem J.* **1993**, *293*, 849–858.
11. Ferro, D. R., Provasoli, A., Ragazzi, M., Torri, G., Casu, B., Gatti, G., Jacquinet, J. C., Sinay, P., Petitou, M., Choay, J. *J. Am. Chem. Soc.* **1986**, *108*, 6773–6778.
12. Sanderson, P. N., Huckerby, T. N., Nieduszynski, I. A. *Biochem. J.* **1987**, *243*, 175-81.
13. Ferro, D. R., Provasoli, A., Ragazzi, M., Casu, B., Torri, G., Bossennec, V., Perly, B., Sinaÿ, P., Petitou, M., Choay, J. *Carbohydr. Res.* **1990**, *195*, 157–167.
14. Gallagher J. T., Walker, A. *Biochem. J.* **1985**, 230, 665–674.
15. Lindahl, U., Kusche-Gullberg, M., Kjelle´n, L. *J. Biol. Chem.* **1998**, *273*, 24979–24982.
16. Esko, J.D., Lindahl, U. *J. Clin. Invest.* **2001**, *108*, 169–173.
17. Lindahl, U., Kusche, M., Lidholt, K., Oscarsson, L. G. *Ann. N.Y. Acad. Sci.* **1989**, *556*, 36–50.
18. Jacobsson, I., Lindahl, U., Jensen, J. W., Rodén, L., Prihar, H., Feingold, D. S. *J. Biol. Chem.* **1984**, *259*, 1056–1063.
19. Rong, J., Habuchi, H., Kimata, K., Lindahl, U., Kusche-Gullberg, M. *Biochemistry* **2001**, *40*, 5548–5555.
20. Lindahl, U., Bäckström, G., Thunberg, L., Leder, I. G. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 6551–6555.
21. Capila, I., Linhardt, R. J. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 391–412.

22. Xu, D., Esko, J. D. *Annu. Rev. Biochem.* **2014**, *83*, 129–157.
23. Björk, I., Lindahl, U. *Mol. Cell. Biochem.* **1982**, *48*, 161–82.
24. Petitou, M., Casu, B., Lindahl, U. *Biochimie* **2003**, *85*, 83–89.
25. Jin, L., Abrahams, J. P., Skinner, R., Petitou, M., Pike, R. N., Carrell, R. W. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 14683–4688.
26. Desai, U. R., Petitou, M., Björk, I., Olson, S. T. *J. Biol. Chem.* **1998**, *273*, 7478–7487.
27. Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., Zako, M. *Annu. Rev. Biochem.* **1999**, 68, 729–777.
28. Bishop, J., Schuksz, M., Esko, J. *Nature* **2007**, *446*, 1030–1037.
29. Ori, A., Wilkinson, M. C., Fernig, D. G. *Front. Biosci.* **2008**, *13*, 4309–4338.
30. Ori, A., Wilkinson, M. C., Fernig, D. G. *J. Biol. Chem.* **2011**, *286*, 19892–19904.
31. Caldwell, E. E., Nadkarni, V. D., Fromm, J. R., Linhardt, R. J., Weiler, J. M. *Int. J. Biochem. Cell. Biol.* **1996**, *28*, 203–216.
32. Fromm, J. R., Hileman, R. E., Caldwell, E. E. O., Weiler, J. M., Linhardt, R. J. *Arch. Biochem. Biophys.* **1997**, *343*, 92–100.
33. Nordenman, B., Björk, I. *Biochim. Biophys. Acta.* **1981**, *672*, 227–238.
34. Esko, J. D., Selleck, S. B. *Annu. Rev. Biochem.* **2002**, *71*, 435–471.
35. Casu, B., Guerrini, M., Torri, G. *Curr. Pharm. Des.* **2004**, *10*, 939–49.
36. Sasisekharan R., Raman R., Prabhakar V. *Annu. Rev. Biomed. Eng.* **2006**, *8*, 181–231.
37. Cardin A. D., Weintraub H. J. *Arterioscler. Thromb. Vasc. Biol*. **1989**, *9*, 21–32.
38. Hartenian, E., Nandakumar, D., Lari, A., Ly, M., Tucker, J. M., Glaunsinger, B. A. *J. Biol. Chem.* **2020**, *295*, 12910-12934.
39. Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Perrett, H. R., Painter, C. D., Narayanan, A., Majowicz, S. A., Kwong, E. M., McVicar, R. N., Thacker, B. E., Glass, C. A., Yang, Z., Torres, J. L., Golden, G. J., Bartels, P. L., Porell, R. N., Garretson, A. F., Laubach, L., Feldman, J., Yin, X., Pu, Y., Hauser, B. M., Caradonna, T. M., Kellman, B. P., Martino, C., Gordts, P. L. S. M., Chanda, S. K., Schmidt, A. G., Godula, K., Leibel, S. L., Jose, J., Corbett, K. D., Ward, A. B., Carlin, A. F., Esko, J. D. *Cell* **2020**, *183*, 1043−1057.
40. Mycroft-West, C. J., Su, D., Pagani, I., Rudd, T. R., Elli, S., Gandhi, N. S., Guimond, S. E., Miller, G. J., Meneghetti, M. C. Z., Nader, H. B., Li, Y., Nunes, Q. M., Procter, P., Mancini, N., Clementi, M., Bisio, A., Forsyth, N. R., Ferro, V.,

Turnbull, J. E., Guerrini, M., Fernig, D. G., Vicenzi, E., Yates, E. A., Lima, M. A. Skidmore, M. A. *Thromb. Haemost.* **2020** *120*, 1700–1715.

41. Remuzzi, G., Schiaffino, S., Santoro, M. G., FitzGerald, G. A., Melino, G., Patrono, C. *Front. Pharmacol.* **2022**, 13, 987816.
42. Thachil J. *J. Thromb. Haemost.* **2020**, *18*, 1020–1022.
43. Hippensteel, J. A., LaRiviere, W. B., Colbert, J. F., Langouet-Astrié, C. J., Schmidt E. P. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **2020**, *319*, L211–L217.
44. Vicenzi, E., Canducci, F., Pinna, D., Mancini, N., Carletti, S., Lazzarin, A., Bordignon, C., Poli, G., Clementi, M. *Emerg. Infect. Dis.* **2004**, *10*, 413–418.
45. Vlodasky, I., Sanderson, R. D., Ilan, *N. Adv. Exp. Med. Bio.* **2020**, 1221.
46. Casu, B., Vlodasky, I., Sanderson, R. D. *Pathophysiol. Haemost. Thromb.* **2008**, *36*, 195–203.
47. Monaco, S., Tailford, L. E., Juge, N., Angulo, J. *Angew Chem Int Ed Engl.* **2017**, *56*, 15289–15293.
48. Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Mohnen, D., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., Seeberger, P. H. (Eds.) *Essentials of Glycobiology* **2022**.
49. Claridge, T. D. W. (2016). High-Resolution NMR Techniques in Organic Chemistry (Third Edition). Elsevier Science.
50. Meyer, B., Peters, T. *Angew. Chem. Int. Ed.* **2003**, *42*, 864–890.
51. Carlomagno, T. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 245–266.
52. Mayer, M., Meyer, B. *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
53. Mayer, M., Meyer, B. *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.
54. Angulo, J., Enriquez-Navas, P. M., Nieto, P. M. *Chem. Eur. J.* **2010**, *16*, 7803–7812.
55. Jayalakshmi, V., Krishna, N. R. *J. Magn. Reson.* **2002**, 155, 106–118.
56. Nepravishta, R., Walpole, S., Hicks, T., Muñoz-García, J. C., Angulo, J. *ChemRxiv.* **2022**.
57. Liepinsh, E., Otting, G. *Magn. Reson. Med.* **1996**, *35*, 30–42.
58. Dalvit, C., Fogliatto, G., Steward, A., Veronesi, M., Stockman, B. *J. Biomol. NMR* **2001**, *21*, 349.
59. Campbell, A. P., Sykes, B. D. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 99–122.

60. Moseley, H. N. B., Curto, E. V., Krishna, N. R. *J. Magn. Reson. Ser. B* **1995**, *108*, 243–261.
61. Cramer, C. J. (2004). Essential of Computational Chemistry – Theories and Models (Second Edition). Wiley.
62. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., Tan, W. China Novel Coronavirus Investigating and Research Team *N. Engl. J. Med.* **2020**, *382*, 727–733.
63. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B. *Lancet* **2020**, *395*, 497–506.
64. Klok, F. A., Kruip, M. J. H. A., van der Meer, N. J. M., Arbous, M. S., Gommers, D. A. M. P. J., Kant, K. M., Kaptein, F. H. J., van Paassen, J., Stals, M. A. M., Huisman, M. V., Endeman, H. *Thromb. Res.* **2020**, *191*, 145–147.
65. Nalbandian, A., Sehgal, K., Gupta, A., Madhavan, M. V., McGroder, C., Stevens, J. S., Cook, J. R., Nordvig, A. S., Shalev, D., Sehrawat, T. S., Ahluwalia, N., Bikdeli, B., Dietz, D., Der-Nigoghossian, C., Liyanage-Don, N., Rosner, G. F., Bernstein, E. J., Mohan, S., Beckley, A. A., Seres, D. S., Choueiri, T. K., Uriel, N., Ausiello, J. C., Accili, D., Freedberg, D. E., Baldwin, M., Schwartz, A., Brodie, D., Garcia, C. K., Elkind, M. S. V., Connors, J. M., Bilezikian, J. P., Landry, D. W., Wan, E. Y. *Nat Med.* **2021**, *27*, 601–615.
66. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N. H., Nitsche, A., Müller, M. A., Drosten, C., Pöhlmann, S. *Cell* **2020**, *181*, 1–10.
67. Zhang, Q., Chen, C. Z., Swaroop, M., Xu, M., Wang, L., Lee, J., Wang, A. Q., Pradhan, M., Hagen, N., Chen, L., Shen, M., Luo, Z., Xu, X., Xu, Y., Huang, W., Zheng, W., Ye, Y. *Cell Discov.* **2020**, *6*, 80.
68. Yu, M., Zhang, T., Zhang, W., Sun, Q., Li, H., Li, J. P. *Front. Mol. Biosci.* **2021**, *7*, 628551.
69. Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., Veesler, D. *Cell* **2020**, *181*, 281–292.
70. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., Graham, B. S., McLellan, J. S. *Science* **2020**, *367*, 1260–1263.

71. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X. *Nature* **2020**, *581*, 215–220.
72. Sztain, T., Ahn, S. H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., McCool, R. S., Kearns, F. L., Acosta-Reyes, F., Maji, S., Mashayekhi, G., McCammon, J. A., Ourmazd, A., Frank, J., McLellan, J. S., Chong, L. T., Amaro, R. E. *Nat. Chem.* **2021**, *13*, 963–968.
73. de Haan, C. A., Li, Z., te Lintelo, E., Bosch, B. J., Haijema, B. J., Rottier, P. J. *J. Virol.* **2005**, *79*, 14451–14456.
74. Milewska, A., Zarebski, M., Nowak, P., Stozek, K., Potempa, J., Pyrc, K. *J. Virol.* **2014**, *88*, 13221–13230.
75. Krufka, A., Guimond, S., Rapraeger, A. C. *Biochemistry* **1996**, *35*, 11131–11141.
76. Lanzi, C., Cassinelli G. *Biochem. Pharmacol.* **2020**, *178*, 114084.
77. Liu, L., Chopra, P., Li, X., Bouwman, K. M., Tompkins, S. M., Wolfert, M. A., de Vries, R. P., Boons, G. J. *ACS Cent. Sci.* **2021**, *7*, 1009−1018.
78. Kim, S. Y., Jin, W., Sood, A., Montgomery, D. W., Grant, O. C., Fuster, M. M., Fu, L., Dordick, J. S., Woods, R. J., Zhang, F., Linhardt, R. *J. Antivir. Res.* **2020**, *181*, 104873.
79. Paiardi, G., Richter, S., Oreste, P., Urbinati, C., Rusnati, M., Wade, R. C. *J. Biol. Chem.* **2022**, *298*, 101507.
80. Chittum, J. E., Sankaranarayanan, N. V., O'Hara, C. P., Desai, U. R. *ACS Med. Chem. Lett.* **2021**, *12*, 1710−1717.
81. Schuurs, Z. P., Hammond, E., Elli, S., Rudd, T. R., Mycroft-West, C. J., Lima, M. A., Skidmore, M. A., Karlsson, R., Chen, Y. H., Bagdonaite, I., Yang, Z., Ahmed, Y. A., Richard, D. J., Turnbull, J., Ferro, V., Coombe, D. R., Gandhi, N. S. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2806–2818.
82. Torri, G., Casu, B., Gatti, G., Petitou, M., Choay, J., Jacquinet, J. C., Sinaÿ, P. *Biochem. Biophys. Res. Commun.* **1985**, *128*, 134–140.
83. Hricovíni, M., Torri G. *Carbohydr. Res.* **1995**, *268*, 159–175.
84. Mikhailov, D., Mayo, K. H., Vlahov, I. R., Toida, T., Pervin, A., Linhardt, R. J. *Biochem. J.* **1996**, *318*, 93–102.
85. Cagno, V., Tseligka, E. D., Jones, S. T., Tapparel, C. *Viruses*, **2019**, 11, 596.
86. Nie, C., Pouyan, P., Lauster, D., Trimpert, J., Kerkhoff, Y., Szekeres, G. P., Wallert, M., Block, S., Sahoo, A. K., Dernedde, J., Pagel, K., Kaufer, B. B., Netz, R. R., Ballauff, M., Haag, R. *Angew. Chem. Int. Ed.* **2021**, *60*, 15870–15878.

87. Tree, J. A., Turnbull, J. E., Buttigieg, K. R., Elmore, M. J., Coombes, N., Hogwood, J., Mycroft-West, C. J., Lima, M. A., Skidmore, M. A., Karlsson, R., Chen, Y. H., Yang, Z., Spalluto, C. M., Staples, K. J., Yates, E. A., Gray, E., Singh, D., Wilkinson, T., Page, C. P., Carroll, M. W. *Br. J. Pharmacol.* **2021**, *178*, 626–635.

88. Guimond, S. E., Mycroft-West, C. J., Gandhi, N. S., Tree, J. A., Le, T. T., Spalluto, C. M., Humbert, M. V., Buttigieg, K. R., Coombes, N., Elmore, M. J., Wand, M., Nyström, K., Said, J., Setoh, Y. X., Amarilla, A. A., Modhiran, N., Sng, J. D. J., Chhabra, M., Young, P. R., Rawle, D. J., Lima, M. A., Yates, E. A., Karlsson, R., Miller, R. L., Chen, Y. H., Bagdonaite, I., Yang, Z., Stewart, J., Nguyen, D., Laidlaw, S., Hammond, E., Dredge, K., Wilkinson, T. M. A., Watterson, D., Khromykh, A. A., Suhrbier, A., Carroll, M. W., Trybala, E., Bergström, T., Ferro, V., Skidmore, M. A., Turnbull, J. E. *ACS Cent. Sci.* **2022**, *8*, 527–545.

89. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., Olson, A. J. *J. Comput. Chem.* **2009**, *30*, 2785–2791.

90. Case, D. A., Darden, T. A., Cheatham, T. E. III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Walker, R.C., Zhang, W., Merz, K. M., Roberts, B., Hayik, S., Roitberg, A., Seabra, G., Swails, J., Goetz, A. W., Kolossváry, I., Wong, K. F., Paesani, F., Vanicek, J., Wolf, R. M., Liu, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Cai, Q., Ye, X., Wang, J., Hsieh, M.-J., Cui, G., Roe, D. R., Mathews, D. H., Seetin, M. G., Salomon-Ferrer, R., Sagui, C., Babin, V., Luchko, T., Gusarov, S., Kovalenko, A., Kollman P.A. AMBER 12, 2012, University of California, San Francisco.

91. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.

92. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926–935.

93. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

94. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case D. A. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

95. Weis, A., Katebzadeh, K., Söderhjelm, P., Nilsson, I., Ryde, U. *J. Med. Chem.* **2006**, *49*, 6596–6606.

96. Miller, B. R., 3rd, McGee, T. D., Jr, Swails, J. M., Homeyer, N., Gohlke, H., Roitberg, A. E. *Chem. Theory Comput.* **2012**, *8*, 3314–3321.
97. Wang, J., Morin, P., Wang, W., Kollman, P. A. *J. Am. Chem. Soc.* **2001**, *123*, 5221–5230.
98. Hricovíni, M., Guerrini, M., Bisio, A. *Eur. J. Biochem.* **1999**, *261*, 789–801.
99. Ragnoli, B., Da Re, B., Galantino, A., Kette, S., Salotti, A., Malerba, M. *Int. J. Mol. Sci.* **2023**, 24, 8945.
100. Tang, N., Bai, H., Chen, X., Gong, J., Li, D., Sun, Z. *J. Thromb. Haemost.* **2020**, *18*, 1094–1099.
101. Martínez-Botía, P., Bernardo, Á., Acebes-Huerta, A., Caro, A., Leoz, B., Martínez-Carballeira, D., Palomo-Antequera, C., Soto, I., Gutiérrez, L. *J. Clin. Med.* **2021**, *10*, 1073.
102. Makarem, A., Zareef, R., Abourjeili, J., Nassar, J. E., Bitar, F., Arabi, M. *Front. Pharmacol.* **2023**, *14*, 1159363.
103. Mauri, L., Boccardi, G., Torri, G., Karfunkle, M., Macchi, E., Muzi, L., Keire, D., Guerrini, M. *J. Pharm. Biomed. Anal.* **2017**, *136*, 92–105.
104. Ögren, S., Lindahl, U. *J Biol Chem.* **1975**, *250*, 2690–2697.
105. Vlodavsky, I., Friedmann, Y., Elkin, M., Aingorn, H., Atzmon, R., Ishai-Michaeli, R., Bitan, M., Pappo, O., Peretz, T., Michal, I., Spector, L., Pecker, I. *Nat. Med.* **1999**, *5*, 793–802.
106. Hulett, M. D., Freeman, C., Hamdorf, B. J., Baker, R. T., Harris, M. J., Parish, C. R. *Nat. Med.* **1999**, *5*, 803–809.
107. Kussie, P. H., Hulmes, J. D., Ludwig, D. L., Patel, S., Navarro, E. C., Seddon, A. P., Giorgio, N. A., Bohlen, P. *Biochem. Biophys. Res. Commun.* **1999**, *261*, 183–187.
108. Fairbanks, M. B., Mildner, A. M., Leone, J. W., Cavey, G. S., Mathews, W. R., Drong, R. F., Slightom, J. L., Bienkowski, M. J., Smith, C. W., Bannow, C. A., Heinrikson, R. L. *J Biol Chem.* **1999**, *274*, 29587–29590.
109. Toyoshima, M., Nakajima, M. *J Biol Chem.* **1999**, *274*, 24153–24160.
110. Davies, G., Henrissat, B. *Structure* **1995**, *3*, 853–859.
111. Pikas, D. S., Li, J. P., Vlodavsky, I., Lindahl, U. *J. Biol. Chem.* **1998**, *273*, 18770–18777.
112. Okada, Y., Yamada, S., Toyoshima, M., Dong, J., Nakajima, M., Sugahara, K. *J. Biol. Chem.* **2002**, *277*, 42488–42495.

113. Peterson, S. B., Liu, J. *J. Biol. Chem.* **2010**, *285*, 14504–14513.
114. Wu, L., Viola, C. M., Brzozowski, A. M., Davies, G. J. *Nat. Struct. Mol. Biol.* **2015**, 22, 1016–1023.
115. Hulett, M. D., Hornby, J. R., Ohms, S. J., Zuegg, J., Freeman, C., Gready, J. E., Parish, C. R. *Biochemistry* **2000**, *39*, 15659–15667.
116. Levy-Adam, F., Abboud-Jarrous, G., Guerrini, M., Beccati, D., Vlodavsky, I., Ilan, N. *J. Biol. Chem.* **2005**, *280*, 20457–20466.
117. Okada, Y.; Yamada, S.; Toyoshima, M.; Dong, J.; Nakajima, M.; Sugahara, K. *J. Biol. Chem.* **2002**, *277*, 42488–42495.
118. Rivara, S., Milazzo, F. M., Giannini, G. *Future Med. Chem.* **2016**, *8*, 647–680.
119. Naggi, A., Casu, B., Perez, M., Torri, G., Cassinelli, G., Penco, S., Pisano, C., Giannini, G., Ishai-Michaeli, R., Vlodavsky, I. *J. Biol. Chem.* **2005**, *280*, 12103–12113.
120. Ferro, V., Dredge, K., Liu, L., Hammond, E., Bytheway, I., Li, C., Johnstone, K., Karoli, T., Davis, K., Copeman, E., Gautam, A. *Semin. Thromb. Hemost.* **2007**, 33, 557–568.
121. Zhou, H., Roy, S., Cochran, E., Zouaoui, R., Chu, C. L., Duffner, J., Zhao, G., Smith, S., Galcheva-Gargova, Z., Karlgren, J., Dussault, N., Kwan, R. Y., Moy, E., Barnes, M., Long, A., Honan, C., Qi, Y. W., Shriver, Z., Ganguly, T., Schultes, B., Venkataraman, G., Kishimoto, T. K. *PLoS one*, **2011**, *6*, e21106.
122. Ferro, V., Liu, L., Johnstone, K. D., Wimmer, N., Karoli, T., Handley, P., Rowley, J., Dredge, K., Li, C. P., Hammond, E., Davis, K., Sarimaa, L., Harenberg, J., Bytheway, I. *J. Med. Chem.* **2012**, *55*, 3804–3813.
123. Casu B, Guerrini M, Naggi A., Perez M., Torri G., Ribatti D., Carminati P., Giannini G., Penco SD., Pisano C., Belleri M. Rusnati M., Presta M. *Biochemistry* **2002**, *41*, 10519–10528.
124. Ni, M., Elli, S., Naggi, A., Guerrini, M., Torri, G., Petitou, M. *Molecules* **2016**, *21*, 1602.
125. Hammond, E., Li, C.P., Ferro, V. *Anal. Biochem.* **2010**, *396*, 112–116.
126. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shaw, D. E., Shelley, M., Perry, J. K., Francis, P., Shenkin, P. S. *J. Med. Chem.* **2004**, *47*, 1739–1749.
127. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., Banks, J. L. *J. Med. Chem.* **2004**, *47*, 1750–1759.

128. Lever, R., Mulloy, B., & Page, C.P. (2012). Heparin - A Century of Progress. Handbook of Experimental Pharmacology.
129. Choay, J., Petitou, M., Lormeau, J. C., Sinay¨, P., Casu, B., Gatti, G. *Biochem. Biophys. Res. Commun.* **1983**, *116*, 492–499.
130. Lindahl, U., Thunberg, L., Backstrom, G., Riesenfeld, J., Nordling, K., Bjork, I. *J. Biol. Chem.* **1984**, *259*, 12368–12376.
131. Craig, P. A., Olson, S. T., Shore, J. D. *J. Biol. Chem.* **1989**, *264*, 5452–5461.
132. Olson, S. T., Björk, I., Sheffer, R., Craig, P. A., Shore, J. D., Choay, J. *J. Biol. Chem.* **1992**, *267*, 12528–12538.
133. Olson, S. T., Richard, B., Izaguirre, G., Schedin-Weiss, S., Gettins, P. G. *Biochimie* **2010**, *92*, 1587–1596.
134. Hricovíni, M., Guerrini, M., Bisio, A., Torri, G., Petitou, M., Casu, B. *Biochem. J.* **2001**, *359*, 265–272.
135. Guerrini, M., Guglieri, S., Casu, B., Torri, G., Mourier, P., Boudier, C., Viskov, C. *J. Biol. Chem.* **2008**, *283*, 26662–26675.
136. Stancanelli, E., Elli, S., Hsieh, P. H., Liu, J., Guerrini, M. *Chembiochem.* **2018**, 19, 1178–1188.
137. Levine, B. G., Stone, J. E., Kohlmeyer, A. *J. Comput. Phys.* **2011**, *230*, 3556–3569.
138. Humphrey, W., Dalke, A. Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33–38.