cicCartuja centro de investigaciones científicas isla de la cartuja

Consejo Superior de Investigaciones Científicas
Junta de Andalucía
Universidad de Sevilla

Tesis Doctoral:

Insights on the Structure and Dynamics of Glycosaminoglycans and their Interactions with Langerin: NMR and Computational Studies

Juan Carlos Muñoz García

Sevilla, 2013

UNIVERSIDAD DE SEVILLA

CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Laboratorio de Glicosistemas

Instituto de Investigaciones Químicas

Insights on the Structure and Dynamics of Glycosaminoglycans and their Interactions with Langerin: NMR and Computational Studies

por

Juan Carlos Muñoz García

Disertación presentada por la **Universidad de Sevilla** para aspirar al **Título de Doctor en Química**

Juan Carlos Muñoz García

Sevilla, 2013

Dr. PEDRO MANUEL NIETO MESA, Investigador Científico (CSIC), y Dr. JESÚS ANGULO ÁLVAREZ, Doctor Contratado "Ramón y Cajal" (CSIC)

CERTIFICAN

Que el trabajo titulado ***Insights on the Structure and Dynamics of Glycosaminoglycans and their Interactions with Langerin: NMR and Computational Studies*** se ha llevado a cabo bajo nuestra dirección y asesoramiento en los laboratorios del Instituto de Investigaciones Químicas (IIQ, CSIC-US) del Centro de Investigaciones Científicas Isla de la Cartuja (CICCartuja), Sevilla, constituyendo la memoria presentada por Juan Carlos Muñoz García (Licenciado en Química) para optar al grado de Doctor en Química.

Sevilla, julio de 2013

Los directores

Dr. Pedro M. Nieto Mesa

Dr. Jesús Angulo Álvarez

El trabajo de investigación recogido en la presente memoria se ha llevado a cabo fundamentalmente en el *Insituto de Investigaciones Químicas* (IIQ), que forma parte del *Centro de Investigaciones Científicas Isla de la Cartuja*, centro mixto del *Consejo Superior de Investigaciones Científicas* (CSIC) y la *Universidad de Sevilla* (US). El doctorando ha sido financiado mediante una beca predoctoral JAE-predoc del CSIC obtenida por concurrencia competitiva, durante el periodo 2009-2013.

Dicha beca predoctoral también ha financiado dos estancias cortas del doctorando, la primera en los laboratorios del Prof. Robert J. Woods, en el *Complex Carbohydrate Research Center* (CCRC) de la *Universidad de Georgia* (Athens, Georgia, EEUU), y la segunda en el grupo liderado por la Dr. Anne Imberty, del *Centre de Recherches sur les Macromolécules Végétales* (CERMAV-CNRS), Grenoble (Francia).

Algunos de los experimentos de RMN se han realizado en el Laboratorio de RMN de Barcelona (LRB), adscrito a los *Centres Científics i Tecnològics* (CCiT) de la *Universidad de Barcelona*, haciendo uso del espectrómetro de 800 MHz. Esto ha sido posible gracias a la financiación recibida por el antiguo *Ministerio de Ciencia e Innovación*.

A Encarni

A mis padres

Nota al lector

El trabajo de investigación recogido en presente memoria de Tesis Doctoral ha sido estructurado en 5 capítulos, los dos primeros cómo contextualización de las bases químicas y biologícas en las que se apoya el estudio (Capítulo 1), así como de los aspectos metodologícos empleados (Capítulo 2). En los Capítulos 3 y 4 se aportan los resultados derivados de la investigación, en primer lugar aquellos relacionados con la estructura y dinámica de oligosacáridos tipo heparina (Capítulo 3), y posteriormente los estudios de reconocimiento molecular de tales oligosacáridos con receptores (proteínas) de relevancia biológica (Capítulo 4). Finalmente, en el Capítulo 5 se recogen las conclusiones más relevantes.

Con el objetivo de optar a la mención de "Doctor Internacional" por la Universidad de Sevilla, los capítulos 1, 2, 3 y 4 han sido redactados en inglés, mientras que el Capítulo de conclusiones y el *Resumen de la Tesis* se han escrito en castellano.

Dada la diversidad de procedimientos utilizados, fundamentalmente en lo que respecta a las técnicas computacionales empleadas, y con el objetivo de facilitar la aportación de detalles específicos, cada desarrollo metodologíco ha sido descrito cómo subcapítulo adicional dentro de los capítulos 3 y 4. Por otro lado, la numeración de las ecuaciones, tablas y figuras es independiente en cada capítulo, mientras que las referencias bibliográficas se recogen conjuntamente al final de la memoria.

Esta Tesis Doctoral consta además de un Apéndice, al que se hace referencia con frecuencia en los capítulos 3 y 4, que recoge espectros, tablas e imágenes 3D de alta resolución que facilitan la compresión del análisis y discusión de resultados expuestos en la memoria. Sin embargo, por motivos de espacio dicho Apéndice se adjunta cómo documento independiente en la version electrónica, pero no en la memoria impresa.

Abbreviations

actDC	activated Dendritic Cell
AMBER	Assisted Model Building with Energy Refinement
AA	All Atom
CCR5	C- C chemokine Receptor 5
CHARMM	Chemistry at HARvard Macromolecular Mechanics
COSY	^{1}H-^{1}H COrrelation SpectroscopY
CLR	C-type Lectin Receptor
CRD	Carbohydrate Recognition Domain
CXCR4	CXC chemokine Receptor 4
DAMP	Danger-Associated Molecular Patterns
DCs	Dendritic Cells
DC-SIGN	Dendritic Cell-Specific ICAM-3 Grabbing Non-integrin
DPFGSE	Double Pulsed Field Gradient Spin Echo
DQF-COSY	Double Quantum Filter COSY
DMEC	Dermal Microvascular Endothelial Cell
DSF	Differential Scanning Fluorimetry
EC	Endothelial Cell
ECD	Extra Cellular Domain
EDTA	Ethyleneiaminetetraacetic Acid
ELISA	Enzyme-Linked ImmunoSorbent Assay
FP	Fluorescence Polarization
FRET	Fluorescence Resonance Energy Transfer
GAG	Glycosaminoglycan
GLIDE	Grid-Based Ligand Docking with Energetics
GROMOS	GROningen MOlecular Simulation
HIV	Human Immunodeficiency Virus
HSQC	^{1}H-^{13}C Heteronuclear Single Quantum Coherence
IC_{50}	half maximal Inhibitory Concentration

ICAM	InterCellular Adhesion Molecule
IFD	Induced Fit Docking
IL	InterLeukin
ISPA	Isolated Spin Pair Approximation
ITAM	Immunoreceptor Tyrosine-based Activation Motif
ITC	Isothermal Titration Calorimetry
MCMM	Monte Carlo Multiple Minimum
MD	Molecular Dynamics
Mo	Monocyte
MST	Microscale Thermophoresis
Mϕ	Macrophages
NLR	NOD (Nucleotide-binding Oligomerization Domain)-Like Receptor
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect SpectroscopY
OPLS	Optimized Potential for Liquid Simulations
PAMP	Pathogen-Associated Molecular Pattern
PDB	Protein Data Bank
QM	Quantum Mechanics
PRR	Pathogen Recognition Receptor
RMSD	Root-Mean-Square Deviation
ROE	Rotating-frame Overhauser Effect
ROESY	Rotating-frame Overhauser Effect SpectroscopY
sLex	sialyl Lewis X
SPR	Surface Plasmon Resonance
STD-NMR	Saturation Transfer Difference-NMR
STD-AF	Saturation Transfer Difference-Amplification Factor
tar-MD	time-averaged restrained Molecular Dynamics
TCR	T-Cell Receptor
TLR	Toll-Like Receptor

TOCSY	^{1}H-^{1}H TOtal Correlation SpectroscopY
Tris	Tris(hydroxymethyl)aminomethane
tr-NOESY	transferred-NOESY
TROESY	Transverse Rotating-frame Overhauser Effect SpectroscopY

Index

Chapter 1

Introduction and objectives

1.1 Carbohydrates: complex macromolecules

Carbohydrates are the most abundant type of biomacromolecule existing in Nature, either alone or forming glycoconjugates with proteins (proteoglycans or glycoproteins) and lipids (glycolipids or lipopolysaccharides), and thus they are available in large quantities from natural sources. They function as energy storage (starch, glycogen), starting material in biosynthesis, structural constituents of plants (cellulose), and they are also the major components of shells, insects or crabs (chitin), bacterial cell walls (lipopolysaccharides), or virus capsids (HIV). Isolation, purification and chemical modification of carbohydrates have been areas of great interest and exploitation during the last decades.

Most of the first studies on carbohydrates focused on plant polysaccharides (e.g. cellulose, starch, pectins) due to their wide range of applications. More recently, the role of carbohydrates in biological events was recognized[1], and glycobiology emerged as a new and challenging research area at the interface of biology and chemistry. In this context, carbohydrate-mediated recognition events are of key importance in biological phenomena, playing a pivotal role to the study of protein-carbohydrate interactions. Actually, the binding protein partners of carbohydrates encompass a wide variety of macromolecules involved in functions such as recognition, biosynthesis, modification, hydrolysis, and so on.

A relevant part within the research field of glycobiology encloses the determination of the structures and functions of complex sugars (glycans). This has become a critical facet of postgenome science, proteomics in particular, since many proteins are post-translationally modified by glycosylation, and these modifications alter and regulate biological activities. Thus, glycans represent a major class of post-translational modifications that dramatically enhance the functional diversity of proteins (**figure 1**). During the last years, the scientific standpoint of glycobiologists has increasingly moved towards the concept of *glycome*, i.e., the complete set of glycan structures expressed by specific cells, tissues or organisms. This has in turn led to a need for analysis of larger numbers of glycan structures, which has accelerated the development of technologies with high-throughput potential[2]. The emerging *omics* domain of *glycomics* has dropped behind that of genomics and proteomics, mainly because of the inherent difficulties in analysing glycan structure and function[3].

The term *glycomics* is formed by the prefix *glyco-*, which means sweetness or sugar, followed by *-omics* (i.e., field of study) to be consistent with the naming convention established by genomics (which deals with genes) and proteomics (dealing with proteins). The definition of glycomics has evolved to cover a range of scientific disciplines that are applied to study the structure and function of carbohydrates (sugars) in biological systems.

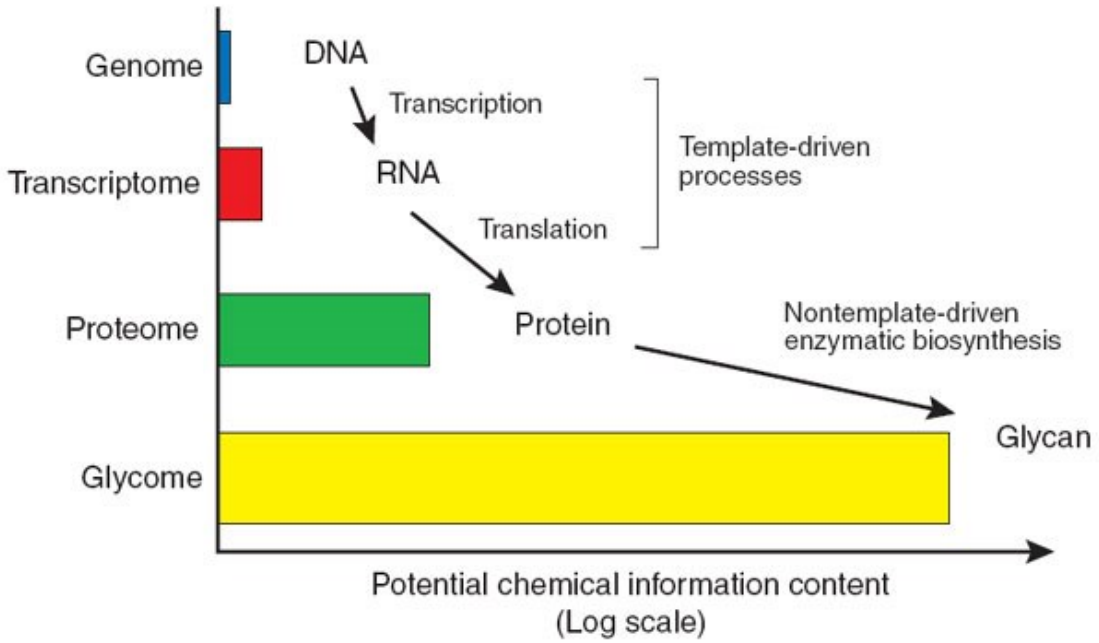

Figure 1. Glycome enhancement of the molecular and functional diversity of the proteome. Protein expression is based on a genetically encoded template, but post-translational modifications of proteins dramatically promote their functional diversity. The glycome represents the main class of posttranslational modifications, providing biological access to vast information space at minimum genetic cost. *Source: Turnbull and Field 2007*[2].

The nine common sugars found in mammalian cells (**figure 2**, top) can be combined in a myriad number of ways to form complex carbohydrate structures. The glycan collection (glycome) of a given cell or organism is thus many orders of magnitude more complex than the genome or the proteome. Thanks to the rapid development of enabling technologies such as high-throughput mass spectroscopy, glycan microarrays and carbohydrate chemistry, deciphering the complexity resulting from this diversity is increasingly possible (**figure 2**, bottom). In this regard, bioinformatics is a fundamental technology of growing importance in managing and integrating the diverse data sets from the different technologies (**figure 2**, bottom).

The developing field of glycomics is earning its place alongside other established "omics" fields such as genomics and proteomics, as it was anticipated by experts in the field:

"We envisage that the collective enterprise of glycomics over the next decade will begin the process of decoding the glycome, thereby yielding many new insights into its myriad functions and producing diverse advances in the biomedical arena". Nature Chemical Biology (2007) **3**; 74-77.

"The knowledge gained from glycomics will be as important as a basis for the pharmaceutical industry as that discovered in the field of genomics and proteomics during the last 30 years". Chem. Eur. J. (2005) **11**; 3194 – 3206.

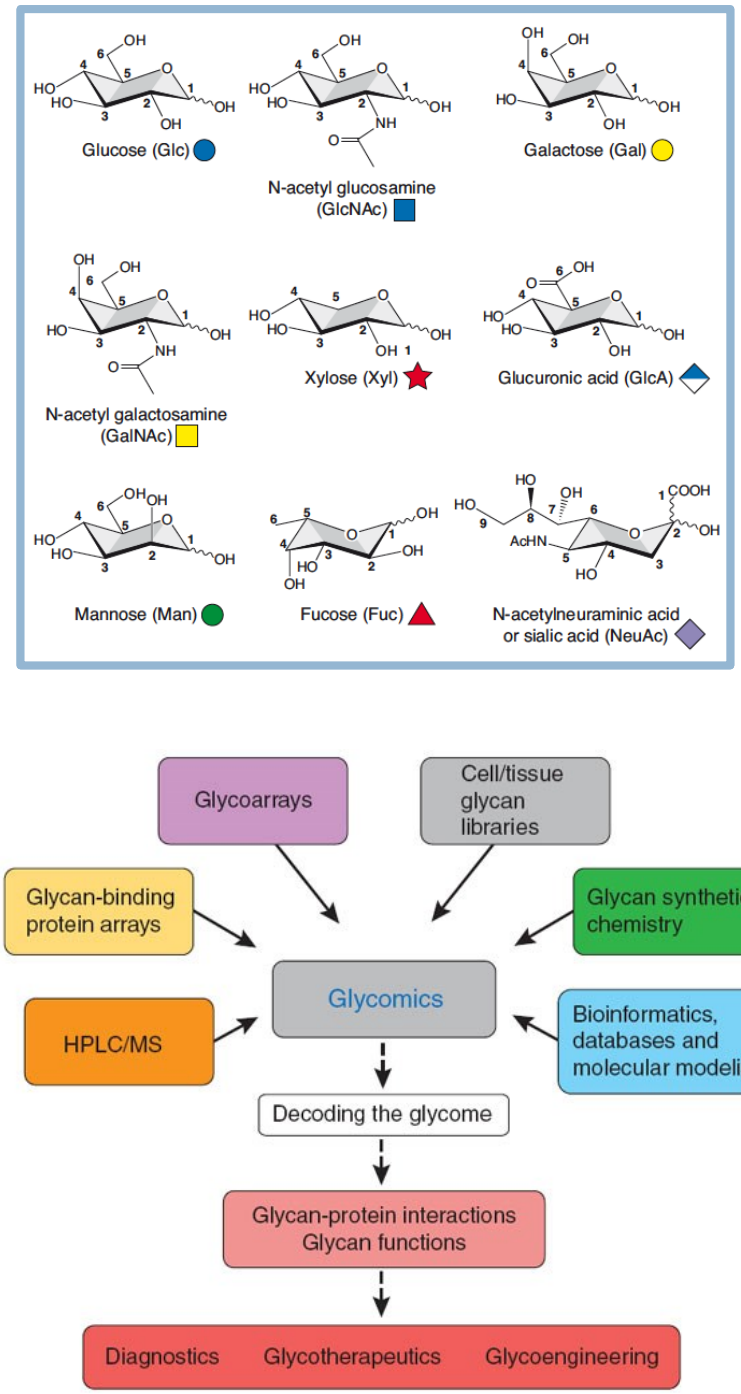

Figure 2. (Top) The nine common sugar "letters" of mammalian glycomics. (Bottom) Combinations of cutting-edge technologies that aid in deciphering the glycocode, leading to new insights and biomedical applications. Currently, they are being exploited to endeavour large-scale analyses of the structure-function relationships of the glycome. *Sources: Weiss and Lyer 2007[4] (top), Turnbull and Field 2007[2] (bottom).*

Many exciting applications of glycomics approaches have become evident, including diagnostics, new routes to glycotherapeutics and defined recombinant protein drugs.

1.1.1 Chemical structure

In contrast to other important biomolecules such as proteins and nucleic acids, carbohydrates can form branched structures by substitution of one or several hydroxyl groups, making them extremely complex and heterogeneous. The structure of an oligosaccharide is determined by the monosaccharide sequence, the glycosidic linkage sites, the stereochemistry of the glycosidic linkages (α or β) and the degree and type of substitution of hydroxyl groups (such as O-methylation or O-sulphation)[5]. Also, it is important to emphasize that water plays a central role in defining oligosaccharide conformation by influencing the geometry of the glycosidic linkages[6].

Carbohydrates, or less commonly named "hydrated carbons", are compounds that often have the empirical formula $C_n(H_2O)_n$. A monosaccharide is an aldehyde or a ketone containing at least two additional hydroxyl groups. While two monosaccharides connected by a glycosidic bond are named a disaccharide, carbohydrates with three to ten units are typically called oligosaccharides, and larger structures are known as polysaccharides. Carbohydrates are chiral and optically active, and the majority of the naturally occurring ones present the D configuration[7]. Carbohydrates with five or six carbon atoms, pentoses and hexoses, respectively, can form intramolecular hemiacetals between the carbonyl group and the hydroxyl group on carbon 4 or 5. The resulting rings are called hexopyranoses (six-membered rings) or pentafuranoses (five-membered rings). The cyclic and the acyclic forms exist in equilibrium, with the hemiacetals (rings) being the most abundant forms. Upon hemiacetal formation, the former carbonyl carbon becomes a new stereocenter, called the anomeric center, with the hydroxyl group either equatorial or axial (β and α configuration in D-glucose, respectively). The equilibrium between the cyclic and acyclic forms allows the two anomeric forms to interconvert, a process called mutarotation[8].

Hexopyranose rings can exhibit different conformations. **Figure 3** illustrates the conformational itinerary among the main canonical conformations (or puckers) of the hexopyranose ring. For the majority of carbohydrates, the energetically most favoured ring conformation is the chair conformation, which exists in two distinct forms, the $^4C_1$ and the $^1C_4$ chairs (C=chair), where the numbers refer to the atoms above (superscript) and below (subscript) a reference plane. The $^4C_1$ conformation is generally favoured for D-sugars due to fewer non-bonding interactions between the ring substituents. For example, the $^4C_1$ conformation is the only one observed by NMR spectroscopy for D-glucose[9]. Also, hexopyranoses can exist in boat (B), half chair (H) and skew-boat (S) conformations. For instance, L-iduronic acid presents three low-energy conformations in equilibrium: $^1C_4$, ^2S_O and $^4C_1$[10].

Glucopyranose exists in a 64:36 mixture between the α and β forms in aqueous solutions. Just considering steric hindrance, it would be expected for electronegative substituents to prefer the equatorial disposition; however, the opposite occurs. This higher than expected occurrence of the axial form is due to the endo-anomeric effect[11], which identifies the preference of the electronegative

substituent at the anomeric carbon (C1 in aldoses) for an axial configuration (α-anomer), rather than for the equatorial orientation (β-anomer). This preference has its basis in the electronic structure of the O5–C1–O1 atomic sequence. The widely accepted justification for the anomeric effect is that it originates from n→σ* hyperconjugation between the lone pair of electrons placed at the non-bonding orbital (n) of the ring oxygen atom (O5) and the antibonding σ* orbital of the adjacent C1–O1 bond , although there are other interpretations (**figure 4**)[12]. For pyranoses, stabilizing hyperconjugation is maximized in the α-anomer[13]. On the other hand, the exo-anomeric effect manifests the preference of the Og-Cx glycosidic bond (Og: glycosidic oxygen; Cx: adjacent carbon to the right) to adopt a *gauche* orientation with respect to the C1-O5 bond (C1: anomeric carbon; O5: oxygen ring), thus it contains itself the the rotameric preference about the C1–Og bond (known as Φ torsion). The exo-anomeric effect arises again from hyperconjugation within the O5–C1–O sequence, but this time it is between the lone pair of electrons placed at the non-bonding orbital (np) of the Og oxygen and the antibonding σ* orbital of the O5–C1 bond (**figure 4**).

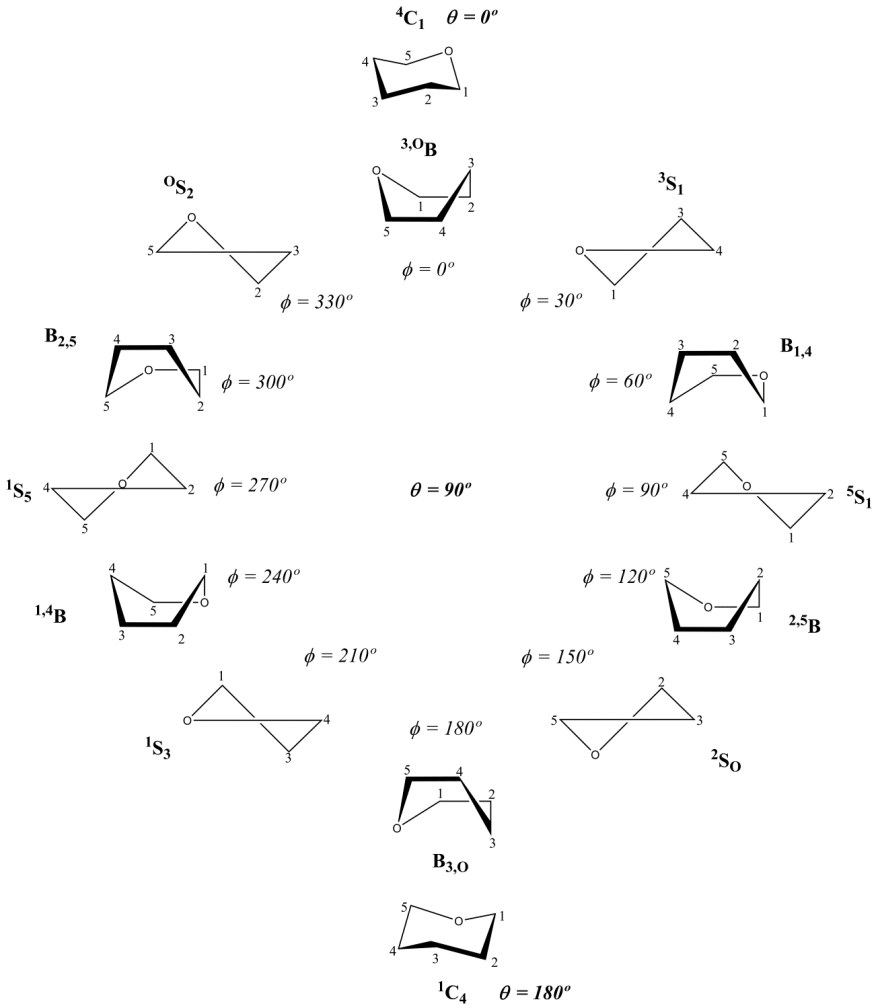

Figure 3. Diagram of the pseudorotational itinerary of the pyranose ring according to Jeffrey and Yates[14], based on the Cremer-Pople ring-puckering coordinates θ and ϕ. The polar $^4C_1$ (θ=0º) and $^1C_4$ (θ=180º) chairs, together with the 12 equatorial puckers (θ=90º) are shown. The envelope (E) and half-chair (H) conformations are not pictured.

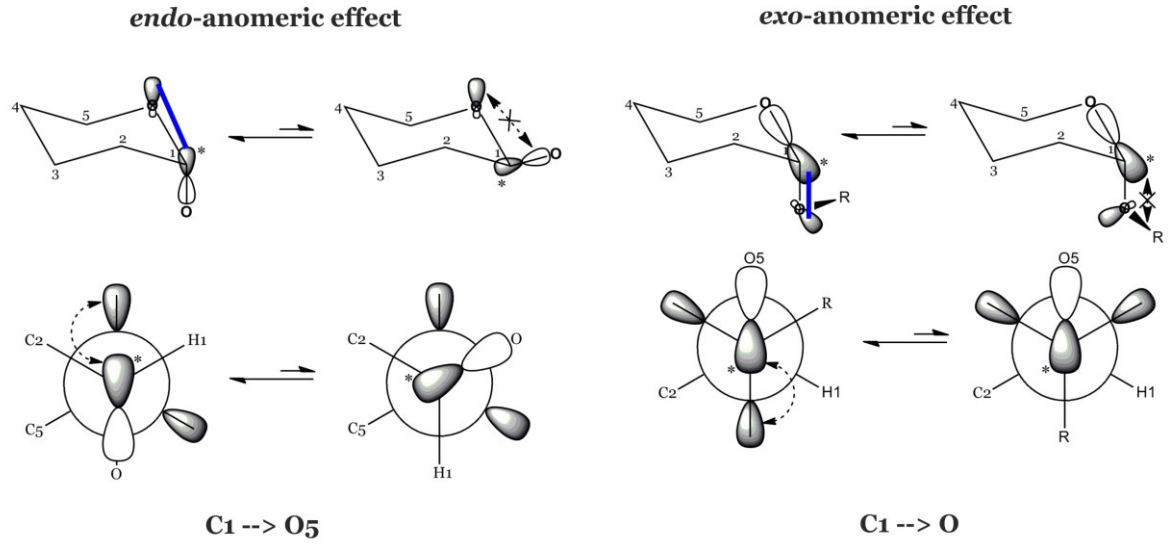

Figure 4. Schematic representation of the stereoelectronic genesis of the endo- and exo-anomeric effect.

The diversity of carbohydrate structures results from the broad range of monomers (>100) of which they are composed and the different ways in which these monomers are joined (glycosidic bonds). Thus, even a small number of monosaccharide units can provide a large number of different oligosaccharides (also referred to as glycans), including branched structures, a unique feature among biomolecules. For example, the number of all possible linear and branched isomers of a hexasaccharide exceeds 1012[15].

Providing a structural basis for the multitude of biological roles played by carbohydrates, it is imperative to accurately determine their spatial (conformation) and dynamic properties in aqueous solution (the importance of dynamics in structural biology was highlighted by the prediction that ~25% of mammalian proteins are fully disordered[16]. This goal promises to enable structure-based design of new medicines and materials, but its realization remains challenging due to experimental and computational difficulties of probing carbohydrate motions, which occur over a broad range of time scales. For instance, glycosidic linkages liberate on nanosecond time scales while pyranose ring conformational exchange (or puckering) and anomerization are microsecond[17] and millisecond time scale phenomena, respectively.

The recognition mechanism of carbohydrates depends on their conformation, which is determined by 1) the sequence of the monosaccharides in the glycan, 2) the anomeric centres (i.e., α or β), 3) the linkage positions (i.e., 1–3, 1–4, 1-6), and 4) the chemical modifications to the core glycan (i.e., sulphation, phosphorylation, methylation, acetylation, etc.). The strength of this interaction is also determined by the carbohydrate conformation and orientation with respect to the binding site.

Carbohydrates and their derivatives possess many hydroxyl groups and thus a large number of rotatable bonds. Due to the many hydroxyl groups, these compounds are usually highly water soluble,

and their logP is often negative (more soluble in the hydrophilic than the hydrophobic phase). The surface of carbohydrates and their derivatives is composed of hydrophobic and hydrophilic patches formed by nonpolar aliphatic protons and polar hydroxyl groups (**figure 5**). This leads to anisotropic solvent densities around carbohydrate molecules. In aqueous environments, favourable interactions of water molecules with the hydrophilic patches result from electrostatic interactions and hydrogen bonding, whereas the interaction of water with hydrophobic surface patches is unfavourable. The equilibrium between hydrophobic and hydrophilic patches forms the basis for properties such as carbohydrate solubility in water and molecular recognition.

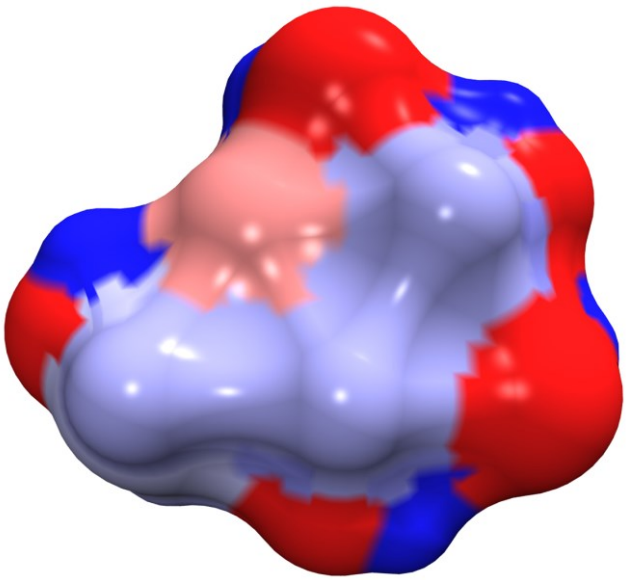

Figure 5. Electrostatic potential surface of β-D-glucose. Intense blue and intense red colours indicate positive and negative charge, respectively, of the hydrophilic patches. In pale red and pale blue colours appear patches with hydrophobic character.

Another essential feature of carbohydrates is their conformational flexibility[18]. Compared to drug-like molecules, carbohydrates are typically much more flexible. In addition, unlike proteins and oligonucleotides (which are built by linear assembly of residues), glycans are often branched structures. In analogy with proteins, the relative orientations of two consecutive monosaccharide units in a disaccharide moiety are expressed in terms of the glycosidic linkage torsional angles Φ and Ψ around the glycosidic bonds (**figure 6**), which in the present thesis are defined as Φ ≡ H1–C1–Og–Cx and Ψ ≡ C1–Og–Cx–Hx (H1: hydrogen atom attached to the anomeric carbon; C1: anomeric carbon; Og: glycosidic oxygen; Cx: carbon belonging to the following residue, linked to the glycosidic oxygen). It should be noted, however, that a different definition is usually found in literature, i.e., Φ ≡ O5–C1–Og–Cx and Ψ ≡ C1–Og–Cx–C(x+1) (O5: oxygen ring). Thus, the overall conformation of the oligosaccharide chain is determined by the set of "preferred" Φ and Ψ dihedral angles within it (except for 1-6 linkages; read below). In addition, the degree of flexibility of the chain manifests on the narrower or wider distribution frequency curves of these torsions (conformational space sampled).

Additional degrees of freedom are associated with hydroxyl group rotations and rotation around the C5–C6 ω-angle, when present (**figure 6**). Unexpected rotameric preferences are also seen for this exocyclic hydroxymethyl group in hexopyranoses. In contrast to expectations based solely on steric effects, this single bond displays a strong preference for rotamers in which O6 and O5 are in a *gauche* orientation (**figure 7**). In contrast to the anomeric effect, the *gauche* effect is principally caused by solvation and electrostatic interactions [28], rather than steric or stereoelectronic effects. Rotamer preferences for this bond can profoundly impact the conformational properties of oligosaccharides containing 1–6 linkages, which are common in mammalian and bacterial cell-surface glycans [85–87]. The three *gauche-gauche* (*gg*), *gauche-trans* (*gt*) and *trans-gauche* (*tg*) rotamers around the ω torsion can be identified by NMR from the set of NOEs observed and the $^{3}J_{HH}$ values measured (**figure 7**). Thus, while the *gg* rotamer produces two NOE signals (H5-H6*proR* and H5-H6*proS*), only one is observed for the other two (*gt*, H5-H6*proS*; *tg*, H5-H6*proR*). In addition, the *gg* rotamer presents two small $^{3}J_{HH}$ couplings, whereas the *gt* conformer gives rise to a small and large $^{3}J_{HH}$ for the H5-H6proS and H5-H6proR, respectively, the opposite applies for the *tg* rotameric state (**figure 7**).

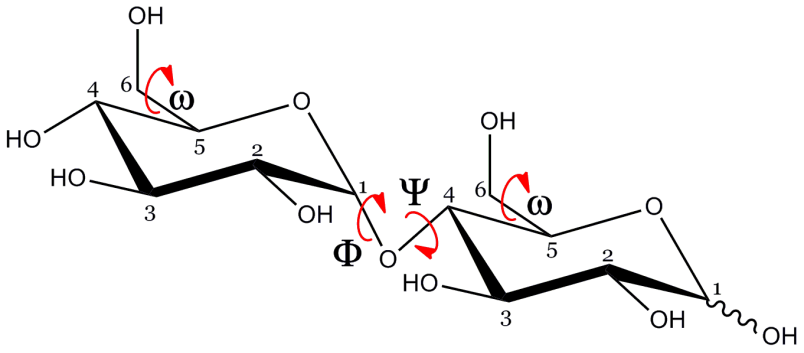

Figure 6. Picture showing the Φ, Ψ and ω dihedral angles that define the overall conformation of oligosaccharides. Maltose has been used as a template.

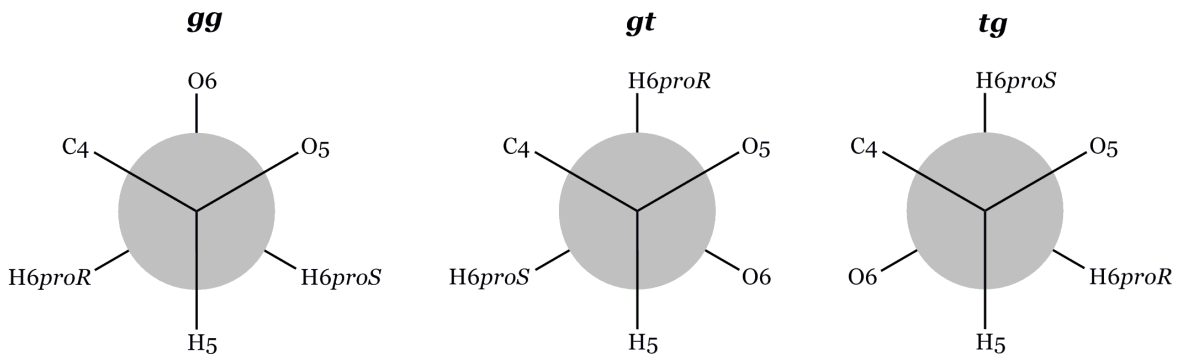

Figure 7. Newman projections of the gauche–gauche (*gg*), gauche–trans (*gt*) and trans–gauche (*tg*) rotameric conformers of the ω dihedral angle. The *gg*, *gt* and *tg* conformers are defined relative to the O5–C5–C6–O6 and C4–C5–C6–O6 torsions.

The energetically favourable conformations of a carbohydrate dimer may be identified on energy plots called (Φ, Ψ) maps (**figure 8**), which are somewhat similar to the Ramachandran plots used to visualize the backbone dihedral angles of the constituent amino acids in proteins. For complex carbohydrates, as they may adopt conformations belonging to different minima, it is necessary to thoroughly sample the conformational space. However, while this may be feasible for glycosidic bonds, the number of degrees of freedom increases rapidly when, in addition to this, we take into account the orientation of the hydroxyl groups, and, as a consequence, the computational cost augments.

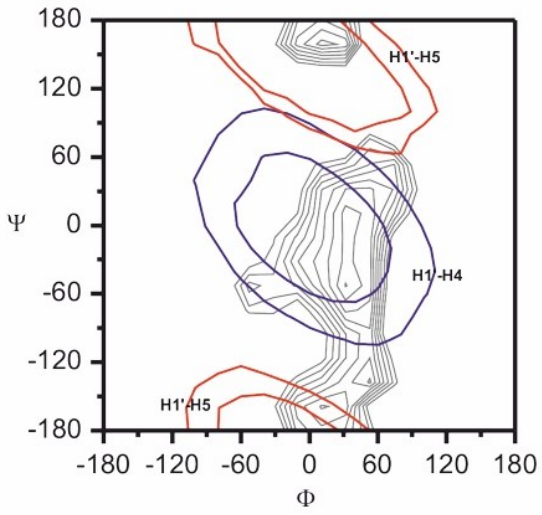

Figure 8. Example of relaxed (Φ, Ψ) map for the L-IdoA-α(1-4)-D-GlcNS6S glycosidic linkage. Energy levels are shown each 0.5 kcal/mol. Isodistance curves corresponding to the H1'-H4 and H1'-H5 distances appear for $r_{ij} \leq 2.5$ and $r_{ij} \leq 3$Å.

Carbohydrates can also be found as different types of glycoconjugates. In nature, glycoconjugates are ubiquitously found as part of proteins, bacteria and viruses. The extent to which a glycoconjugate is glycosylated can vary largely, with a carbohydrate content from 1% to 99%. The major classes of glycoconjugates found in nature are glycoproteins, proteoglycans, glycolipids and lipopolysaccharides (**figure 9**).

A glycoprotein is a protein which has one or several oligosaccharides covalently linked to it. Glycoproteins are classified according to the linkage type of the oligosaccharide to the protein. The linkage can be to either an asparagine residue, called N-linked glycoproteins or to a serine or a threonine residue, called O-linked glycoproteins (**figure 10**). When attached to the cell membrane via an ethanolamine phosphate group they are called GPI anchors. The properties of a protein change with different glycosylation patterns. Often, glycoproteins are embedded in the outer membrane of cells and work as antennas sticking out of the surface. They glycan parts play a variety of roles in processes such as cell adhesion and the three-dimensional folding of proteins[5, 19].

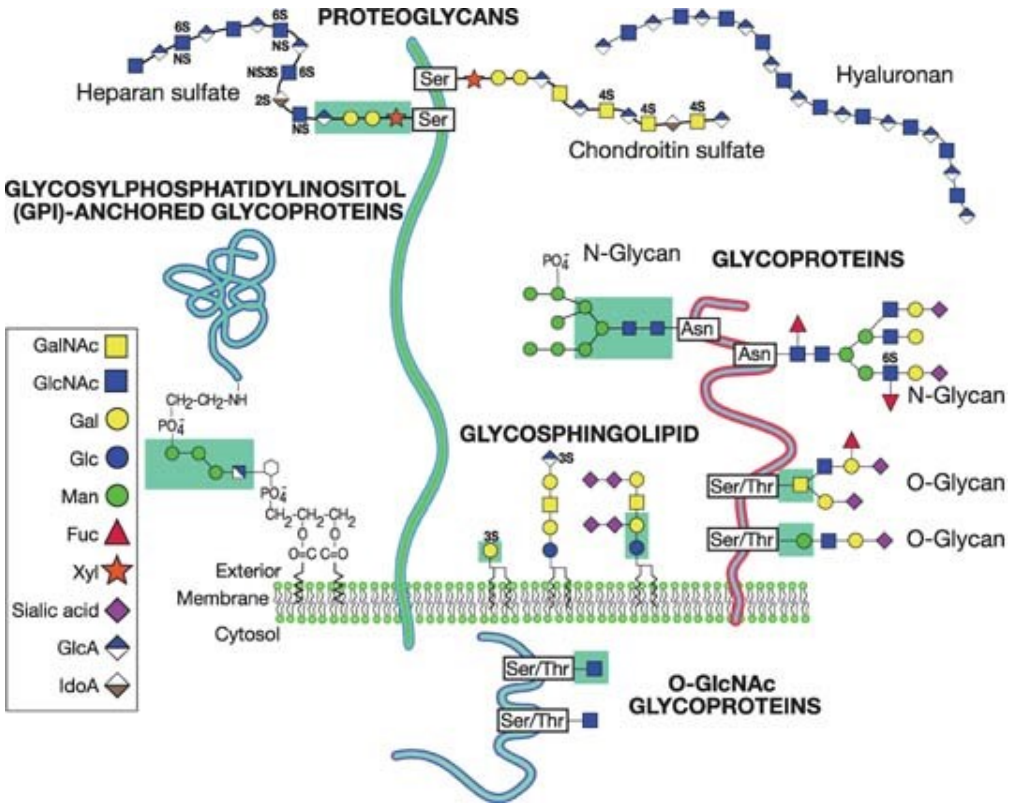

Figure 9. Known glycan-protein and glycan-lipid linkages in nature. *Source: Hidekatsu Iha and Masao Yamada, 2013*[20].

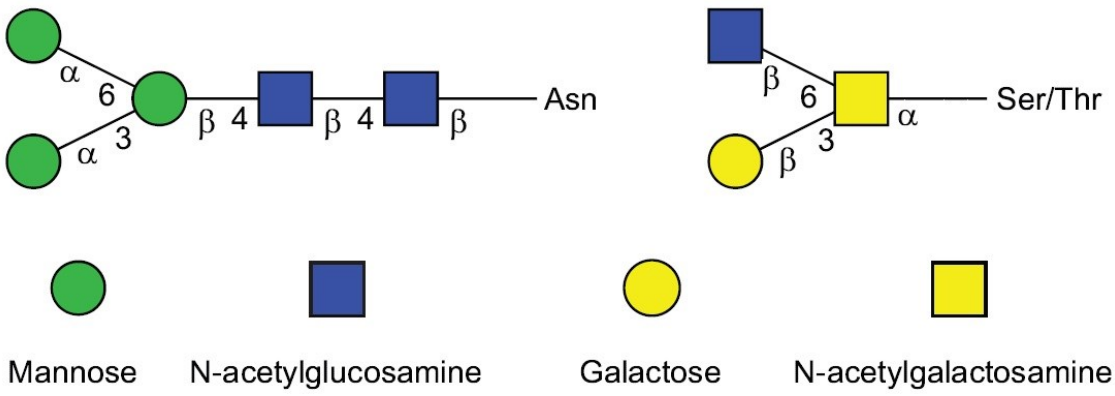

Figure 10. (*Left*) Schematic picture in CFG (Consortium for functional genomics) notation of the uniform core pentasaccharide of N-glycans, consisting of two D-GlcNAc and three mannose moieties. (*Right*) Example of O-glycan (α-D-GalNAc residue is bound to the peptide chain).

Proteoglycans are a class of proteins placed on the surface of cells, making up a major part of the extracellular matrix, which provides structural support. They are also involved in binding cations and water, and in regulating the movement of molecules through the matrix. Proteoglycans are glycoproteins consisting of linear polysaccharides covalently attached along the length of a single core

protein. At least one of the polysaccharide chains of a proteoglycan must be a glycosaminoglycan (GAG). Proteoglycans vary greatly in size depending on the molecular weight of the core protein and the number and length of the polysaccharide chains, and can contain as much as 95% of carbohydrates.

Glycolipid is the term used for any compound containing one or more monosaccharide residues bound by a glycosidic linkage to a hydrophobic moiety such as an acylglycerol, a sphingoid, a ceramide (*N*-acylsphingoid) or a prenyl phosphate.

Lipopolysaccharides (LPS) are complex molecules that act as endotoxic O-antigens and are found in the cell walls of Gram-negative bacteria (S-lipopolysaccharides) and in fungus. A lipid part (Lipid A) forms a complex with a core polysaccharide part through a glycosidic linkage (**figure 11**). The core part is linked to a third external region of a highly immunogenic and variable O-chain polysaccharide or O-antigen made up of repeating oligosaccharide units. The latter region of the LPS molecule is responsible for bacterial serological strain specificity.

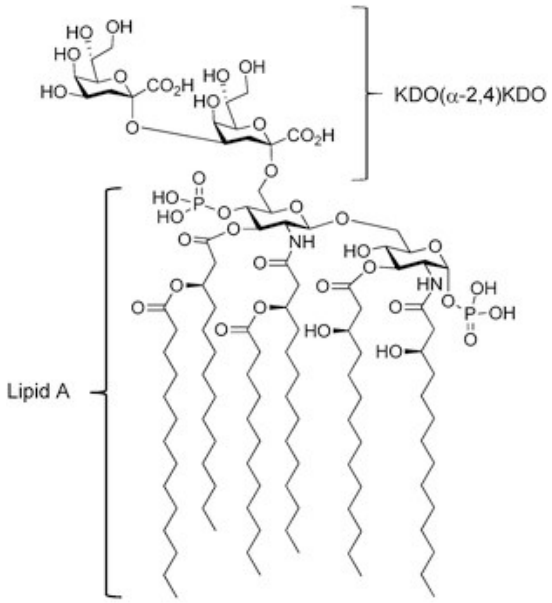

Figure 11. Example of lipopolysaccharides (LPS) chemical structure, from *E. coli*.

1.1.2 Synthetic carbohydrates

Using synthetic carbohydrates instead of the naturally occurring ones (biosynthesized) is essential for the appropriate characterization and interpretation of their molecular and binding properties. The reason is the non-template driven nature of their biosynthesis, which gives rise to a wide heterogeneity in both the chain length and the substitution pattern. This highly complicates or impedes to establish structure-function relationships.

Synthetic carbohydrates are used in many different ways. For instance, to evaluate their three-dimensional structure, for testing the biological activity and to map carbohydrate binding sites in proteins[5]. They are also crucial in the development of new carbohydrate-based drugs and vaccines as well as drug delivery systems.

To enable and facilitate analysis with NMR spectroscopy it is sometimes necessary to synthesize isotopically labelled carbohydrates. Oligosaccharides are traditionally synthesized in solution from the linking of a glycosyl donor with a glycosyl acceptor. The preparation of monosaccharide building blocks often involves tedious protecting group chemistry, although, more recently, chemoenzymatic methods and automated solid-phase synthesis have facilitated the synthesis of oligosaccharides[21]. To make carbohydrates more useful as drugs they can be prepared as pseudosugars where either the ring oxygen is replaced with a methylene group or the glycosidic oxygen is replaced by nitrogen, carbon or sulphur making them hydrolytically stable[22]. Glycodendrimers are another example of carbohydrate-containing molecules. They are branched symmetrical polymers, often synthesized by click-chemistry[23]. Dendrimers have found many different uses, such as in host-guest chemistry, catalysis, in materials science and as drug delivery systems[24].

1.1.3 Biological importance

In nature, carbohydrates are present, among others, in the cytoplasm and in the extracellular matrix of cells (ECM). The ECM is a complex network of two main classes of macromolecules: (1) polysaccharide chains called *glycosaminoglycans* (GAGs; see next section), which are usually found covalently linked to proteins in the form of proteoglycans (with the exception of hyaluronic acid; see **figure 12**), and (2) fibrous proteins, including collagen, elastin, fibronectin and laminin, which have structural and adhesive functions, secreted locally by cells and assembled into an organized mesh. The proteoglycan molecules in connective tissue form a highly hydrated, gel-like substance in which the fibrous proteins are embedded. The polysaccharide gel resists compressive forces on the matrix while permitting the rapid diffusion of nutrients, metabolites, and hormones between the blood and the tissue cells. The collagen fibres both strengthen and help organize the matrix, while elastin fibres give it resilience. Briefly, the ECM accomplishes multiple functions such as:

1. Filling the space between cells, binding cells and tissue together.
2. Providing a lattice through which cells can move.
3. Filtration.
4. Support and elasticity.
5. Influences cell growth and differentiation (many cells interact with their ECM).
6. Serves as a structural element in tissues and also influences their development and physiology.

Carbohydrates are involved in numerous biological functions, such as recognition in axonal growth or path-finding[25], blood anticoagulation[26], cell-cell recognition[27], antibody-antigen interactions,[28]

structure factors in extra-cellular matrices[29], and post- or co-translational modifications of polypeptides[30]. Correct glycosylation patterns are essential for normal cell and organism function, and aberrant glycosylation is associated with numerous human diseases[31].

Polysaccharides make up a substantial part of bacterial cell-walls, the greatest part being peptidoglycans giving the membrane mechanical strength. Other polysaccharides, such as lipopolysaccharides, capsular polysaccharides and exopolysaccharides, are to a large extent covering the cell-walls of bacteria. Microbial surface polysaccharides play an important role in bacteria-host interactions.

In Gram-negative bacteria the lipid bilayer contains lipopolysaccharides (LPS's). LPS's consist of three regions, the lipid-A, the core region and the O-antigen. When the lipid-A is released from the membrane it is toxic and harmful to mammals. The O-antigen consists of repeating units with 2-8 carbohydrate moieties, and is very diverse. The O-antigen is the part of the bacteria recognized by the immune system[5]. Bacteria can also produce capsular polysaccharides, which is an extracellular coat surrounding the bacteria associated with virulence[19]. An additional kind of polysaccharide produced by bacteria is exopolysaccharides (EPS). They are high molecular weight polymers generally composed of repeating units of D-glucose, D-mannose, D-galactose, L-fucose, L-rhamnose, and D-glucuronic acid[32]. Exopolysaccharides are used in a number of industrial products, e.g., as food additives and in medical applications[33].

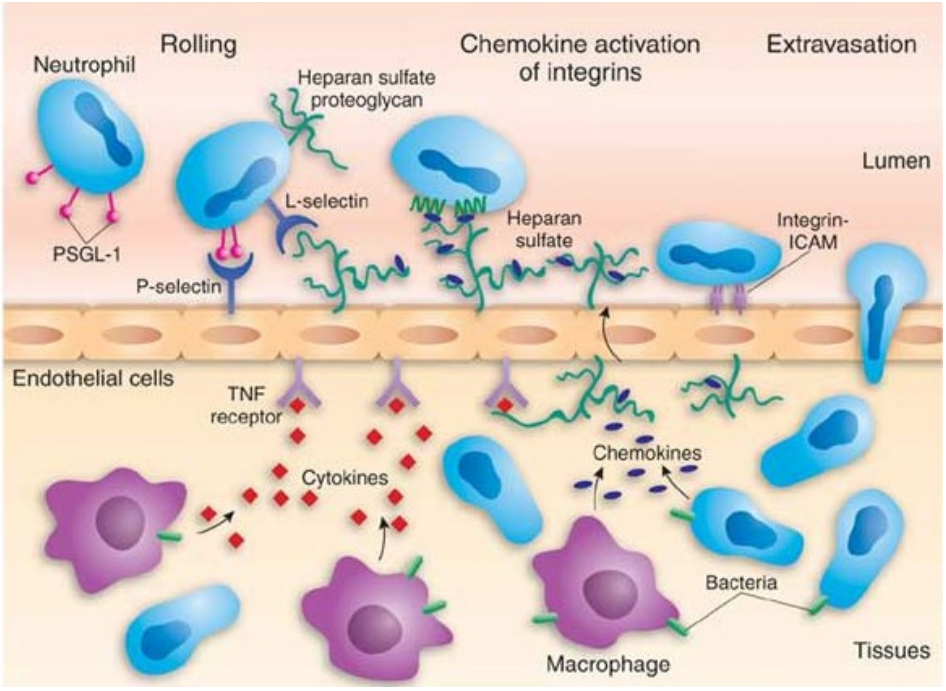

Figure 12. Multiple roles of heparan sulphate (HS) in leukocyte entry into sites of inflammation. HS on activated endothelium binds L-selectin on leukocytes (granulocytes, monocytes and lymphocytes) during the rolling phase of leukocytes over the endothelium. Endothelial HS binds and presents chemokines to chemokine receptors on leukocytes, which leads to activation of leukocytes and movement of leukocytes towards the site of inflammation, while HS is also involved in the transport of chemokines across the endothelial cell barrier.

1.2 Glycosaminoglycans (GAGs)

Glycosaminoglycans (GAGs) are long unbranched polysaccharide chains consisting of repeating disaccharide units. By definition, one of the two sugars in this repeating unit is a D-hexosamine residue (i.e., D-glucosamine, D-GlcN, or D-galactosamine, D-GalN), giving glycosaminoglycans their name. The other monosaccharide is an uronic acid (D-glucuronic acid, D-GlcA, or L-iduronic acid, L-IdoA). Thus, GAGs are divided into different classes according to the nature of their repeating unit (**table 1**): hyaluronan (HA), keratan sulphate (KS), chondrotin sulphate (CS), heparan sulphate (HS), heparin (HEP) and dermatan sulphate (DS)[19].

HS and CS are synthesized in the Golgi apparatus, where the individual GAG chains are O-linked to a core protein, forming a large proteoglycan (PG)[34]. Keratan sulphate, on the other hand, can be either N-linked or O-linked to the core protein of the PG[35]. HA is not synthesized in the Golgi from the core protein but rather by an integral plasma membrane synthase, which secretes the nascent chain immediately[36]. The biosynthesis of GAGs is a complex non-template-driven process involving several enzymes that assemble the GAG polymer and then sulphate them at specific positions. The GAG attachment sites on the core protein of the PG have a consensus Ser-Gly/Ala-X-Gly motif. Importantly, this versatility of GAGs biosynthesis yields a large number of possible substitution patterns that make them into highly dense information carriers.

The study of molecular recognition processes between carbohydrates and protein receptors has attracted considerable attention during the past years[37]. Among those processes, the interaction of glycosaminoglycans (GAG) and signalling proteins (**figure 12**), due to its biological relevance and large number of cases, has attracted the attention of many research groups and structural details of those interactions have been extensively studied[38]. There are many examples where the specificity of the interaction between the GAG and signalling proteins relies on the substitution pattern[39]. Among them it should be mentioned the interaction with the fibroblast growth factor (FGF) family[40], chemokines[41] and cytokines, antithrombin-III (AT-III)[42], lipases, apolipoproteins and ECM and plasma proteins[43].

Table 1. Types of GAGs and their disaccharide building blocks[44].*

Category	Disaccharide repeating unit	Chemical substitutions
Heparin (HEP)	L-IdoA2X-α(1-4)-D-GlcNY,3X,6X-α(1-4)	X: SO_3^-, Y: Ac or SO_3^-
Heparan sulphate (HS)	D-GlcA-β(1-4)-D-GlcNY,3X,6X-α(1-4)	X: SO_3^-, Y: Ac or SO_3^-
Chondroitin sulphate (CS)	D-GlcA2X-β(1-3)-D-GalNAc,4X,6X- β(1-4)	X: SO_3^-
Dermatan sulphate (DS)	L-IdoA2X-α(1-3)-D-GalNAc,4X,6X- β(1-4)	X: SO_3^-
Keratan sulphate (KS)	D-Gal6X-β(1-4)-D-GlcNAc,6X-β(1-3)	X: SO_3^-
Hyaluronic acid (HA)	D-GlcA-β(1-3)-D-GlcNAc-β(1-4)	None

* Note that the abbreviations stand for: L-iduronic acid, L-IdoA; D-glucuronic acid, D-GlcA; D-glucosamine, D-GlcN; D-galactosamine, D-GalN; D-galactose, D-Gal. The acetyl ($COCH_3$) and sulphate (OSO_3^-) groups are abbreviated using *Ac* and *S*, respectively. Also note that for HEP/HS and CS/DS pairs, only the major disaccharide repeating unit has been indicated, although the disaccharide sequence of its partner is also present.

1.2.1 Heparin and the singular role of L-iduronic acid

Heparin is a linear polymer consisting of disaccharide repeating units of 1→4-linked hexopyranosyluronic acid and 2-amino-2-deoxyglucopyranose (glucosamine) residues[45]. The uronic acid residues typically consist of 90% L-idopyranosyluronic acid (L-iduronic acid, L-IdoA) and 10% D-glucopyranosyluronic acid (D-glucuronic acid, D-GlcA)[1]. Heparin has the highest negative charge density of any known biological macromolecule. This is the result of its high content of negatively charged sulphate (OSO_3^-) and carboxylate (COO^-) groups[46]. Indeed, the average heparin disaccharide contains 2.7 sulphate groups. The most common structure occurring in heparin is the trisulphated disaccharide L-IdoA2S-α(1-4)-D-GlcNS,6S shown in **figure 13**. However, other substitution patterns also participate, leading to the microheterogeneity of heparin. For example, the amino group of the glucosamine residue may be substituted with an acetyl or sulphate group or unsubstituted. Also, the 3- and 6- positions of the glucosamine residues can either be substituted with an O-sulphate group or unsubstituted. The uronic acid, which can either be L-iduronic or D-glucuronic acid, may also contain a 2-O-sulphate group. Glycosaminoglycan heparin has a molecular weight range of 5-40 kDa, with an average molecular weight of about 15 kDa and an average negative charge of approximately -75. The

[1] The terms "hexopyranosyluronic" (idopyranosyluronic or glucopyranosyluronic) and "uronic" (iduronic or glucuronic) will be interchangeably used in the present Doctoral Thesis. The first term is, indeed, more restrictive as it exclusively refers to the α and β anomers in the ring form, whereas the second term allude to both the cyclic and acyclic sugars. Since GAGs are polymers formed by the repetition of cyclic sugars (rings), "hexopyranosyluronic" is indeed the most precise term.

large structural variability of heparin, coming form its high heterogeneity and polydispersity, makes it an extremely challenging molecule to characterize.

The structural complexity of heparin can be considered at several levels. At the proteoglycan (PG) level, different numbers of polysaccharide (or glycosaminoglycan) chains (possibly having different saccharide sequences) can be attached to the various serine residues present in heparin's core protein. During their biosynthesis, heparin chains are attached to a unique core protein, serglycin, found only in mast cells and some hematopoietic cells. Tissue proteases act on this core protein to release peptidoglycan heparin, a small peptide to which a single long polysaccharide chain (100 kDa) is attached. This peptidoglycan is short-lived as it is immediately processed by a β-endoglucuronidase to a number of smaller (about 15 kDa) polysaccharide chains called glycosaminoglycan (GAG) heparin[47]. Most of the chemical and physical properties of heparin are related to GAG structure or sequence, conformation, chain flexibility, molecular weight and charge density.

Heparan sulphate is structurally related to heparin but it is much less substituted with sulphate groups than heparin and has a more varied structure (or sequence). Like heparin, heparan sulphate is a repeating linear copolymer of a uronic acid 1→4 linked to a glucosamine residue[48]. Although D-glucuronic acid predominates in heparan sulphate, it can contain substantial amounts of L-iduronic acid. Heparan sulphate generally contains only about one sulphate group per disaccharide, but individual HS may have a higher content of this group. Also, heparan sulphate chains often contain domains of extended sequences having low or high sulfation[49]. While heparan sulphate contains all of the structural variations found in heparin (and vice versa), the frequency of occurrence of the minor sequence variants is greater than in heparin, making HS structure and sequence much more complex. HS chains are also polydisperse, but are generally longer than heparin chains, having average molecular weight of about 30 kDa ranging from 5 to 50 kDa[50]. Heparan sulphate is biosynthesized, as a proteoglycan, through the same pathway as heparin; however, unlike heparin, the HS GAG chain remains connected to its core protein. Heparan sulphate is ubiquitously distributed on cell surfaces and is also a common component of the extracellular matrix[49, 51]. Two types of core proteins, the syndecans (an integral membrane protein) and the glypicans (a GPI-anchored protein), commonly carry heparan sulphate GAG chains and correspond to the two major families of heparan sulphate PGs[51-52]. The HS chains on these heparan sulphate PGs bind a variety of proteins and mediate various physiologically important processes, including blood coagulation, cell adhesion, lipid metabolism and growth factor regulation[53]. Although structurally similar, heparin and heparan sulphate GAGs can often be distinguished through their different sensitivity towards a family of GAG-degrading microbial enzymes, the heparin lyases[54].

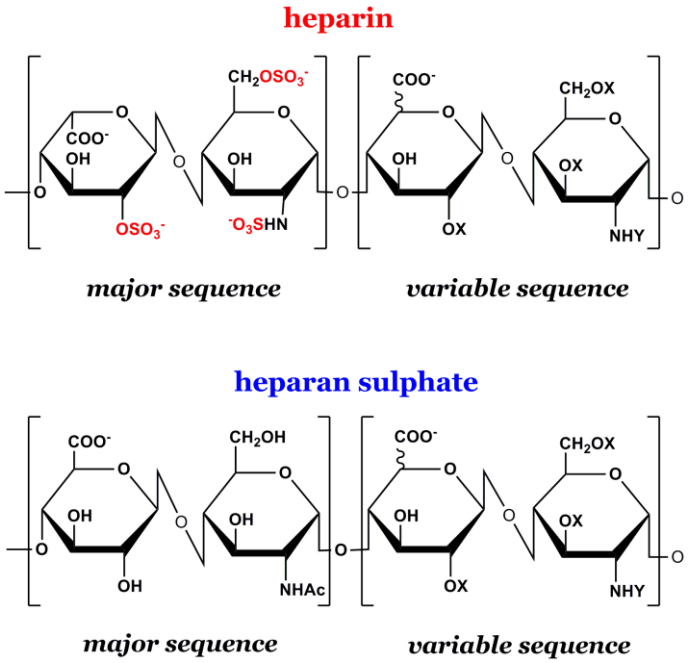

Figure 13. Major and minor disaccharide repeating units in heparin and heparan sulphate (X=H/SO_3^-, Y=Ac/SO_3^-/H). *Adapted from Capila and Linhardt 2002*[55].

Conformation

Experimental structure determination methods such as X-ray crystallography[56], NMR spectroscopy[57], and fluorescence energy-transfer spectroscopy[58] have been applied in studies of carbohydrate conformation, either free or complexed with proteins. While NMR spectroscopy has been extensively used to characterize the dynamics of glycans in solution[59], interglycosidic linkage conformations are notoriously difficult to determine by NMR spectroscopy because of the paucity of Nuclear Overhauser Effects (NOEs)[60], the uncertainties in the Karplus-type equations employed to interpret scalar J-coupling constants[61], and the potential for the linkage to populate multiple rotamer states[62]. Moreover, NMR techniques employed to determine the structural properties of polysaccharides or protein-carbohydrate complexes are limited by molecular weight constraints. Alternatively, X-ray crystallography can be a powerful source of structural information. However, the presence of multiple glycoforms often prevents crystallization of glycoproteins, and the inherent flexibility of oligosaccharides is the presumed reason for the notable absence of X-ray structures for any but the smallest systems.

Theoretical methods, such as Monte Carlo and molecular dynamics (MD) simulations, are employed increasingly to augment the experimental approaches in determining the conformational properties of carbohydrates, and biomolecules in general. The level of interest in applying classical simulations to oligosaccharides arises from experimental limitations and is demonstrated by the numerous force fields and parameter sets that have been derived for carbohydrates[63].

The conformation of heparin has been extensively investigated since its discovery[45a, 64]. Owing to its nature and topology, i.e., a rigid helix with a complete turn every four residues[65], discontinuous interactions with the same side of a protein surface are expected, grouping each three contiguous sulphate groups on opposite sides (**figure 14**)[66]. Additionally, the number and distribution of sulphate groups should play some role in the specificity of the interaction[67]. Another structural key aspect of the heparin or HS structure is that while it is very rigid from the backbone perspective (global conformation), at the same time it is quite flexible at the local level, i.e., when the conformational equilibrium of the iduronate ring is considered.

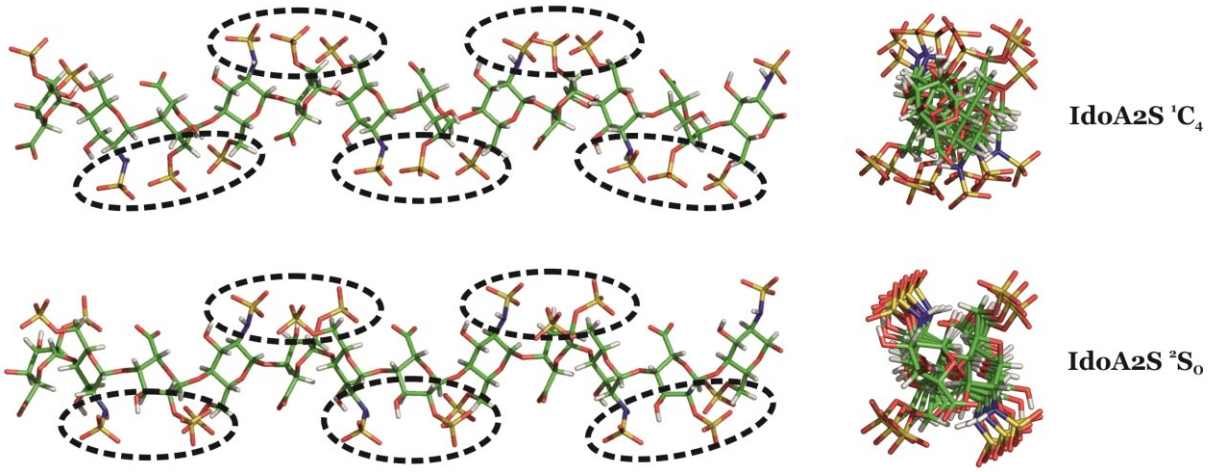

Figure 14. Molecular conformation of heparin determined by NMR and molecular modelling from a dodecasaccharide representative of the regular region of heparin, with all the iduronate residues adopting either the $^1C_4$ (top) or the ^2S_O (bottom) puckering[65]. Each three contiguous sulphate groups are marked within dashed-lined circles.

L-iduronic acid, biosynthesized in the polymeric form through a single epimerization at the C5 position of D-glucuronic acid, confers unique properties to iduronate-containing biomolecules. In the manner of most of L-hexopyranoses, it could be expected L-IdoA residue to adopt a $^1C_4$ chair conformation as its sole most stable conformer. However, more than one conformation is accessible[68] in solution and, furthermore, the equilibrium between them can be modulated[69]. This unique feature of the iduronate ring is related to its ability to adopt several conformations of comparable energies[66b, 70], which explains the particularly good ability of iduronate-containing GAGs to control the activity of proteins such as chemokines, growth factors or blood coagulation enzymes[71].

Since changes in the ring conformation alter both the dihedral angles between vicinal hydrogen atoms and the distances between them, one can employ NMR spectroscopy to track such ring puckers by monitoring the spin-spin vicinal coupling constants ($^3J_{HH}$) and the proton-proton NOEs. Thus, over the past decades, extensive solution NMR experiments[10, 72] as well as theoretical calculations[73] have been performed to better understand the conformational flexibility of the iduronate ring. As a result, the picture describing the L-IdoA ring puckering, initially thought to be depicted by the equilibrium

between the $^{1}C_4$ and $^{4}C_1$ chair conformers, have been further completed when evidences about the $^{2}S_O$ skew-boat pucker playing a critical role in the control of blood coagulation appeared[74]. This $^{2}S_O$ conformer (or pucker) can be easily identified by NMR because it produces an intra-ring exclusive NOE cross-peak corresponding to the close contact between H2 and H5 protons (**figure 15**), which are not at an NOE distance either in the $^{1}C_4$ (**figure 15**) nor in the $^{4}C_1$ chairs.

Measured $^{3}J_{HH}$ couplings of iduronate as a monosaccharide[75], or as the non-reducing terminal of oligosaccharides[72, 75], indicated a mixture of both the $^{4}C_1$ and $^{1}C_4$ chairs with an additional contribution of the $^{2}S_O$ skew-boat. Specifically, an internal L-IdoA2S ring in heparin-like molecules shows a conformational equilibrium between the $^{1}C_4$ chair and the $^{2}S_O$ skew-boat puckers, with a negligible or non-existent population of the $^{4}C_1$ chair[75-76]. In this regard, it has been previously described that the $^{1}C_4$ and $^{2}S_O$ conformers may interconvert with little changes to the geometry of the glycosidic linkages to adjacent residues in the polysaccharide chain[66b, 77] (C4-O4 and C1-O1 bonds present similar orientations in both forms), and thus anticipating the idea, later demonstrated, that the global conformation of the oligo- or polysaccharide is independent of the iduronate conformational plasticity[77-78]. Concerning the iduronate flexibility, the existence of fast pseudorotational interconversion along the boats and skew-boats conformational space must also be considered (**figure 3**)[66b, 79].

The skew-boat $^{2}S_O$ occupancy in the L-IdoA and L-IdoA2S conformational equilibria is almost certainly biologically significant (inhibition of the coagulation cascade is thought to be initiated by antithrombin binding heparin with the iduronate residue in $^{2}S_O$ conformation, and synthetic heparins presenting $^{2}S_O$-biased iduronate analogues are highly potent[80]. Recently, it has been reported the first complete exploration of the low-energy conformations for iduronate ring (L-IdoA and L-IdoA2S) by MD simulations (**figure 16**)[17]. This study predicted that iduronate residues undergo microsecond puckering equilibrium ($^{1}C_4$-$^{4}C_1$ conformations exchange of the iduronate ring on the microsecond time scale) and that this depends on substitution pattern (L-IdoA 2-O-sulfation stabilizes the $^{1}C_4$ conformer) and epimerization (C5 epimerization leads to the $^{4}C_1$ chair)[17]. These observations have been of fundamental importance as almost all historical carbohydrate simulations are sub-microsecond in duration. Furthermore, they have revealed how enzymatic chemical modifications (epimerization and sulphation) fine-tune the free energy landscape of the iduronate ring and thereby mediate protein selectivity. According to the authors, the theoretical free-energy landscape obtained for the iduronate ring (also for the D-GlcA residue; see **figure 16**), not amenable to experiment, provides a new route for the development of so-needed carbohydrate mimetic biomaterials and pharmaceuticals[17].

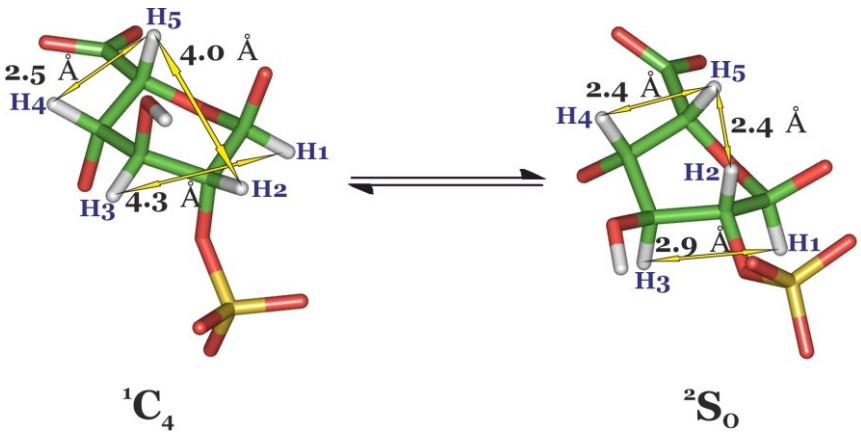

Figure 15. 3D representation of the $^1C_4$-^2S_O conformational equilibrium of the iduronate ring. Note the variation of the H2-H5 and H1-H3 distances, outside the NOE range in the $^1C_4$ chair conformer and within the NOE distance in the ^2S_O skew-boat pucker. Only the H4-H5 distance does not vary between both puckers.

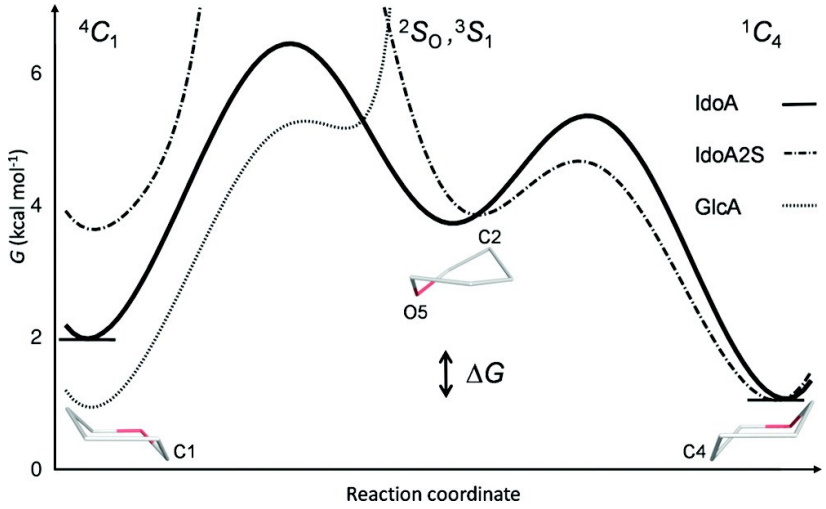

Figure 16. One-dimensional free energy (*G*) landscapes derived from equilibrium populations of each L-IdoA, L-IdoA2S and D-GlcA monosaccharide MD simulation. *Source: Sattelle et al. 2010*[17].

As hypothesis, it has been widely accepted that the driving force that determines the conformational equilibrium of the iduronate residue in heparin oligosaccharides is the electrostatic repulsion between anionic charges on neighbouring residues[69a]. Particularly for the regular region of heparin (iduronate in the L-IdoA2S form), due to the presence of three contiguous sulphate groups, NSO_3^-(GlcN)-2OSO_3^-(IdoA)-6OSO_3^-(GlcN), aligned on the same side of the helix (**figure 14**), it is tempting to speculate that the electrostatic "stress" existing on that part of the molecule might be at the origin of the singular conformational plasticity of the iduronate ring. Whereas the flexibility of the backbones of polysaccharides is usually associated only with rotation of the monosaccharide residues around the glycosidic bonds, the extra-flexibility induced by the presence of an equilibrium of two or more conformations of monosaccharide residues is a peculiar characteristic of iduronate-containing

glycosaminoglycans which could contribute to their binding properties and biological "versatility". These features contrast with the poor binding and biological properties of other glycosaminoglycans having approximately the same degree of sulphation and molecuIar weight, but the more rigid glucuronic acid as the major uronic acid.

Different factors act as modulators of the conformational equilibrium of the iduronate ring. Thus, the type of counterion may shift the equilibrium towards one of the puckers by interactions in specific sites of the polysaccharide[75, 81]. This is the case of Ca^{2+} ions, which drive the equilibrium towards the $^1C_4$ chair in heparin-like oligosaccharides[81b]. Furthermore, and more interesting in the framework of the present study, this equilibrium is highly sensitive to intramolecular factors such as the 2-O-sulfation of the iduronate residue and the sulfation pattern of the adjacent GlcN rings[68]. Thus, for internal L-IdoA or L-IdoA2S residues in heparin and heparan sulphate sequences, though only the $^1C_4$ and ^2S_O puckers participate in the conformational equilibrium[75-76], this is displaced towards the ^2S_O conformation when IdoA2S is 4-O-substituted with a 3-O-sulphated GlcNS residue (54-69% according to $^3J_{HH}$ at room temperature[75]). On the other hand, the IdoA2S conformational equilibrium is shifted towards the $^1C_4$ chair pucker as long as it is at the non-reducing terminal[75]. In addition, in the case of a terminal non-sulphated L-IdoA residue, the $^4C_1$ form also contributes significantly to the equilibrium[75].

Regarding the global conformation of heparin, which is determined by the geometry of its glycosidic linkages (Φ and Ψ torsions), rigid and flexible behaviours have been observed for the GlcN-IdoA and IdoA-GlcN (**figure 8**) linkages, respectively[82]. In addition, the conformational space sampled by the Φ torsion is restricted by the conditions imposed by the *exo*-anomeric effect, and so it commonly presents a narrow distribution of values around the *syn* geometry (**figures 8** and **17**). On the other hand, the Ψ torsion, not limited by the anomeric effect, behaves rigidly (*syn*) within the GlcN-IdoA linkages, but may provide a significant flexibility to IdoA-GlcN linkages[82]. This additional flexibility of the Ψ torsion is characterized by the appearance of *anti-ψ* conformations (±180º) together with the more common *syn-ψ* disposition (**figure 17**). From the NMR viewpoint, while the *syn-ψ* geometry can be identified by the presence of the H1'-H4 and H1'-H6 NOE cross-peaks, the *anti-ψ* conformation produces the H1'-H3, H1'-H5 and H5'-H6 exclusive NOEs (**figure 17**). This allows to experimentally identify the existence of conformational flexibility around the IdoA-GlcN linkages in solution.

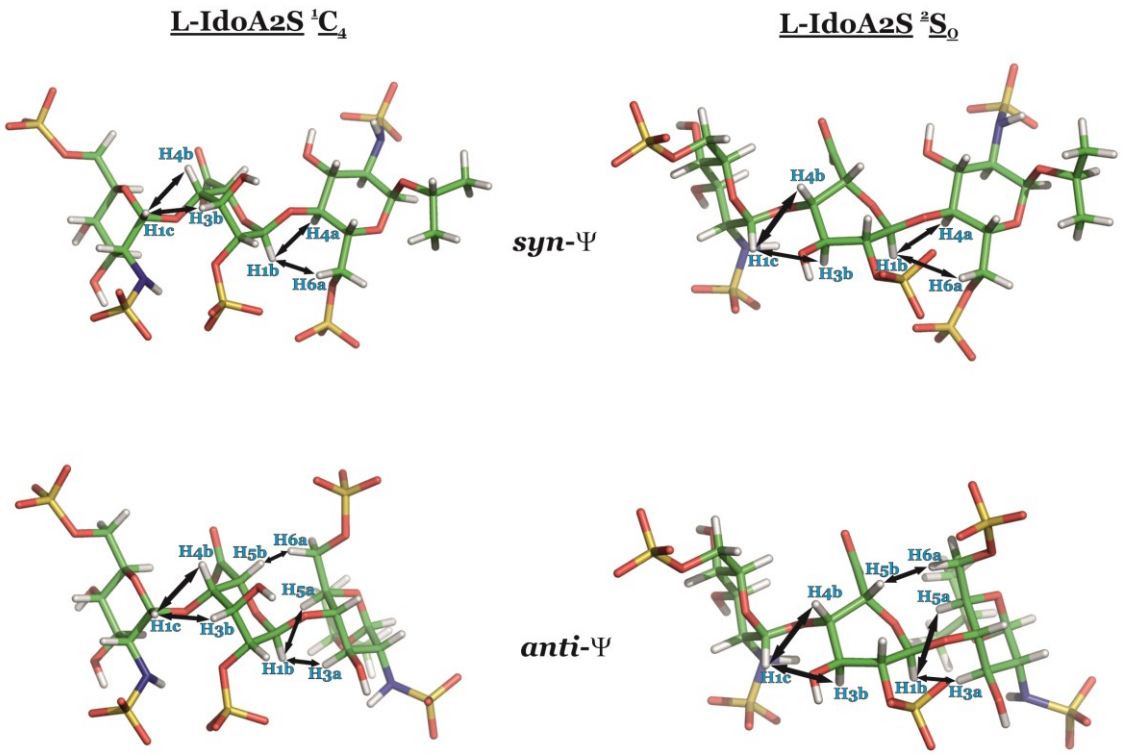

Figure 17. Interglycosidic NOE distances in heparin-like fragments with the iduronate ring adopting both the $^1C_4$ (left) and ^2S_O (right) conformations. Note the different of set NOEs observed for the minor and major *anti*-Ψ (bottom) and *syn*-Ψ (top) conformations, respectively, around the flexible IdoA-GlcN glycosidic linkage. Reducing-end GlcN, IdoA2S and the non-reducing rings are called a, b and c, respectively.

1.2.2 Heparin binding proteins: the acidic Fibroblast Growth Factor (FGF-1) case

FGF-1 is a member of the Fibroblast Growth Factor family that interacts with heparin/heparan sulphate (HEP/HS) polysaccharides and the membrane receptors FGFRs, thus triggering a signal that leads to different cellular essential functions such as the regulation of embryonic development, homeostasis and regenerative disorders[83]. The formation of a FGF1-HEP/HS-FGFR2 ternary complex is the key step for the activation of the FGF signalling pathway. Dimerization of the receptors and subsequent autophosphorylation activates a mitogenic response through an enzymatic cascade[40, 84]. Previously, our group addressed the study of the factors that govern the activation of FGF-1 via heparin binding by measuring the induced mitogenic activities of synthetic oligosaccharides[67, 85].

As the helical structure of heparin drives the sulphate groups towards opposite sides of its molecular axis, the multimerization of FGF molecules might be, at first, favoured [86]. In fact, there is a crystallographic structure of the complex between heparin (hexasaccharide) and FGF-1 (PDB code 1AMX)[87], which corresponds to a FGF-1 dimer linked by a regular heparin chain.

The hexasaccharide shown in **figure 18** (**Hexa2**, see *Chapter 3*) was previously prepared in order to shed some additional light on the HEP-FGF1 binding mode. This hexasaccharide sequence presents an axially non-symmetric sulphate distribution in order to prevent the potential formation of FGF1 dimers[85b, 85d]. NMR data in solution were consistent with a 1:1 HEP/HS-FGF1 complex (PDB code 2ERM; see **figure 19A**)[88], with the binding site corresponding to a swallow depression on the surface of the growth factor. Also, it was observed that the FGF1-induced mitogenic activity of this hexasaccharide was higher than that of an hexasaccharide presenting the regular sulphation pattern of heparin[67]. This result permitted to discard the dimerization of FGF1 mediated by heparin as an absolute requirement for the observed bioactivity. The high ability of **Hexa2** to activate FGF-1 is probably due to the combination of its sequence and sulphation pattern, which allow it to fully occupy both **a** and **b** sub-sites (**a**: primary; **b**: secondary) of the protein with 3 and 2 charged groups (SO_3^-), respectively (**figures 18** and **19A**). Probably, the existence of one additional sulphate group pointing towards the protein increases the stability of the FGF1-carbohydrate complex in comparison with a hexasaccharide of the regular region of heparin. This feature can also explain why the biological activity of **Hexa2** is similar to that of a synthetic octasaccharide of heparin regular region[67]. Furthermore, other investigations[89] suggested that primary amino acid differences within the heparin binding sites of FGFs and FGFRs, together with ligand-induced variations in the orientation of the D2 receptor domain, led to the formation of distinct positively charged canyons for individual FGF–FGFR combinations, with these canyons serving as molecular ”sieves” that select an optimal heparin sulphation pattern (**figure 19B**).

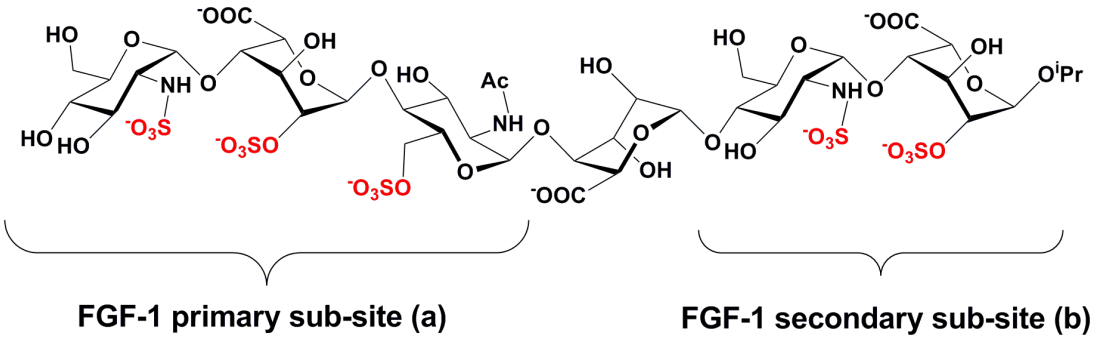

Figure 18. Scheme representation of the axially asymmetric heparin-like hexasaccharide that maximizes the interaction with FGF-1 (**Hexa2**). Note that, according to the NMR structure (PDB entry 2ERM), while the non-reducing part of the molecule occupies the primary sub-site (**a**), the reducing moiety binds to the secondary sub-site (**b**).

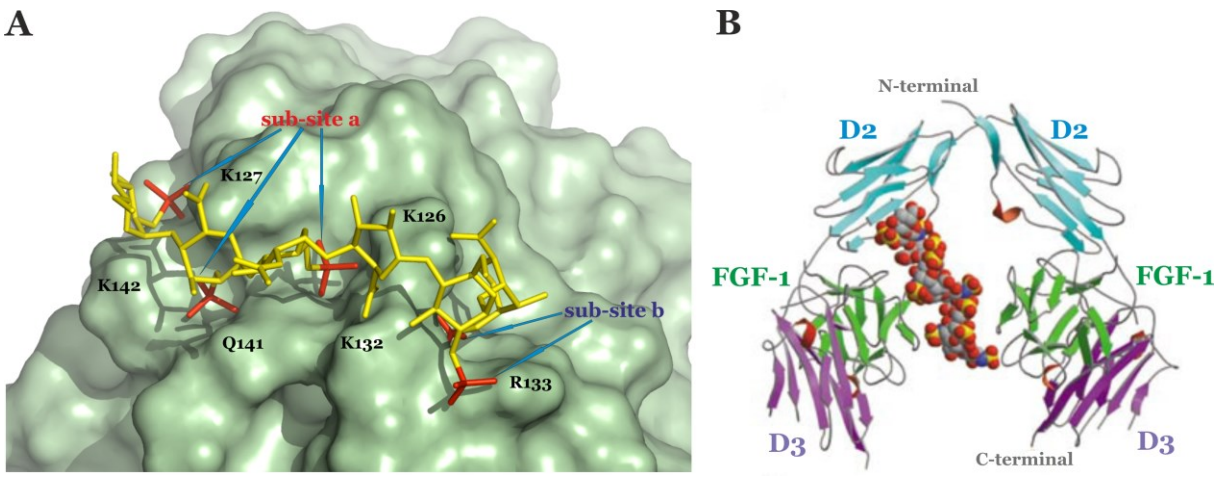

Figure 19. (**A**) Close 3D view of the binding site of FGF1 bound to **Hexa2** (PDB code 2ERM). Both sub-binding sites **a** and **b** are marked with blue arrows. The amino acid residues involved in the key electrostatic interactions with the hexasaccharide are labelled. (**B**) View of the FGF1-FGFR2-heparin ternary complex. FGFR2 domains 2 (D2) and 3 (D3) are shown in cyan and magenta, respectively, and FGF1 in green. The heparin molecule appears in CPK representation. *Adapted from Pellegrini et al. 2000*[84b].

1.2.3 Hyaluronic acid

The polysaccharide hyaluronan (HA), a ubiquitous extracellular matrix (ECM) component, is synthesized by many different cell types as a large co-polymer of β(1→4)-D-GlcA-β(1→3)-D-GlcNAc disaccharide repeating units (**figure 20**), typically in the MDa mass range. This large HA has general functions in matrix structural integrity, water and cation homeostasis in all tissues, and specialized functions in some tissues, such as a lubricant in synovial fluid[90]. HA binds to many different hyaladherins[91], HA-binding proteins, involved in remodelling and organizing ECM in a tissue-specific fashion[92]. HA binding to surface receptors activates cell-signalling events important for development, wound healing, and metastasis of some cancers[93]. Hyaluronic acid plays a pivotal role in the assembly process and is unique among the glycosaminoglycans because it is not sulphated or otherwise chemically modified during its biosynthesis[90]. It can be found at relatively high concentrations in soft connective tissue, for example in vitreous humor, synovial fluid and umbilical cord, where its large molecular domain results in viscoelastic solutions that flow at high concentration and fill space effectively. Also, HA is a scaffolding molecule onto which proteoglycans are non-covalently bound to form large macromolecular aggregates[94] whose high negative charge density serves to provide pressure in cartilage by drawing in water and ions.

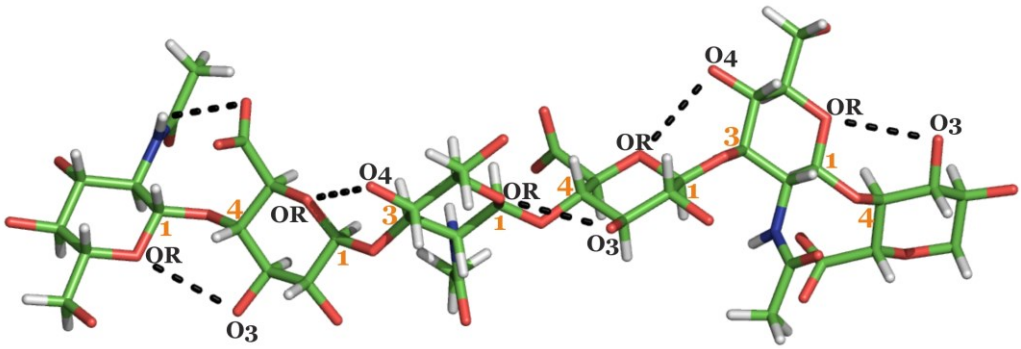

Figure 20. Average solution structure of hyaluronan, determined by NMR and computer modelling (PDB code 2BVK)[95]. Hydrogen-bond contacts are shown with dashed lines.

Initial studies on the conformation of HA, by application of X-ray fibre diffraction techniques, indicated a range of helical morphologies in response to changes in environmental factors, such as counterion, fibre humidity and pH. In particular, left-handed 3-fold and 4-fold helical conformations were reported in the presence of potassium, sodium and calcium ions. In addition, all of the fibre diffraction refinements featured intramolecular hydrogen bonds between amide and carboxylate groups of adjacent residues. However, the presence and precise nature of intramolecular hydrogen bonds has been difficult to characterize in aqueous solution[96]. Regarding NMR spectroscopy applied to HA, the first studies were unsuccessful in unequivocally characterizing the solution conformation of HA[96d, 97] because the overlap present in NMR spectra severely limited the amount of resolved and specific information that could be obtained. Also, measurements are usually averaged over the whole chain length, and thus vastly more problematic to interpret. These technical challenges and the consequent paucity of data have resulted in several different models published for the secondary and tertiary structure of HA[98]. Furthermore, much of these data come from NMR experiments that were performed nearly 20 years ago, and some of them not even in water[96b, 99]. Since that time, NMR and computational approaches have advanced considerably and now permit a new molecular description for the local solution conformation of HA and the role of intramolecular hydrogen bonds and water molecules to be developed. In a more recent structural study of HA carried out by Almond et al.[95], the solution conformation of this polysaccharide was investigated using molecular dynamics (MD) simulations and high-field nuclear magnetic resonance (NMR). In contrast to older studies, in this work the authors carried out MD simulations including explicit water molecules and sodium ions, while NMR experiments utilized ^{15}N-enriched oligosaccharides to allow residue-specific information to be obtained. The resultant average conformation (PDB code 2BVK)[95] was predicted to be almost a contracted left-handed 4-fold helix; i.e. similar to that observed for sodium hyaluronate fibers by X-ray diffraction, but with the acetamido side-chain *trans* to H2 proton (**figure 20**). The glycosidic linkages and acetamido side-chains were predicted to have standard deviation rotations of 138º and 188º around their mean conformations in free solution, respectively, and were not observed to be stabilized by strong intramolecular hydrogen bonds as X-ray fibre diffraction refinements describe for the solid-state. Rather, weak and transient hydrogen bonds that are in rapid exchange with solvent

molecules were predicted. These observations were quantitatively consistent with demanding residue-specific NMR data and corresponded to an HA molecule that is rod-like as an oligosaccharide and behaves as a stiffened random coil at large molecular mass, in close agreement with previous hydrodynamic observations. Such information is essential for the construction of viable models for the complicated interactions that are responsible for tissue assembly and remodelling within the extracellular matrix.

1.3 C-type lectins

C-type lectins constitute a superfamily of Metazoan proteins (i.e., those belonging to multicellular animals) containing C-type lectin-like domains (CTLDs). The CTLD term refers to protein domains that are homologous to the Carbohydrate Recognition Domains (CRDs) of the C-type lectins. Although originally C-type lectins were identified as the structures that bind carbohydrates in a Ca^{2+}-dependent manner (thereof the term C-type), not all the members of this family recognize carbohydrates and not all need Ca^{2+} ions for ligand binding. However, for simplicity, the terms "C-type lectin" and "C-type lectin receptor (CLR)" are used interchangeably in this thesis to refer to those lectins containing a CRD within their structures, thus showing calcium-dependent specificity for carbohydrates.

The mammalian CLRs are divided into 17 types based on their phylogenetic relationships and domain structures. Among them, only two classes (type I and II transmembrane proteins) are produced by dendritic cells, DCs, and Langerhans cells, LCs (**figure 21**). Type I C-type lectins (MMR and DEC-205) contain an amino-terminal cysteine-rich repeat (S–S), a fibronectin type II repeat (FN) and 8–10 carbohydrate recognition domains (CRDs), which bind ligands in a Ca^{2+}-dependent manner (**figure 21**). Type II C-type lectins contain only one CRD at their carboxy-terminal extracellular domain. The cytoplasmic domains of the C-type lectins are diverse and contain several conserved motifs that are important for antigen uptake: a tyrosine-containing coated-pit intracellular targeting motif, a triad of acidic amino acids and a dileucine motif (**figure 21**). Other type II C-type lectins contain other potential signalling motifs (ITIM, ITAM or proline-rich regions (P)).

Most of the CLR family members function as adhesion receptors, and only CLRs of type II, V and VI are present mostly on myeloid lineage immune cells and function as Pathogen Recognition Receptors (PRRs). PRRs are a series of innate immune receptors that include membrane bound toll-like receptors (TLRs), cytoplasmic Nod-like receptors (NLRs) and a RNA helicase family of receptors that recognize pathogen-associated molecular patterns (PAMPs) or danger-associated molecular patterns (DAMPs) and initiate immune responses against pathogens or repair responses in damaged tissues.

Type II CLRs (which function as PRRs) are mostly expressed by different DC subsets, including LCs (**table 2**). They bind pathogens through the recognition of mannose, fucose, glucose and other carbohydrate structures. The combination of CLRs on DCs enables the recognition of most classes of human pathogens. Pathogen recognition by CLRs leads to its internalization, degradation and subsequent antigen presentation[100].

The CLRs can be immune activating or inhibitory depending on their ability to associate with certain signalling molecules or the presence of specific motifs in their cytoplasmic tails. Most of the type II CLRs are predicted to be activating as in their transmembrane regions they have a positively charged residue which allows association with adaptor proteins. The activating CLRs may harbour the immunoreceptor tyrosine-based activation motifs (ITAMs). Upon ligand binding, clustering of CLRs occurs and ITAMs are phosphorylated, which initiates a downstream signalling cascade eventually

leading to activation of various cellular responses. Activating CLRs include dectin-2, DCAR, BDCA2, Mincle and DC-SIGN[101].

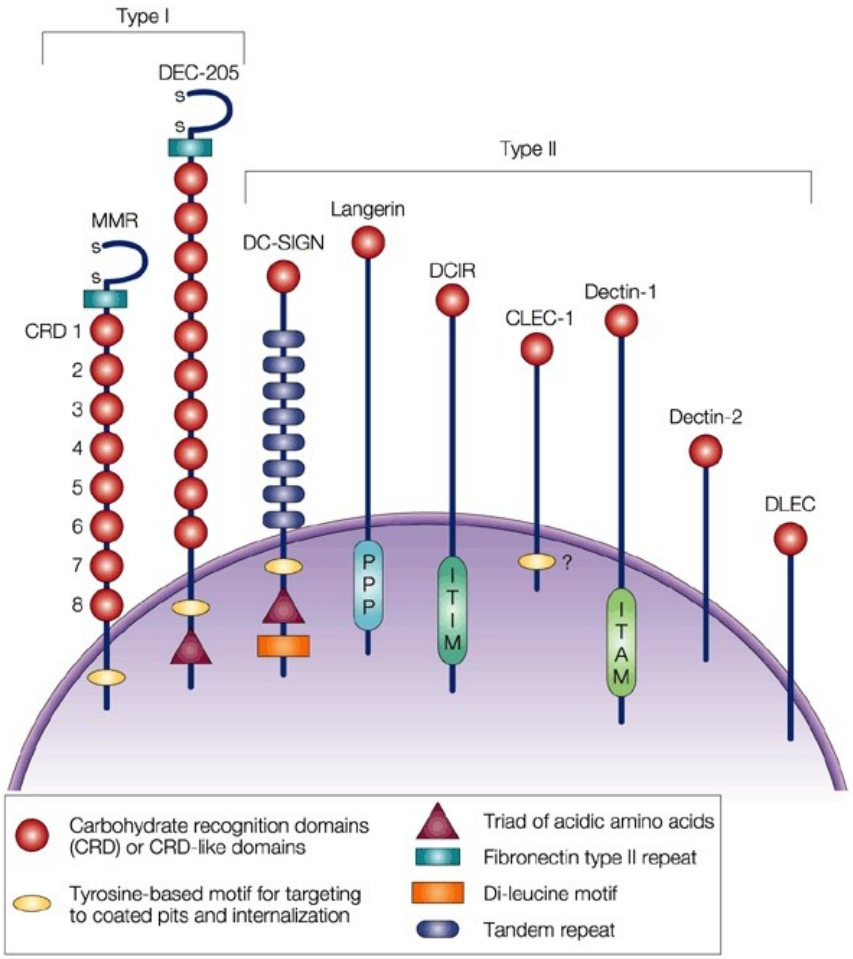

Figure 21. Types (I and II) of C-type lectins or lectin-like molecules that are produced by dendritic cells and Langerhans cells. *Legend: CLEC-1, C-type lectin receptor 1; DCIR, dendritic cell immunoreceptor; DC-SIGN, dendritic-cell specific ICAM-3 grabbing non-integrin; DLEC, dendritic cell lectin; ITAM, immunoreceptor tyrosine-based activation motif; ITIM, immunoreceptor tyrosine-based inhibitory motif; MMR, macrophage mannose receptor. Source: Figdor et al. 2002*[102].

Table 2. Main features of some C-type lectins produced by DCs and LCs.

C-type lectin	Type	Amino acids	Production	Ligand/s	Function/s	Key antibodies
MMR (CD206)	I	1456	DCs, LCs, Mo, Mϕ, DMECs	Man, Fuc, sLex	Antigen uptake[103]	MG38, anti-human[104]
DEC-205 (CD205)	I	1722	DCs, LCs, actDCs, thymic ECs	?	Antigen uptake[105]	
Dectin 1	II	247	DCs, LCs	β-glucan[106]	T-cell interaction[107]	
Dectin 2	II	209	DCs, LCs	?	Antigen uptake[108]	
Langerin (CD207)	II	328	LCs	Man, Glu, Gal6S, Fuc, heparin*	Formation of Birbeck granules[109], HIV-1 barrier[110]	DCGM4, anti-human
DC-SIGN (CD209)	II	404	DCs	HIV-1 (gp-120), SIV, mannan, ICAM-2, ICAM-3	T-cell interaction[111], HIV-1 pathology[112], migration[113], antigen uptake	AZN-D1, anti-human

Abbreviations: actDCs, activated dendritic cells; DMECS, dermal microvascular endothelial cells; Mo, monocytes; Mϕ, macrophages; sLex, sialyl Lewis X; ECs, endothelial cells. *To date, heparin is the first Langerin ligand reported to interact in a calcium-independent manner[114].

1.3.1 Structure of C-type lectin receptors (CLRs)

As we have introduced above, the common feature of all CLRs is that they possess at least one compact globular structure with a characteristic fold designated “C-type lectin-like fold” or “C-type lectin-like domain (CTLD)” that is unusual to any other known proteins. For the majority of CLRs that function as PRRs, the CTLDs bind sugars, usually in Ca^{2+}-dependent manner, and therefore this domain is commonly called a “carbohydrate recognition domain” (CRD).

All CRDs possess a characteristic "double-loop" fold. The whole domain can be regarded as a loop with two flanking α helices (α1 and α2) and two antiparallel β-sheets: N- and C-terminal β strands β1 and β5 constitute the basal β-sheet, and the top β-sheet is formed by strands β2, β3, and β4 (**figure 22**). The long loop region enters and exits the core domain at the same location, and is involved in Ca^{2+}-dependent carbohydrate binding and, for some CRDs, in domain-swapping dimerization. Four highly conserved cysteine residues form two disulphide bridges at the bases of the loops: C1-C4 bridge links α1 and β5, and C2-C3 bridges β3 strand and a loop upstream the β5 strand (**figure 22**).

The long loop region among different CRDs varies, and those that possess it are designated "canonical", while those that lack it are called "compact". The presence or absence of a short extension at N-terminus, a β1-hairpin, further subdivides CRDs to long or short forms, respectively. Two additional cysteine residues at the beginning of CRDs sequence are characteristic for the long form CRDs. The corresponding disulphide bridge C0-C0′ stabilizes the β-hairpin (**figure 22**).

There may be up to four Ca^{2+}-binding sites in the CRDs, and their occupancy depends on the sequence of a particular CRD. Sites 1, 2, and 3 are located within the long loop, while the fourth site (Ca-4) participates in the salt bridge formation between helix α2 and β1/β5 sheet[115] (**figure 22**).

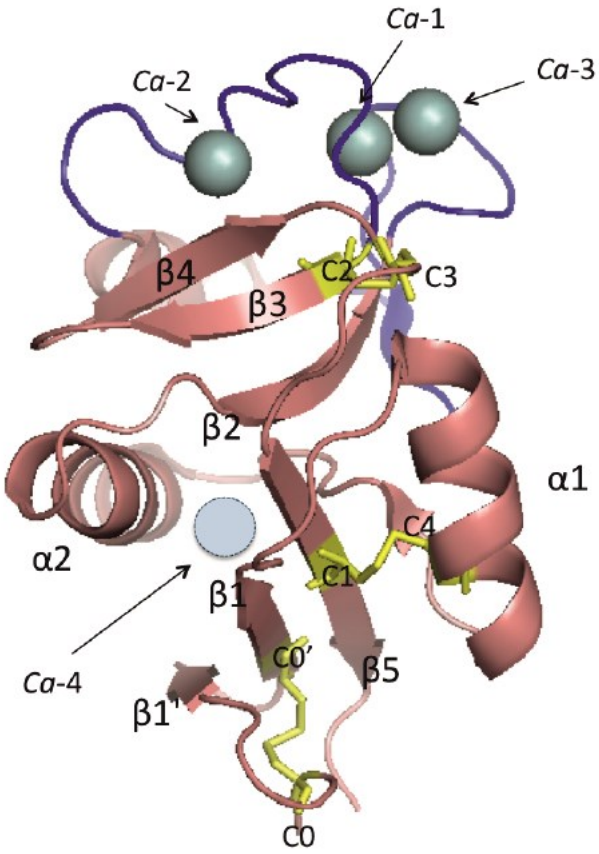

Figure 22. Cartoon representation of a common CRD structure in CLRs (DC-SIGN CRD; PDB code 1K9I). The long loop is shown in blue and the disulphide bridges in yellow sticks. The Ca^{2+}-binding sites 1, 2 and 3 existing in this lectin are shown as cyan spheres, while the location of the fourth Ca^{2+}-site (Ca-4, absent in this particular structure) has been drawn and appears as a cyan circle. *Source: Doctoral Thesis of Ieva Sutkeviciute*[116].

1.3.2 Structural features of glycan binding to CRDs

Amino acid residues with carbonyl side chains involved in Ca^{2+} coordination in site 2 form two characteristic motifs in the CLR sequence that, together with the calcium ion itself, are directly involved in monosaccharide binding (**figure 23A**). The first group of residues, the "EPN motif, is contributed by the long loop region and contains two residues with carbonyl side chains separated by a proline in *cis* conformation (**figure 23B**). The carbonyl side chains provide two Ca^{2+}-coordination bonds, form hydrogen bonds with the monosaccharide and determine binding specificity. The *cis*-proline is highly conserved and maintains the backbone conformation that brings the adjacent carbonyl side chains into the positions required for Ca^{2+} coordination. The second group of residues, the "WND motif" (**figure 23**, **B** and **C**), is contributed by the β4 strand. Although only the asparagine and aspartate residues of this motif are involved in Ca^{2+}-coordination, the tryptophan amino acid immediately preceding them is a highly conserved contributor to the hydrophobic core[117] and is a useful landmark for detecting the motif in a sequence. In the MBP-A structure shown in **figure 23B**, ASN205 and ASP206 provide three Ca^{2+}-coordination bonds (two from the side chains, one from the backbone carbonyl of Asp) and also form hydrogen bonds with the sugar. Additionally, another carbonyl side chain is involved in site 2 formation, which belongs to the residue preceding the second conserved cysteine at the end of the long loop region (Glu193 in MBP-A; see **figure 23B**) and forms one coordination bond with the Ca^{2+} ion.

The overall network of the hydrogen-bond donors and acceptors in binding site 2 (**figure 23**) determines the binding orientation of the carbohydrate and also which hydroxyls of the carbohydrate it can accept, i.e. the monosaccharide specificity. The EPN motif has a configuration that accommodates mannose-type monosaccharides (**figure 23B**), whereas the QPD motif determines specificity for galactose-like monosaccharides (**figure 23C**). In both of these motifs the *cis* configuration of the two carbonyl sidechains separated by proline is crucial for Ca^{2+}-coordination and sugar binding. Besides the restrictions imposed by the H-bond network, other structural elements in the binding sites introduce selectivity to particular ligands within the mannose or galactose groups.

As no other Ca^{2+}-binding site except for site 2 is known to be involved in sugar binding, and as the site 2 motifs can be confidently detected in the sequence, it is common in the literature to associate the predicted Ca^{2+}-dependent carbohydrate binding properties of an uncharacterized sequence with the presence of these motifs[118]. Although this is a useful simplification, it should be noted that the absence of the motifs associated with Ca^{2+}-binding site 2 does not indicate that the CTLD is incapable of binding Ca^{2+}, as there are two independent sites (1 and 4). Also, the presence of these motifs does not guarantee lectin activity for the CLR, as there are numerous examples of C-type lectins that contain the conserved motifs but are not known to bind monosaccharides.

The other three Ca^{2+} binding sites (1, 3 and 4) play a structural stabilization role, as removal of Ca^{2+} increases susceptibility to proteolysis and changes physical properties of the domain. Ca^{2+} binding site 2 is also important for structural stability of the domain. It has been shown that pH-induced loss of

Ca^{2+} causes the destabilization of the loops, which has an important physiological role for CRDs of endocytic receptors as internalization of ligand-bound receptor to acidic lysosomes and consequent Ca^{2+} loss leads to the release of the ligand for further processing, while receptor is recycled to the cell surface[115].

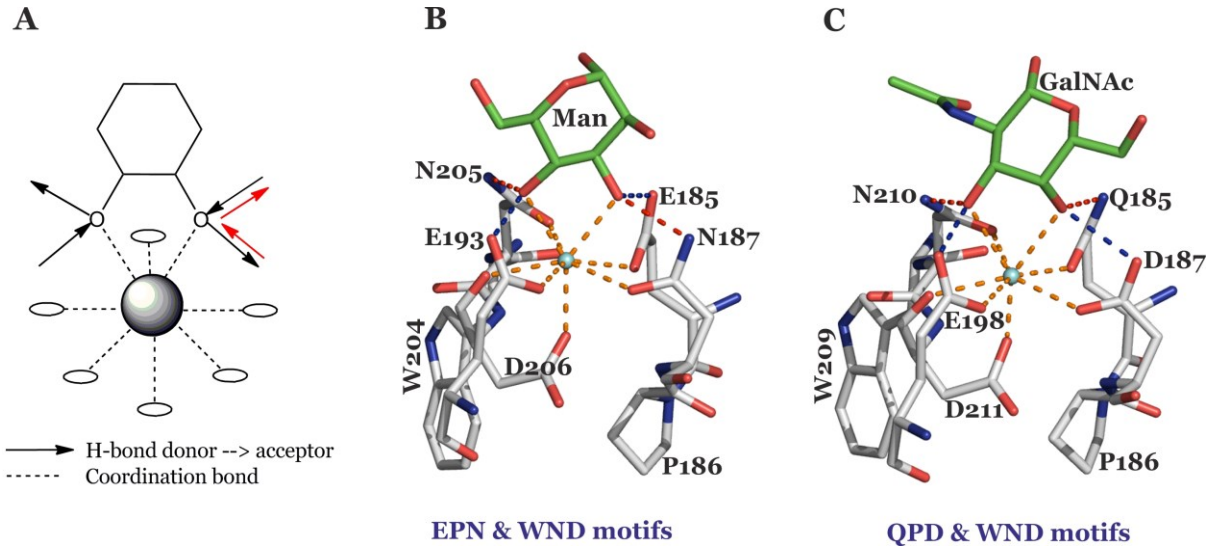

Figure 23. Ca^{2+}-dependent monosaccharide binding by CLRs. (**A**) Schematic representation of a Ca^{2+}-hexose-CLR complex. Two hydroxyl oxygens and the ring of the hexose are shown. The Ca^{2+} atom is shown as a large grey sphere, and oxygens as empty circles and ovals. Protein groups that act as hydrogen donors and acceptors are not shown. Black arrows show the direction of hydrogen bonds in "mannose-specific" CLRs, while red arrows indicate opposite directions in "galactose-specific" CLRs. (**B**) Mannose residue bound to MBP-A CRD (PDB code 2MSB). (**C**) GalNAc residue bound to MBP-A mutant CRD (PDB code 1BCJ). In (B) and (C), the coordination bonds are orange, and the hydrogen bonds where sugar hydroxyl acts as acceptor or donor are red or blue, respectively. The Ca^{2+} ions are shown as a cyan spheres. *Adapted from the Doctoral Thesis of Ieva Sutkeviciute*[116].

1.4 Langerin: a natural barrier to HIV-1 infection

Langerin, a transmembrane type II C-type lectin with carbohydrate specificity, is almost exclusively expressed on epidermal Langerhans cells (LGs), i.e., a subset of dendritic cells (DCs) with very singular structural features. Due to its placement at the epithelium level, these cells were reported a few year ago to constitute the first recognition barrier to HIV-1 particles[110]. Interestingly, instead of the common antigen-presenting cell role of DCs, LCs internalize and subsequently degrade HIV-1 virions upon Langerin sequestration of the virus[110]. Thus, Langerin has been presented as as a natural barrier to HIV-1 transmission by LCs[110].

The host cell infection by HIV starts by binding of HIV envelope proteins "Env" (i.e, a trimmer of gp120 and gp41 heterodimers, where gp41 initially is hidden) to its primary receptors $CD4^+$ T lymphocytes, a member of the immunoglobulin superfamily that enhances T-cell receptor (TCR)-mediated signalling (**figure 24**), being an absolute requirement for the infection. CD4+ T lymphocytes also express chemokine receptors CCR5 and CXCR4 that are exploited by HIV to enter the cells, hence also called HIV co-receptors (**figure 24**). Env binding event induces rearrangements in its gp120 subunit, which ultimately result in V3 loop repositioning and bridging sheet exposure that are essential for co-receptor engagement. The attachment of the virion can be relatively nonspecific, so that, for instance, HIV Env can interact with negatively charged cell-surface heparan sulphate proteoglycans[119]. More specific adhesion includes interactions between the envelope protein and α4β7 integrin[120] or CLRs such as DC-SIGN[112] and Langerin[110]. Either way of adhesion has been proposed to bring HIV envelope into close proximity with the host CD4 and a co-receptor, leading to the fusion of viral and target cell membranes[121] (**figure 24**).

Subsequent binding to the co-receptors, CCR5 or CXCR4 depending on the virus strain R5 or R4, triggers the membrane fusion potential of Env, and usually is followed by the "surfing" of the virus particle to the site where productive membrane fusion may occur[122]. It is thought that HIV might usurp the host cell machinery to reach cell surface sites where membrane fusion can occur[123]. Besides, HIV may need to be endocytosed by the host cell for productive membrane fusion to occur[124]. Upon formation of Env-gp41 complex, the co-receptor undergoes conformational changes that expose its hydrophobic fusion peptide (**figure 24**), which then inserts into the host cell membrane and folds to form a six-helix bundle. The latter is the driving force that brings the opposing membranes into close proximity, resulting in the formation of a fusion pore[125] (**figure 24**). Once the virus enters the cell, it can start its replication and productive infection, which ultimately lead to the depletion of $CD4^+$ T lymphocytes in the body, and thus the acquired immunodeficiency syndrome (AIDS).

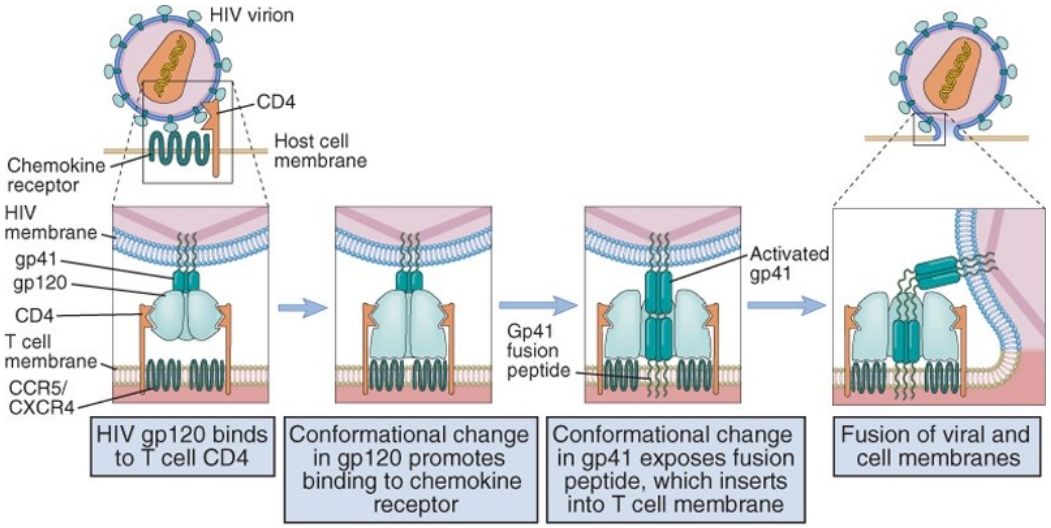

Figure 24. Schematic representation of HIV entry to a target cell. *Source: Abbas et al. 2011*[126].

Importantly, although the major target of HIV is $CD4^+$ T cells, earlier studies have shown that DCs are crucial for HIV-1 infection enhancement and dissemination in mucosa, in the case of sexual HIV transmission, since the virus hijacks them to achieve the productive infection of the CD4+ T cells and the burst of the disease[127]. The stromal DCs that express a C-type lectin DC-SIGN (DC-Specific ICAM3 Grabbing Non-integrin) and reside in mucosa of vagina and ectocervix, have been repeatedly reported to be exploited by HIV-1 to enhance its infectivity of T cells.

There exist several HIV infection mechanisms (**figure 25**) in which it is noticeable that DC-SIGN has a very important role in DC-mediated HIV transmission enhancement[112]. After virion binding to DC-SIGN, HIV can be endocytosed into DCs and, subsequently, the intact viral particles can be stored either in multivesicular bodies[112, 128], as integrated provirus (following productive infection of DCs) or as DC-SIGN-bound virions on the cell surface and protected from degradation[129] (**figure 25**). During infection there is an accumulation of intact viral particles on DC side while HIV receptors (CD4, CCR5) are presented on $CD4^+$ T cell side. This situation greatly facilitates HIV-1 transfer from DCs to T cells. Indeed, it has been demonstrated that blocking DC-SIGN prevents HIV-1 binding and subsequent *trans* infection of $CD4^+$ T cells[130].

While *trans* infection is responsible for the early stage of infection (24h after HIV exposure), there exists a different pathway of DC-SIGN-bound HIV transmission to T lymphocytes. This pathway is involved in long-term HIV transfer (72h after exposure) and it occurs as a *cis* infection of DCs by transfer of DC-SIGN-bound virus to canonical HIV entry receptors, CD4 and CCR5, which leads to productive infection of DCs and, in turn, to the presentation of increased viral load to the T cells[131].

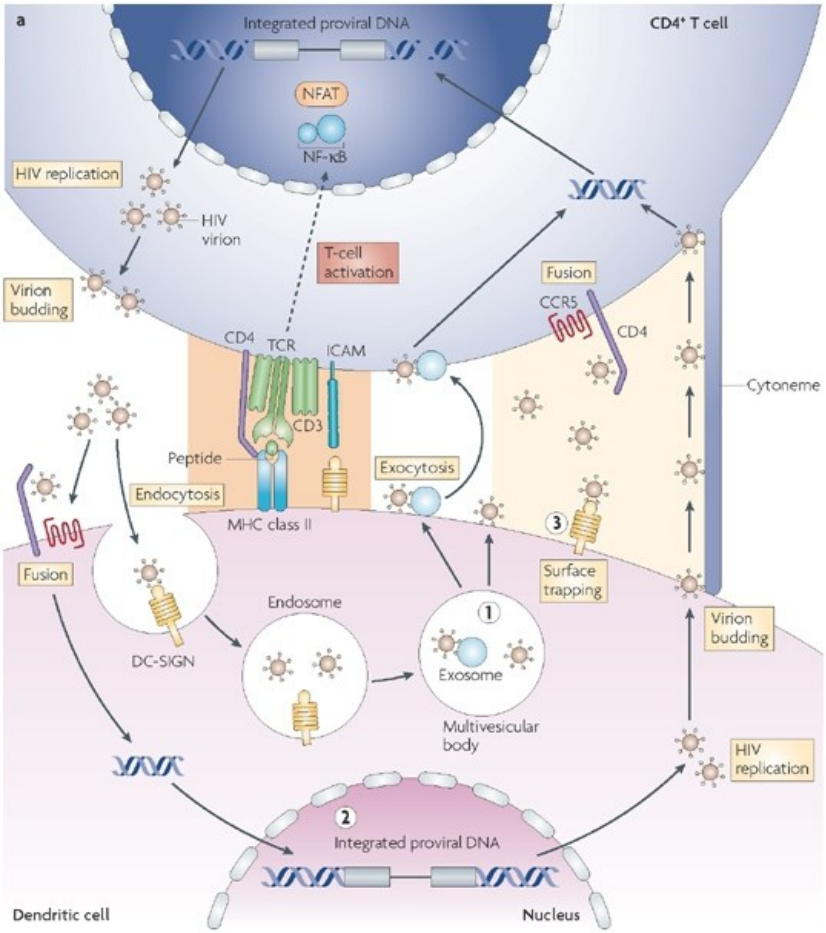

Figure 25. Importance of DCs-T-cells interactions and DC-SIGN implication on HIV-1 transmission. *Source: Hladik and McElrath 2008*[132].

1.4.1 Novel role of epithelial LCs and their transmembrane protein Langerin in HIV infection

LCs constitute a subset of DCs, whose role is not completely clear, that are located in epithelium of mucosal tissues and epidermis[133] (**figure 26A**). Although initially they were assumed to function as antigen-presenting cells (APCs), like dermal or stromal DCs[134], the accumulating evidence of their fail to present antigens from various viruses and parasites to T cells activation, and the observation that LCs induce T regulatory cells, supported the hypothesis that these cells could have an inmunosuppresive tolerogenic role[135] . Notably, Langerhans cells are the only epidermal cells to constitutively express major histocompatibility complex class II molecules[136], CD1a molecules[133a], and Langerin[109] at their cell surface. LCs play a key role in the induction of immune responses against invading pathogens by capturing and processing foreign antigens and migrating to draining lymph nodes to present processed antigens to T cells[137]. A characteristic hallmark of LCs is the unique presence (in the cytoplasm) of a tennis racquet- or rod-shaped membranous structures, known as Birbeck granules (BGs)[138] (**figure 26, B** and **C**), that are thought to be part of the endosomal

recycling pathway (they are subdomains of the endosomal compartment)[109]. The protein Langerin has been shown to be the main component and responsible for the formation of these Birbeck granules[109].

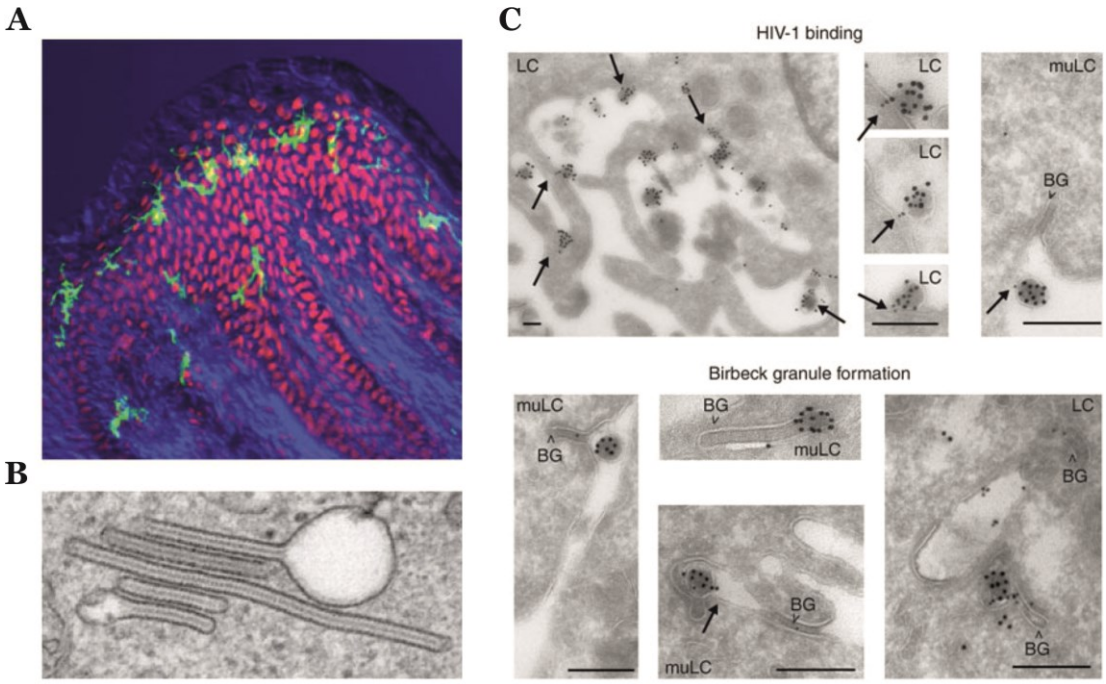

Figure 26. (**A**) Langerhans cells (green) in the foreskin surrounded by nuclei of epithelial cells and other cell types (red). (**B**) Birbeck granules, site of HIV processing in LCs. (**C**) HIV-1 particles (black small circles) internalized into Birbeck granules of LCs (and mutant muLCs) upon capture by Langerin. *Sources: (A, B) Schwartz 2007[139]. (C) de Witte et al. 2007*[110].

As DC-SIGN, Langerin is a transmembrane type II C-type lectin, almost exclusively expressed in humans by epidermal LCs but also present on dermal $CD103^+$ DCs and lymph node resident $CD8^+$ DCs[109, 140]. It contains one calcium-dependent carbohydrate recognition domain with a short cytoplasmic tail with a proline rich motif[141], forms a trimer on the cell surface and, upon crosslinking with either a cell-bound or a soluble ligand, it induces the formation of BGs. As a C-type lectin, Langerin plays a role in pathogen recognition. However, only few pathogens have been demonstrated to interact with Langerin. Both HIV-1[110, 142] and *Mycobacteria leprae*[143] have been identified as pathogens that interact with human Langerin, whereas murine Langerin was shown to bind to *Candida albicans*[144].

LCs express receptors including CD4, CCR5 and Langerin[145]. It must be noted that, different from the other DCs (antigen-presenting cells), LCs play a significant preventive role in HIV infection process. In particular, it has been reported that epidermal LCs expressing Langerin efficiently bind HIV virions, which in turn are directed to Birbeck granules for degradation[110] (**figure 27**). Moreover, a recent study has shown that vaginal LCs may have low or no expression of Langerin, and thus they are

susceptible to HIV infection[146]. These experimental evidences show an important protective function of Langerin in HIV invasion process.

LCs reside in the epithelium (particularly in skin epidermis and mucosal epithelium) while DCs are placed at the sub-epithelium level (**figure 27**). Thus, LCs are the first cells encountered by HIV-1 particles, which are captured by the LCs membrane receptor Langerin, resulting in viral clearance and inhibition of HIV-1 transmission across the mucosal layer (**figure 27**). The disruption of the epithelial barrier through trauma or ulcerations enables HIV-1 to circumvent the LCs barrier and reach the DC-SIGN-expressing DCs, which act as antigen-presenting cells, thus mediating HIV-1 transmission to T cells (**figure 27**). On the other hand, the inhibition of Langerin function by drugs, co-infections or mutations, gives rise to the efficient transmission of HIV-1 to T cells but, in this case, through infection of LCs (**figure 27**). Hence, two C-type lectins such as DC-SIGN and Langerin present very different roles in HIV invasion despite the high homology of their CRDs and the overlap of ligand specificities[147]. While DC-SIGN promotes HIV transmission and infection, Langerin fights against HIV invasion.

Based on the carbohydrate recognition specificity of Langerin CRD for mannose[148], fucose[148], N-acetyl-glucosamine[148] and 6-O-sulphate-galactose[148b] monosaccharides (see next section), it is likely that Langerin has a broader specificity for pathogens than it has been thought so far. Also, it has been very recently demonstrated that the trimeric extracellular domain of Langerin (LgECD) bind GAGs (specifically heparin) in a novel non-calcium dependent binding site[114] (see next section).

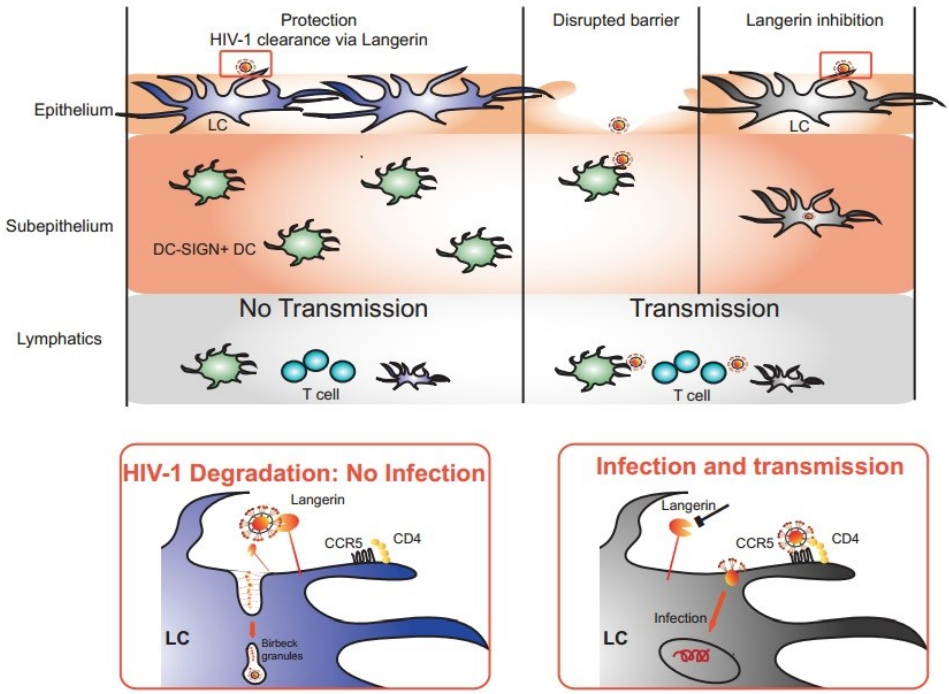

Figure 27. Schematic representation of the implication of LCs on HIV-1 transmission. Note that only LCs are present at the first or more external pathogen-recognition level (epidermis), therefore being the first entity of the immune system to encounter the virus. *Source: de Witte et al. 2007*[110].

1.4.2 Langerin 3D structure and known ligands

Langerin is formed by 328 amino acids (37,5 kDa) and has an overall molecular structure similar to DC-SIGN and other CLRs. It consists of a N-terminal short cytoplasmic domain, a unique transmembrane domain, and a large extracellular domain (ECD) subdivided into a neck domain and a C-terminal carbohydrate-recognition domain, CRD, (**figure 28**)[109]. The CRD of Langerin contains the EPN motif characteristic of mannose-type specificity (also presents the WND motif; see **figure 29**). On the other hand, the cytoplasmic domain of Langerin presents a proline-rich signalling motif (WPREPPP), which could function as a docking site for signal transduction proteins, and indeed, it was demonstrated to be important for Langerin intracellular targeting[149].

Like many other CLRs, Langerin exists as an oligomer (active form), forming trimers stabilized by a coiled-coil of helices in the neck region (**figure 28**). Trimer formation is essential for binding to oligosaccharide ligands because, as is typical for C-type CRDs, the CRD of Langerin has only low affinity for monosaccharides[148a, 150]. In addition, it should be noted that oligomerization of C-type lectins is also important for determining selectivity for particular oligosaccharide structures. For example, in serum mannose-binding protein, three CRDs in the trimeric unit are held in a fixed position via interactions between the CRDs and an helical neck region so that the binding sites are arranged to interact with arrays of sugars in polysaccharides of bacterial cell walls, but not with mammalian high mannose-type oligosaccharides[151].

Although both DC-SIGN and Langerin are C-type II lectins with similar sequences and structures of their CRDs, several important differences exist between these two CRDs (**figure 30**). The overall structure of Langerin CRD is maintained by two conserved disulphide bridges, in contrast to the four S-S bonds in DC-SIGN. Unlike DC-SIGN, Langerin CRD has only one Ca^{2+} ion at the calcium site 2 (conventional sugar binding site; see **figure 22**), and the lack of other Ca^{2+} ions in sites 1 and 3 (**figure 22**) might be at the origin for the high flexibility of the β2-β2´ loop comprising residues 258-262, which also leads to the formation of a large groove specific to Langerin structure. However, the most significant difference between these two lectins is present in their sugar-binding site topologies.

A unique feature of Langerin CRD is the presence of two lysine residues (LYS299 and LYS313) within the sugar binding site that provides it the basic character that allows Langerin to accommodate sulphated sugars in the binding site. On the other hand, while PHE313 is an important side chain in DC-SIGN sugar binding-site as it forms stacking interactions with the sugar ring[152], PHE315 residue is not placed at the appropriate position in Langerin as to participate in such sugar interaction (**figure 30**, upper). The overall topology of Langerin CRD composes only a small binding site strongly constrained by LYS299, while DC-SIGN has a potential to adapt more extended oligosaccharides with a widely open binding site (**figure 30,** down), and thus less extensive secondary contacts take place with Langerin CRD than with DC-SIGN CRD. Also, the linkage between CRD and neck domains of DC-SIGN and Langerin are also different. Thus, while the flexibility in the former is retained, Langerin

trimers present rather rigid CRDs[153], and this probably gives rise to different ligand recognition mechanisms for the two lectins.

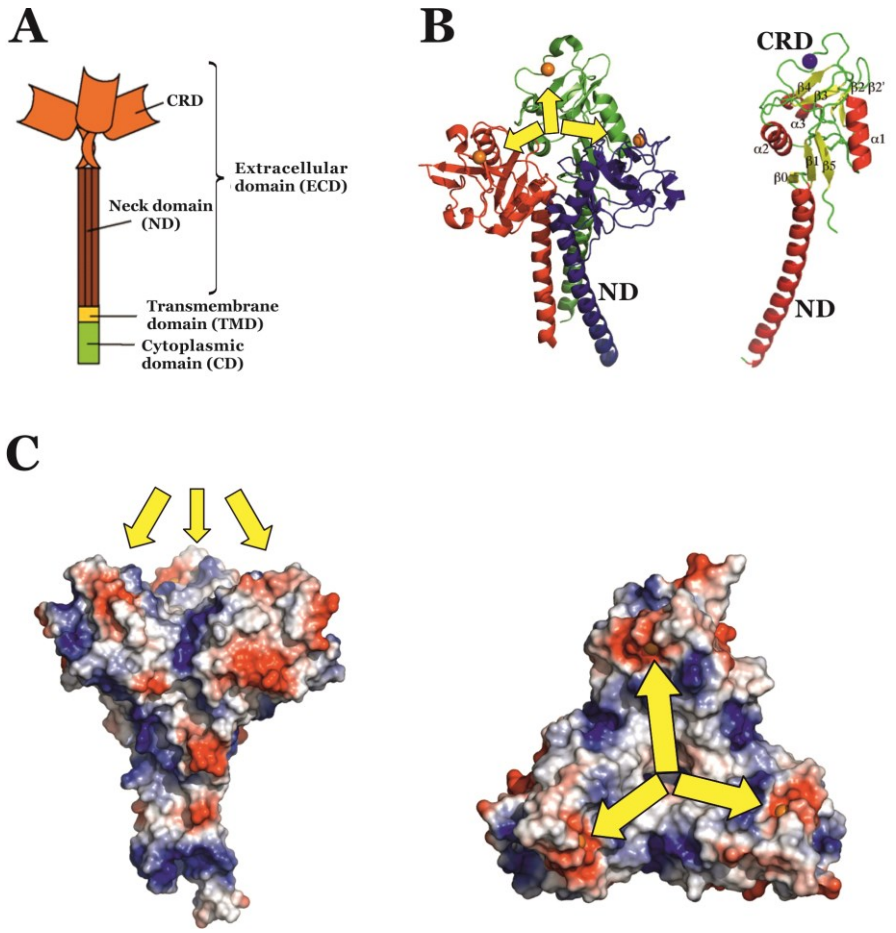

Figure 28. (**A**) Schematic 2D representation of the whole Langerin protein, indicating each of its four domains. (**B**) 3D cartoon model of Langerin extracellular domain, Lg ECD (left; PDB code 3KQG[153]), showing each of the three protomers in a distinct colour and the Ca^{2+} ion as an orange sphere. On the right, secondary structure of protomer A of LgECD, showing α-helices in red, sheets in yellow, loops in green and the calcium ion in blue. (**C**) Surface representation of Lg ECD from the front (left) and above (right) perspectives. Note: yellow arrows point at the CRDs calcium binding site.

Langerin combines the features of several other C-type lectins. The organization of the trimmer suggests that the multiple binding sites are widely spaced and probably rigidly positioned in a manner reminiscent of mannose-binding protein[153]. However, although the binding site in each CRD is relatively open as in mannose-binding proteins, recent glycan array results indicated that Langerin shows preferential binding to specific ligands rather than broad specificity resulting from the binding of just terminal sugar residues, as seen for mannose-binding protein. The structures of the Langerin-ligand complexes reveal that this preferential binding results from a small number of favourable contacts with portions of selected oligosaccharide ligands beyond the monosaccharides sitting in the primary binding site. In this regard, the binding site has some of the features of the binding site in DC-SIGN, although secondary contacts are less extensive.

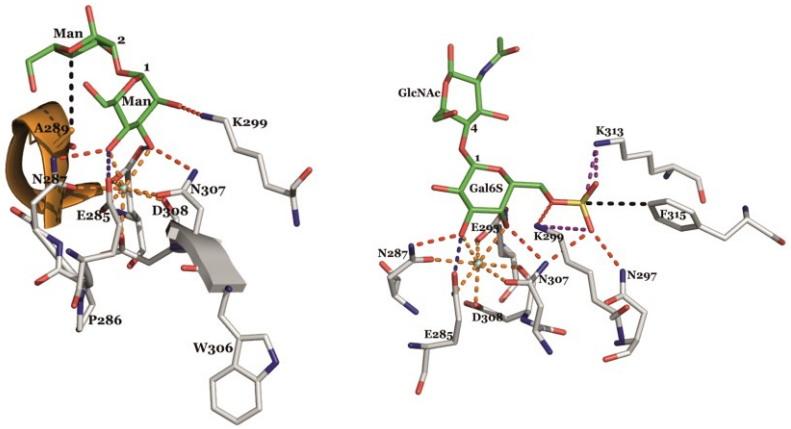

Figure 29. Zoomed 3D view of Langerin CRD complexed to Man-α(1-2)-Man (left; PDB code 3P5F[148b]) and Gal6S-β(1-4)-GlcNAc (right; PDB code 3P5I[148b]). Amino acids and ligands carbons are coloured grey and green, respectively, and the calcium-ion appears as a cyan sphere. With dashed lines are indicated the coordination bonds (orange), the favourable electrostatic (purple) and hydrophobic (black) interactions, and the hydrogen bonds where sugar hydroxyl acts as acceptor (red) and donor (blue).

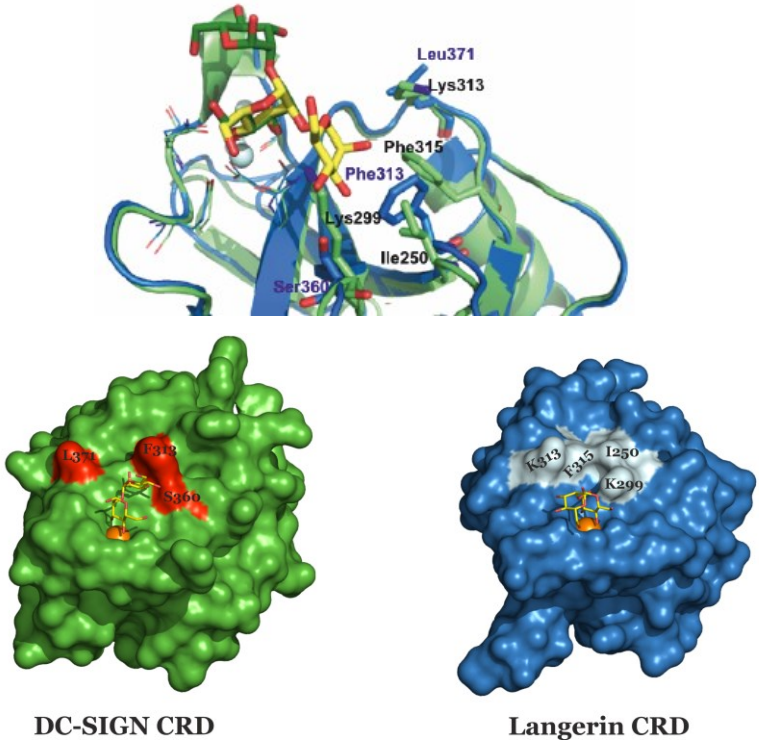

Figure 30. (*Upper*) Zoomed view of the CRDs superimposition of DC-SING (blue; PDB 2IT5[154]) and Langerin (light green; PDB 3P5F[148b]) bound to Man-α(1-2)-Man, shown in yellow (DC-SIGN) and green (Langerin) sticks. The side chains of the EPN and WND motifs are shown in lines, whereas the essential amino acids residues required for sugar binding appear as sticks, and the Ca^{2+} ions as light cyan spheres. Blue and purple labels correspond to DC-SIGN and Langerin side chains, respectively. (*Down*) Comparison of the CRD structure of DC-SIGN (green surface) and Langerin (blue surface), both complexed to Man-α(1-2)-Man (PDBs 2IT5 and 3P5F, respectively). The most significant side chains involved in ligand binding are coloured red (DC-SIGN) and palecyan (Langerin). The ligands and calcium ions appear in yellow-red sticks and orange spheres, respectively.

Two years ago, Feinberg et al. published the crystal structures of the CRD from human Langerin bound to a series of oligomannose compounds, the blood group B antigen and a fragment of β-glucan[148b]. The structures revealed binding to mannose, fucose and glucose residues by Ca^{2+}-coordination of vicinal hydroxyl groups with similar stereochemistry. Oligomannose compounds bind through a single mannose residue, with no other residues contacting the protein directly (**figure 29**). Likewise, a β-glucan fragment, Glc-β(1–3)-Glc-β(1–3)Glc, binds to Langerin through the interaction of a single glucose residue with the Ca^{2+} site. Also, the fucose moiety of the blood group B trisaccharide Gal-α(1-3)-(Fuc-α(1–2))-Gal binds to the Ca^{2+} site, in this case selectively (compared to other fucose-containing oligosaccharides) due to additional favourable interactions of the non-reducing terminal galactose residue.

Surprisingly, binding was observed for the $6SO_4$-Gal-β(1–4)-GlcNAc ligand through Ca^{2+}-coordination of the equatorial and axial 3-OH and 4-OH group, respectively, of the galactose residue (**figure 29**), a heretofore unobserved mode of galactose binding in a C-type carbohydrate-recognition domain bearing the EPN signature motif characteristic of mannose binding sites but not the QPD motif that determines specificity for galactose-like monosaccharides. Thus, it was suggested that the observed stabilizing charge-charge interactions between the 6-sulphate group of galactose and LYS299 and LYS313 residues seem to compensate for the non-optimal binding of galactose at this site (**figure 29**). This result puts on evidence that subtle secondary interactions with Langerin CRD determine the binding selectivity with this protein, which highlights the difficulty of predicting binding specificity from a comparative analysis of closely related CRDs.

Also, binding of Langerin to $Lewis^X$-type carbohydrates containing 6-sulphated galactose has been reported[155]. Since Langerin does not bind $Lewis^X$, high-affinity binding to $6SO_4$-$Lewis^X$ likely arises from interactions of the sulphated galactose similar to those described above for $6SO_4$-Gal-β(1-4)-GlcNAc. In addition, the structure explains why Langerin does not bind to 3-sulphated galactose groups[153, 155], i.e., the 3-OH group is not available for Ca^{2+} coordination and hydrogen-bonding interactions in the conserved site.

In any case, it has to be noted that, in contrast to the other ligands identified by screening on synthetic and pathogen glycan arrays, oligosaccharides bearing terminal $6SO_4$–Gal residues are not found on pathogen surfaces. However, binding to such structures has been suggested to mediate interaction with endogenous sulphated ligands such as keratan sulphate (KS), which contains the $6SO_4$-Gal-β(1-4)-GlcNAc repeating unit[155a]. Langerin is unique among receptors with C-type CRDs because of its ability to bind both mannose/fucose-type ligands and a galactose-based ligand.

Very relevant in the context of the present thesis, it has been very recently demonstrated by the research group of Prof. Fieschi that, apart from the sulphated sugar KS, Langerin performs unique specificity towards a broader range of GAGs. In particular, Langerin is able to bind heparin, HS but also several types of CS, and surprisingly, binding event is independent of the presence of Ca^{2+} ions[114]. Additionally, the measured affinities for heparin and HS were in the nanomolar range, ranking them

as the best ligands ever described for Langerin. Also, the interaction is strictly dependent upon oligomerization, so that Lg ECD trimer must be present for binding to occur[114].

From the above evidences, Langerin-GAGs interactions appear to be completely different from how related CLRs traditionally recognize their ligands. The mandatory requirement for Langerin trimeric form suggests the existence of a unique binding site constituted by the assembly of at least 2 of 3 protomers. Furthermore, experiments with HEP, HS, CS and DS compounds showed that 1) binding to Langerin cannot be simply correlated to a net charge effect and 2) specific sulphation (D-GlcN6S and L-IdoA2S) strengthen the interaction[114].

In the referred work[114], a docking model for Langerin-heparin binding was proposed (**figure 31**) and three main structural features were identified: 1) heparin fragment docking pose did not interact with the calcium ions, 2) Langerin-heparin interactions were not driven by polar forces (salt bridges and hydrogen bonds), and 3) the molecular recognition of heparin fragments depended upon more than one Langerin CRD. Furthermore, the authors suggested that the simultaneous N- and 6-sulphation of GlcN residues seemed to be essential for the interaction by acting as a "bridge" between both CRDs (**figure 31**).

Langerin is thus able to selectively interact with sulphated carbohydrates through two totally distinct modes: 1) a Ca^{2+}-dependent binding mode in the CLR canonical site when OH groups are available at 3 and 4 positions of the sugar ring, and 2) in a Ca^{2+}-independent manner for sulphated GAGs where either the O3 or O4 atom is engaged in the polysaccharide glycosidic linkage.

The identification of the Langerin specificity towards GAGs raises the question of the physiological relevance and role of such an interaction. HS is abundantly present in the tissues hosting LCs and directly exposed at DCs surfaces where it participates to the capture of many pathogens as well as immune activation[156]. Interestingly, a previous work on the biochemistry of LCs trafficking indicated that heparin, and more particularly N-sulphated glucosamine moieties of heparin, could inhibit LCs trafficking[157].

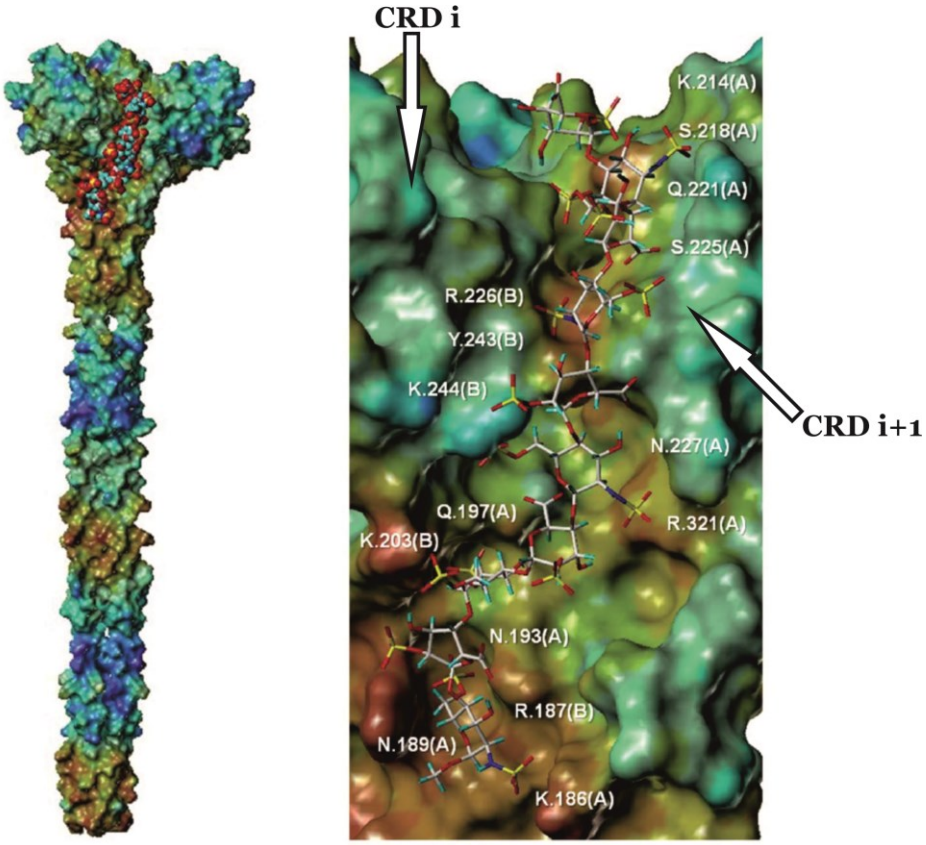

Figure 31. Docking binding mode of heparin decasaccharide (space fill, left; capped stick, right) in complex with Langerin (Connolly surface colour-coded according to the electrostatic potential; from blue to red for negative and positive electrostatic areas). Amino acids mediating the main interactions with the decasaccharide are labelled in white. Note that the ligand extends along the interface between two contiguous CRDs (i and i+1). *Source: Chabrol et al. 2012*[114].

1.5 Objectives of the research

Along years, our research group has gained a deep expertise in the synthesis and structural characterization of chemically pure glycosaminoglycans (in particular, HEP/HS and HA) of defined sequence and structure. Under the framework of a wide-ranging research project aiming to determine the structure and dynamics of different synthetic glycosaminoglycans (GAGs) and to depict the molecular basis governing their recognition with proteins, we have addressed the study of different heparin and hyaluronan-like oligosaccharides in the free state and bound to both the acidic Fibroblast Growth Factor, FGFG-1 (heparin-like oligosaccharides), and the HIV-1 natural barrier Langerin (heparin and hyaluronan-like olisosaccharides). Langerhans cells (Langerin-expressing cells) have been described to mediate the elimination of the virus through the transmembrane cell surface receptor Langerin. Thus, knowing in detail the recognition mechanisms of this protein is especially important to develop appropriate inhibitors for HIV-1 infection, not to inhibit Langerin but, on the contrary, to enhance its interactions with natural receptors and/or to avoid Langerin inhibition when targeting DC-SIGN. The elucidation at atomic resolution of the structural features that characterize

Langerin interactions with different synthetic ligands analogue to those contained in the extracellular matrix (ECM) of cells is of fundamental importance.

To address the objectives of this thesis, we have taken advantage of the experience gained by the group in high-resolution Nuclear Magnetic Resonance (NMR) spectroscopy and computational methods such as Molecular Dynamics (MD) and Computational Docking. In particular Saturation Transfer Difference (STD) NMR spectroscopy and transferred-NOE techniques have been employed for the binding studies. From the computational view, the *know-how* acquired by the PhD candidate during two short stays at Prof. Robert J. Woods´ (Complex Carbohydrate Research Centre, CCRC, Athens, Georgia, USA) and Prof. Anne Imberty´s (Centre de Recherches sur les Macromolécules Végétales, CERMAV, Grenoble, France) groups has been of huge importance for the execution of high-level MD and docking calculations.

In particular, the objectives we have pursued in the present thesis are:

- Study the solution conformation and dynamic behaviour of heparin-like tri- and hexasaccharides.
- Rationalize biological activity data of different FGF-1 oligosaccharides based on the determination of their conformations in solution.
- Identify at atomic level key ligand-receptor contacts for the formation of Langerin complexes with hyaluronan-like disaccharides and heparin-like tri- and hexasaccharides (ligand epitopes).
- Characterize the bioactive conformations of the ligands bound to Langerin, in solution.
- Qualitatively determine the binding affinities of different ligands for Langerin, in solution (competition STD experiments).

Chapter 2

Techniques and tools

As it has been outlined in *Chapter 1*, the physiological processes taking place in cells are the result of highly regulated intermolecular protein–protein and protein–ligand interactions. In particular, the binding of low molecular weight molecules (ligands) to macromolecules such as proteins plays a major role in the regulation of biological processes, e.g., signal transmission and cellular metabolism. The analysis of protein–ligand interactions is crucial, not only for understanding the regulation of biological functions, but also for designing novel bioactive molecules that modulate protein function or inhibit protein-ligand interactions[158].

With respect to carbohydrates, the elucidation of the 3D structures and dynamics properties of oligosaccharides and glycoconjugates, both in the free state and bound to proteins, is a prerequisite for a better understanding of the molecular basis of their associations and interactions, and the relationships between structures and functions, which are involved in the biochemistry of recognition processes and the subsequent rational design of carbohydrate-derived drugs. These have been claimed to be the main challenges in structural glycoscience[159] and many efforts in this direction still have to be done.

A large variety of biophysical techniques have been developed to characterize protein–ligand complexes, e.g., surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), fluorescence polarization assay (FP), fluorescence resonance energy transfer (FRET), enzyme-linked immunosorbent assay (ELISA), differential scanning fluorimetry (DSF), microscale thermophoresis (MST) and electrospray ionization mass spectrometry (ESI-MS) to cite the most popular methods[158d, 160]. SPR and MST techniques permit to obtain kinetic interaction parameters, whereas ITC measures the thermodynamic properties of binding in solution. In silico approaches have been also applied to search for ligands for a protein target (virtual screening) or to propose 3D models of protein–ligand complexes (docking calculations)[161], whereas X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are both experimental techniques for resolving atomic structures[162].

Since the scientific interests of our group are focused on the structural studies of protein-carbohydrate complexes at atomic resolution, it is the X-ray crystallography and NMR spectroscopy the existing tools to our aim. However, one of the disadvantages of X-ray crystallography applied to carbohydrates is that the oligosaccharides, either in their free form or as part of glycoconjugates, are inherently difficult to crystallize, and structural data from X-ray studies are sparse[163]. Even when succeeding in crystal formation, part or the whole glycan is, in most cases, not observed in the high-resolution electron density map[57d] due to the intrinsic high flexibility of carbohydrates. Furthermore, the experimental assessment of carbohydrate recognition by X-ray crystallography is impeded by difficulties of co-crystallizing proteins and carbohydrates. To overcome this limitations, it has been tried, for instance, to stretch the polysaccharide into an oriented fibre[164]. Also, it has been employed electron diffraction to study very small crystals, or needles, that can be obtained from polysaccharides[165]. In any case, the amount of data collected to date is that small that building a model by molecular mechanics is necessary to resolve the 3D structure.

On the other hand, high-field NMR spectroscopy in solution state is one of the most important techniques for probing intermolecular interactions. NMR spectroscopy detects and reveals protein-ligand interactions with a large range of affinities, and it is widely used in pharmaceutical research to identify hits from compound library screening in drug discovery[166]. Protein–ligand complexes are analysed using the so-called protein-observed and ligand-observed NMR experiments in which the NMR parameters of the protein and the ligand, respectively, are compared in their free and bound states[166]. In particular, ligand-observed methods are not limited by the protein molecular size and therefore have great applicability for analysing protein–ligand interactions. The use of these NMR techniques has considerably expanded in recent years, both in chemical biology and in drug discovery.

In protein-observed methods, the chemical shift perturbations of the protein resonances observed upon ligand addition are identified to localize the ligand binding site. This enables one to immediately distinguish specific from non-specific binding. The 3D structure of the protein-ligand complex can be resolved via heteronuclear experiments performed on isotopically labelled (^{13}C, ^{15}N, ^{2}H) protein samples. The structure resolution requires molecular dynamics calculations with experimental NMR restraints resulting from chemical shifts, scalar couplings, nuclear Overhauser effects (NOEs), paramagnetic interactions or residual dipolar couplings[162a, 167]. The major drawbacks are the experimental time and the need for a highly stable and soluble protein. In addition, these methods are limited in routine practice to proteins with low molecular masses (less than 30 kDa) to avoid great effort with regard to both labelling strategies and resonance assignment.

NMR parameters such as transverse, longitudinal, and *cross-relaxation* rates strongly depend on the molecular rotational correlation time τ_c, which is directly related to the molecular weight. Ligand-based NMR experiments rely on the modification of such size-sensitive NMR parameters for the ligand in the presence of a protein receptor[166b, 168]. Considering a diffusion controlled protein-ligand binding of weak to moderate affinity (dissociation constant, K_D, typically between 10^{-8} and 10^{-3} M), the association-dissociation process is fast within the chemical shift time-scale, so that the NMR parameters observed are a simple population-weighted average between the free and bound states. In contrast to protein-observed experiments, ligand-observation is more sensitive with larger receptors and do not require the use of isotopically labelled proteins. Ligand-based methods can be used for the detection of interactions and the measurement of protein–ligand affinities, and can also provide pertinent structural information on the protein–ligand complexes.

2.1 Structure by NMR: the Nuclear Overhauser Effect (NOE)

2.1.1 Origin of the NOE

An accurate definition of the Nuclear Overhauser Effect (NOE) is the change in intensity of one resonance when the spin transitions of a dipolarly coupled nucleus are somehow perturbed from their equilibrium populations. This perturbation is achieved by either saturating a resonance, i.e., equalising the spin population differences across the corresponding transitions (then it is called *steady-state* NOE), or inverting it by reversing the population differences across the transitions (transient NOE). Thus, the magnitude of the NOE observed for spin I when spin S is perturbed ($\eta_I\{S\}$) is expressed as the percentage of relative intensity change between the equilibrium intensity (I_o) and that in the presence of the NOE (I), so that

$$\eta_I\{S\}=\frac{I-I_o}{I_o}\cdot 100 \qquad \textit{Eq. 1}$$

The intensity changes caused by NOE can be either positive ($I>I_o$) or negative ($I<I_o$) depending on the motional properties of the molecule and the signs of the magnetogyric ratios of the spins involved.

To facilitate the understanding of the origin of the NOE, we will consider a system only formed by two homonuclear spin-1/2 nuclei of ^1H (positive magnetogyric ratio), I and S, contained in a rigid molecule that tumbles isotropically in solution, i.e., it does not show any preferential axis about which to rotate. In this idealistic system both protons are not scalarly coupled ($J_{IS}=0$) but they are enough close in space as to share dipolar coupling, this is, magnetic interaction through space between two spins such that both of them are able to sense the presence of the other dipolar-coupled partner. Therefore, upon selective saturation of ^1H nucleus S, the spin populations of nucleus I will be also perturbed and the system will try to come back to the initial equilibrium situation. Although equilibrium recovery takes place by different relaxation mechanisms it is only the *cross-relaxation* pathways, characterized by the W_0 (zero-quantum) and W_2 (double-quantum) transition probabilities (or rates), those responsible for the NOE development (**figure 1A**).

It is important to note that the W_0 and W_2 *cross-relaxation* pathways always compete with one another, with the dominant mechanism dictating the sign of the observed NOE and being dependent on the reorientational dynamics properties of the molecule (**figure 1B**). Thus, a small molecule performing a rapid tumbling in solution, which corresponds to a short correlation time τ_c, will generate fluctuating local magnetic fields of high frequency. On the other hand, a macromolecule, featuring slow tumbling rates (long τ_c) will give rise to low-frequency magnetic fields. These fluctuating local fields are the responsible for inducing *cross-relaxation* when their oscillation frequencies correspond to the W_0 and/or W_2 transitions. Furthermore, those of low-frequency are significantly more efficient than the high-frequency magnetic fields in activating the *cross-relaxation* pathways, due to the

different spectral density function (J(ω))[2] that describe the fast, intermediate and slow motions (**figure 1B**). Therefore, for slow tumbling the lower energy W_0 process predominates such that a negative NOE ($I<I_0$) is efficiently developed (**figures 1B** and **2**). On the contrary, W_2 (higher energy) is the dominant mechanism acting during *cross-relaxation* of fast tumbling molecules, so that a positive NOE ($I>I_0$) is observed (with lower intensity compared to negative NOE; see **figure 2**). For and intermediate tumbling rate (usually medium-size molecules), the NOE will be either positive or negative depending on the dominant *cross-relaxation* process. Also, it has to be noted that there are two limits of motion in terms of the magnitude of the NOE developed. Thus, the rapid tumbling of a small molecule in a low viscosity solvent will favour the W_2 process to a large extent, displaying high positive homonuclear NOEs. This is called *extreme narrowing limit*. In contrast, the slow molecular tumbling of large molecules in high viscosity solvents will stay in the *spin-diffusion limit*, characterised by an enormously favoured W_0 mechanism, and thus, highly negative homonuclear NOEs.

Apart from the *cross-relaxation* mechanisms, only responsible for the NOE growth, single quantum relaxation pathways (W_1) activate to re-establish the equilibrium population differences of the non-saturated nucleus (I in the simplified model) as soon as the NOE begins to develop, so acting against the NOE build-up. Thus, if W_1 relaxation happens to be rather more efficient than W_0 and W_2 pathways together, the macroscopic magnetization will probably come back to the equilibrium before a measurable NOE is developed and this will not be observed. The NOE therefore results from the balance between distinct competing relaxation pathways, with its sign depending on the W_2-W_0 difference and its magnitude on the three W_0, W_1 and W_2 rates (see *Eq. 2*, derived from the so-called Solomon equation).

$$\eta_I\{S\} = \frac{\gamma_S}{\gamma_I}\left[\frac{W_2 - W_0}{W_0 + 2W_1 + W_2}\right] = \frac{\gamma_S}{\gamma_I}\left[\frac{\sigma_{IS}}{\rho_{IS}}\right] \qquad \textit{Eq. 2}$$

So, the ideal conditions for the NOE to be observed are inefficient W_1 processes and efficient W_0 or W_2 transitions.

[2] J(ω) is the frequency distribution of the fluctuating magnetic fields associated with molecular motion. It may be viewed as the probability of finding a component of the motion at a given frequency ω (in rad/s).

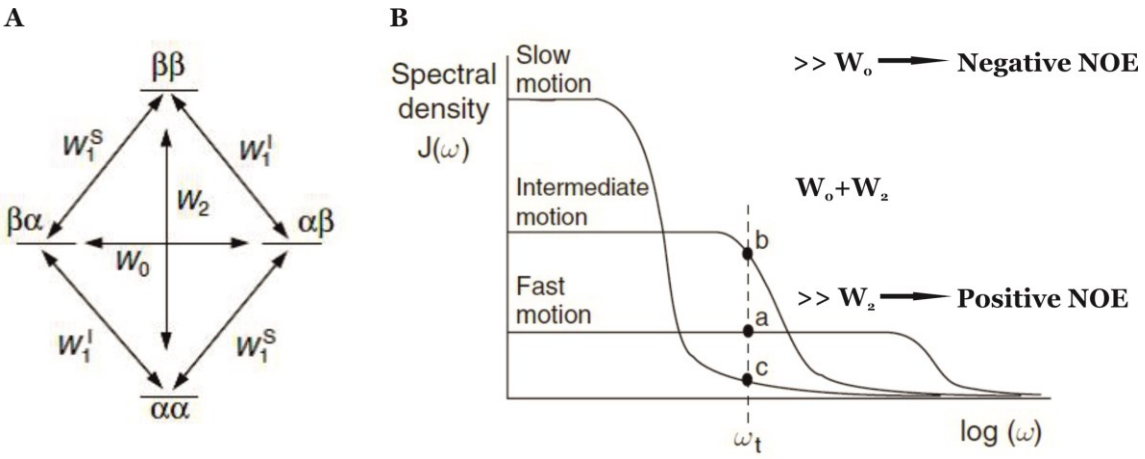

Figure 1. (**A**) The six possible transitions in a two-spin system. Note the two *cross-relaxation* pathways, W0 and W2. (**B**) Evolution of the spectral density function (J(ω)) as a function of the frequency of motion (ω) in logarithmic scale (ω_t corresponds to the frequency of an hypothetical spin transition). *Source: Claridge 2009*[169].

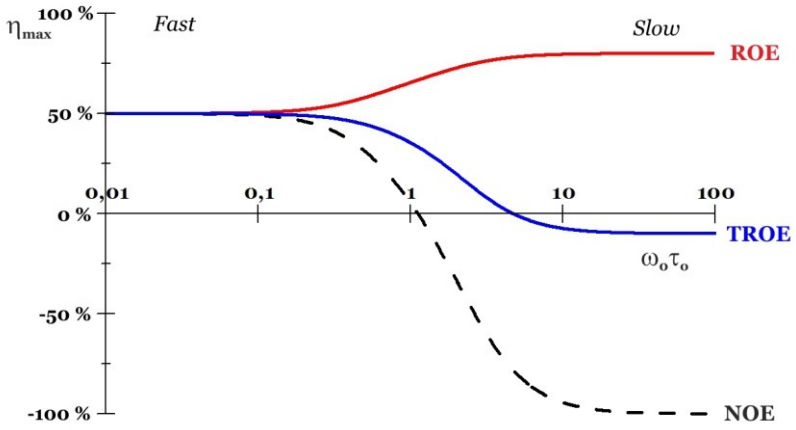

Figure 2. Variation of the maximum theoretical homonuclear steady-state enhancement (η_{max}), for NOE (black dashed line), TROE (blue bold line) and ROE (red bold line) experiments, in a two-spin system as a function of molecular tumbling rates in logarithmic scale (defined by the dimensionless parameter $\omega_o\tau_c$, with ω_o being the spectrometer observation frequency and τ_c the rotational correlation time). The region of fast motion is the *extreme narrowing limit* and that of slow motion is the *spin-diffusion limit*.

2.1.2 Rotating-frame NOE (ROE)

The greatest problem associated with NOE experiments is the *zero-crossing* region around $\omega_o\tau_c \approx 1$ where the conventional (laboratory-frame) NOE observed via steady-state or transient techniques becomes vanishingly small. This typically occurs for mid-sized molecules with masses of around 1000-2000 daltons, depending on solution conditions and spectrometer frequency. With the increasing interest in larger molecules in many areas of organic chemistry research coupled with the wider

availability of higher field instruments, this is likely to be a region visited ever more frequently by the research chemists' molecules. Other than altering solution conditions (such as changing temperature) in an attempt to escape from this, the measurement of NOEs in the rotating-frame provides an alternative solution. In this case, the *cross-relaxation* rate between homonuclear spins is given by an expression that remains positive for all values of τ_c, and the undeniable benefit of ROEs is, quite simply, that they remain positive for all realistic molecular tumbling rates. For small molecules, the magnitude of the ROE matches that of the transient NOE, whilst for larger molecules it reaches a maximum for homonuclear spins of 68%, but under no circumstances does it become zero (**figure 2**). Similarly, the NOE and ROE growth rates are identical for small molecules but differ for very large ones.

For very large molecules, the ROE therefore grows twice as fast as the NOE, and has opposite sign[170]. In essence, ROEs develop whilst magnetisation is held static in the transverse plane, rather than along the longitudinal axis (hence they are sometimes also referred to as transverse NOEs). To generate the required population disturbance of the source spins, the target resonance is subjected to a selective 180^o pulse prior to the non-selective 90^o pulse, such that it experiences a net 270^o flip and is thus inverted relative to all others. Transverse magnetisation is then "frozen" in the rotating frame by the application of a continuous, low power spin-lock pulse to prevent evolution (in the rotating frame) of chemical shifts. The experiment is more frequently performed as the 2D experiment where it is usually termed ROESY (rotating-frame NOE spectroscopy). The situation during the spin-lock may be viewed as the transverse equivalent of events during the transient NOE mixing time (**figure 3**). The action of the spin-lock is to maintain the opposing disposition of magnetisation vectors, which would otherwise be lost through differential chemical shift evolution, and so allows the ROE to develop through cross-relaxation in the transverse plane. Spin relaxation here is characterised by a time constant called $T1\rho$, of very similar magnitude to T_2. In utilising the spin-lock, one has effectively replaced the static B_0 field of the conventional NOE with the far smaller rf B_1 field, and it is this that changes the dynamics of the NOE. Whereas $\gamma B0$ typically corresponds to frequencies of hundreds of megahertz, $\gamma B1$ is typically only a few kilohertz, meaning $\gamma B_1 << \gamma B_0$ and hence ω_1 (the rotating-frame frequencies) $<< \omega_0$. The consequence of this is that $\omega_1\tau_c << 1$ for all realistic values of τ_c, and all molecules behave as if they are within the *extreme narrowing limit*. Thus, ROEs are positive, any indirect effects have opposite sign to direct effects and tend to be weak, and saturation transfer can be distinguished by sign from ROEs, regardless of molecular size and dynamics. Against these obvious benefits are a number of experimental problems, principally TOCSY transfers (particularly contributing to strongly coupled systems such as glycans), also occurring during the spin-lock, and signal attenuation from *off-resonance* effects.

An alternative ROESY sequence is the T-ROESY experiment, which is effective at suppressing TOCSY transfer. This is achieved by substituting the low-power continuous wave lock field used in ROESY sequence by a mixing sequence of $180^o_x\ 180^o_{-x}$ or $180^o_x\ 180^o_{-x}\ 360^o_x\ 360^o_{-x}\ 180^o_x\ 180^o_{-x}$ pulses. However, cross-relaxation rates in T-ROESY are an equal mixture of ROE and NOE components, and

moreover, the cross-relaxation rates measured are four times lower than those in conventional ROESY[171] (**figure 2**).

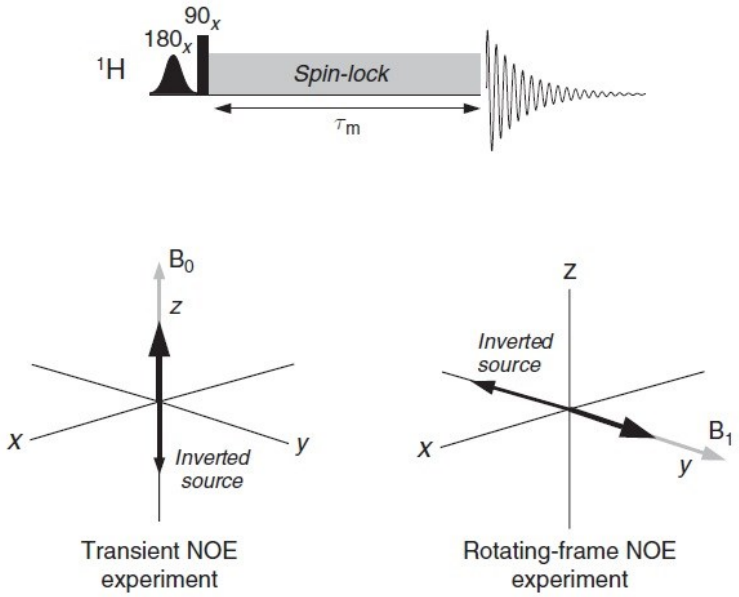

Figure 3. (Up) Scheme of the pulse sequence for observing rotating-frame NOEs. The ROEs develops during the long spin-lock pulse that constitutes the mixing period τ_m. (Down) Situation of the magnetization during the spin-lock. The rotating-frame NOE experiment can be viewed as the transverse equivalent of the transient NOE experiment. *Source: Claridge 2009*[169]

2.1.3 Measuring internuclear distances

As it has been commented in the previous section, the NOE is the result of the dipolar coupling of two spins which are close in the space to each other, so that this is a distance dependent effect. However, the exact NOE dependency with distance and which NOE experiment we must carried out to obtain accurate distance measurements are not straightforward issues. Thus, for instance, steady-state NOEs cannot readily be translated into internuclear separation because they result from a balance between the influences of all neighbouring spins (only relative distances are obtained from these NOEs). It has also been shown that for molecules that exhibit negative enhancements, steady-state measurements may fail to provide any reliable information of spatial proximity, and here one is forced to consider the kinetics of the NOE. Thus, saturation of the target resonance for periods that are far less than those needed to reach the steady-state would allow some NOE to appear, which is then sampled. Repeating the experiment with progressively incremented saturation periods allows the build-up to be mapped. Owing to the use of shortened saturation periods, the enhancements observed with this method are termed truncated driven NOEs or TOEs. Although once popular, this experimental approach is rather less used nowadays and as such shall be considered no further. The more common approach to obtaining kinetic data is to instantaneously perturb a spin system not by saturation but by inverting the target resonance/s (i.e. inverting the population differences across the corresponding transitions) and then allowing the NOE to develop in the absence of further external interference. In this case, the

NOE is seen initially to build for some time but ultimately fades away as spin relaxation restores the equilibrium condition; these enhancements are thus termed transient NOEs. The measurement of transient NOEs gained widespread popularity, initially in the biochemical community, in the form of the 2D NOESY experiment, which remains an extremely important structural tool in this area and increasingly in the analysis of smaller molecules. The 1D transient NOE experiment, also referred to as 1D NOESY, is also widely used within the chemical community as a gradient-selected sequence capable of providing high-quality NOE spectra. Transient experiments, whether 1D or 2D, are more commonly used qualitatively as 'single-shot' techniques, providing an overview of enhancements within a molecule rather than being employed to map the growth of the NOE.

Unlike the steady-state enhancements, the transient enhancements are influenced by only a single internuclear separation (r) with r^{-6} dependence whilst the so-called *initial rate approximation* is valid. Under this approach, the two cross-relaxing spins initially behave as if they were an isolated spin pair and the growth of the NOE has a linear dependence on mixing time. As longer mixing periods are used, the relaxation of spin I begins to compete with cross-relaxation between I and S, so the build-up curve deviates from linearity and the NOE eventually decays to zero (**figure 4**). Thus, for the initial rate approximation to be valid, mixing times significantly shorter than the T1 relaxation time of spin I must be used. Only under these conditions is meaningful distance measurement possible. If, on the other hand, the goal is to qualitatively identify through-space correlations, as is more often the case in routine work, mixing periods comparable to T_1 provide maximum enhancements. Since transient NOEs develop in the absence of an external radiofrequency field, they tend to give rise to weaker positive values (38% maximum) than the steady-state effects (50% maximum; see **figure 2**), so careful choice of timing is crucial to the success of transient experiments with fast-tumbling molecules. With respect to the slow-tumbling molecules, owing to the domination of efficient cross-relaxation both the transient and steady-state experiments give rise to a maximum negative NOE of -100%.

The primary reason why transient NOE methodology is used, instead of the steady-state experiment, is the complications arising from the existence of spin diffusion for molecules within this regime (slow-tumbling). Thus, in this case there is also a demand for the use of short mixing periods.

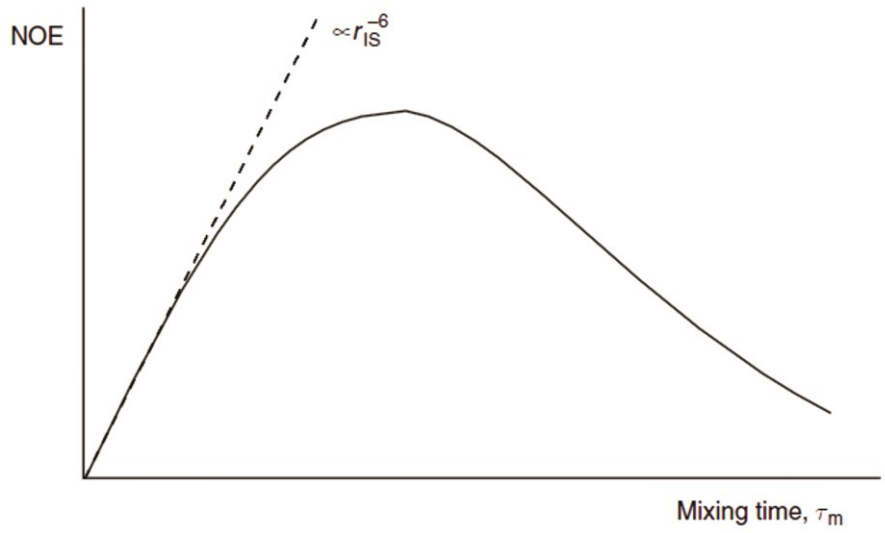

Figure 4. Development of the transient NOE between two spins as a function of mixing time (τ_m). *Source: Claridge 2009*[169]

Assuming the initial rate approximation to be valid (the NOE grows linearly), the magnitude of the enhancement of spin I after inversion of spin S ($\eta_I\{S\}$), after a period τ_m, will be proportional to the cross-relaxation rate, which in turn depends on r_{IS}^{-6}. Thus

$$\eta_I\{S\} = k\sigma_{IS}\tau_m = k' r_{IS}^{-6}\tau_m \qquad \textit{Eq. 3}$$

where the constants of proportionality k and k' contain the overall correlation time of the molecule, τ_c, in addition to a number of known physical constants. If τ_c is known, r_{IS} can be directly derived. However, whilst it is possible to determine τ_c from laborious relaxation time measurements, this is rarely done in practice, being more common to use a known internal distance as a reference. Therefore, if the reference NOE between nuclei A and B, of known internuclear separation r_{AB}, is also measured, then

$$\frac{\eta_I\{S\}}{\eta_A\{B\}} = \frac{r_{IS}^{-6}}{r_{AB}^{-6}} \qquad \textit{Eq. 4}$$

A direct comparison of the two NOE intensities thus provides the unknown internuclear distance. This simple relationship has been extensively used to provide measurements of internuclear separations, particularly in biological macromolecules. From a single experiment, distances can be estimated assuming that the *initial rate approximation* is valid for all interactions. This relies on all internuclear vectors in question possessing the same correlation time (i.e, isotropic tumbling), which of course is not real in all cases, e.g. when the internal motions are the main contributors to the reorientational properties of proton-proton vectors.

The *off-resonance* ROESY experiment

In an anisotropic molecule, the NOESY or ROESY based distances calculated using the *isolated spin pair approximation* using a single distance (and therefore a single correlation time) as reference are not accurate as they depend on the angle between the interprotonic vector and the molecular axis, that governs the correlation time of the vector. As an alternative method it was proposed to use several reference distances selected to cover the range of possible orientations with respect to the rotation axis.[85a-c] Thus, ranges for all the unknown distances could be obtained.[85a-c] However, in the carbohydrate field the accuracy needed to distinguish some of the characteristic features is usually lost (e.g. with heparin oligosaccharides). A more precise method is the *off-resonance* ROESY approach[172], which allows to calculate distances from relaxation data out from the *isolated spin pair approximation* (ISPA) by obtaining simultaneously the correlation time and the distance for each pair of protons. This method relies on calculating several *off-resonance* ROESY values by varying the tilted angle of the effective ROESY spin-lock field to achieve enough amounts of independent data as to extract the correlation time for each vector and from this to calculate the interprotonic distances. This procedure allows to extract simultaneously cross relaxation rates σ_{NOE} and σ_{ROE} and effective correlation times τ_c by measuring a linear combination of NOE and ROE effects controlled by the spin lock offset. In order to quantify the distances, arrays of several series of off-resonance ROESY experiments at several mixing times are recorded at different tilted angles (6, 10, 20 and 30 kHz), and all growth curves for each proton and each tilted angle are linearly fitted. Then, by calculating the growth rate of several series of *off-resonance* ROESY experiments corresponding to different spin lock offsets, the values of σ_{NOE}, σ_{ROE} and τ_c can be independently obtained for each proton pair.

2.2 Ligand-based NMR spectroscopy for binding studies

A critical step in drug discovery is the identification of high-affinity ligands for specific macromolecular targets. This is why, over the last 10 years, NMR spectroscopy has become a powerful tool in the pharmaceutical industry. Furthermore, significant improvements have been carried out in the field of NMR instrumentation in recent years, e.g., digital recording, cryogenic probes, auto-samplers, or higher magnetic fields, allowing to shorten the time to acquire data and to improve the spectral quality. In addition, new experiments and pulse sequences make a vast amount of information available for the drug discovery process. All these techniques take advantage of the fact that upon complex formation between a target molecule and a ligand, significant perturbations of specific NMR-sensitive parameters of either the one or the other can be observed and used to qualitatively detect ligand binding, or quantitatively determine binding affinity or characterize the structure of the complex. Furthermore, some of the techniques allow the identification of either the binding site on the receptor, or which part of the ligand is responsible for interacting with the target.

In the context of protein-glycan interactions, their kinetics properties (typically medium-weak binding affinities) make them very suitable to be studied by NMR techniques based on the observation the ligand (i.e., glycan). Therefore, STD-NMR and transferred-NOESY experiments represent very robust and powerful ligand-based NMR techniques to study, at atomic resolution, the hot spots of glycan-receptor interactions (ligand binding epitope) and the conformation of the carbohydrate in the bound state, respectively.

The suitable performance of STD-NMR and tr-NOESY techniques rely on the study of transient interactions of a weak-binding ligand to a macromolecular receptor, through the observation of intra- and/or inter-molecular NOE enhancements, such that the residence time τ_{res} of the ligand in the free state is much higher than that of the bound state. If binding is too tight, magnetization is lost by the efficient T1 relaxation typical of macromolecules, precluding NOE to develop prior to the acquisition period. Thus, the different timescales playing a role in this type of experiments have to be considered to fully understand what is going on and thus to correctly interpret the results (**figure 5**).

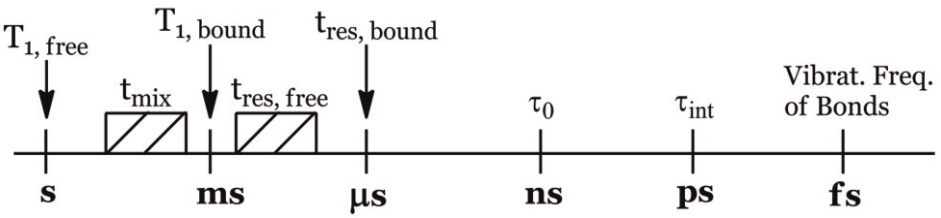

Figure 5. Approximated scheme of the different timescales in which the distinct mechanism taking place during molecular recognition of low-affinity ligands to macromolecular receptors. Note the important differences between $T_{1,\,free}$-$T_{1,bound}$ and $t_{res,free}$-$t_{res,bound}$, allowing the observation of the intermolecular NOEs.

2.2.1 The equilibrium kinetics of binding: the fast exchange approximation

Assuming a protein receptor P with a single binding site for the ligand L, such that they interact to transiently form the complex PL, the binding equilibrium can be described by a model of two states

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

with a temperature dependent dissociation constant K_D

$$K_D = \frac{[P]\cdot[L]}{[PL]} = \frac{k_{off}}{k_{on}} \qquad Eq.\ 5$$

were the unimolecular dissociation rate constant k_{off} is inversely related to the lifetime of the protein-ligand complex or, equivalently, the residence time of the ligand in the bound state t_{res} (**figure 5**), while the bimolecular rate constant k_{on} is the probability of a productive interaction between the protein and the ligand. At any given time, the fraction of bound protein f_{PB} is given by

$$f_{PB} = \frac{[PL]}{[P]+[PL]} \qquad Eq.\ 6$$

Combining equations 5 and 6 yields

$$f_{PB} = \frac{[L]}{[L]+K_D} \qquad Eq.\ 7$$

This is one of the fundamental equations that drives the outcome of the NMR experiment. It shows that increasing the ligand concentration [L], the fractional occupation of the receptor-binding site f_{PB} will rise in a hyperbolic function of [L]. In the limit [L] << K_D, the fractional population of bound protein is directly proportional to the concentration of the free ligand. In the other extreme, [L] >> K_D, the protein receptor is completely saturated by the ligand (f_{PB} = 1). An interesting situation occurs when [L] = K_D, that is, the receptor is half-saturated, which means that half of the receptor molecules exist in a one-to-one complex with the ligand. Ligands with weaker affinity have larger K_D and so, to saturate the receptor-binding site, the addition of more ligand molecules is required than in the case of a ligand with higher affinity. Thus, by adjusting the receptor and ligand concentrations, it is possible to "select" the maximum K_D for an optimum observable NMR signal. This tuning of detection threshold is very useful in ligand-based screening experiments (STD NMR in particular) of large libraries of compounds.

In the free state, both receptor and ligand retain their intrinsic NMR parameters (e.g. chemical shifts, relaxation rates, translational diffusion coefficients). In each other's presence, the mutual binding affinity of ligand and receptor drives an exchange process that toggles both sets of molecules between

the free and complexed states. At equilibrium, the transient presence of the ligand molecule at the protein site is responsible for chemical exchange, characterized by the exchange kinetic constant k_{ex}

$$k_{ex} = k_{on} \cdot [P] + k_{off} \qquad Eq.\ 8$$

All the ligand-based NMR screening experiments assume that the receptor-ligand binding is in the *fast exchange limit*[3]. Indeed, this is a necessary condition for ligand-based NMR techniques to be useful, as in the fast exchange limit the equations are enormously simplified, and the total effect of the protein-ligand complex on the NMR parameter is just the weighted sum of its separate contributions (weighted average). Thus, if Q is the NMR parameter, then

$$\langle Q \rangle = P_B Q_B + P_F Q_F \qquad Eq.\ 9$$

where ⟨Q⟩ is the observed exchange-averaged parameter between its free (F) and bound (B) states, with respective populations P_F and P_B. Observed differences between ⟨Q⟩ and Q_F provide measurable evidence of receptor binding and often an indicator of a "hit" in screening of a large library of compounds. The ability to detect binding with adequate sensitivity depends critically on the bound state contributions (P_BQ_B) being significantly larger than those from the free state (P_FQ_F). Since typical screening conditions where $[L]_T >> [P]_T$ are used automatically make $P_B << P_F$, it is therefore necessary that ⟨Q⟩ is amplified in the bound state (i.e. $Q_B >> Q_F$). On the other hand, if k_{on} is well approximated by a diffusion-limited value (10^7-10^9 $M^{-1}s^{-1}$), then the slowest k_{ex} constants lie within the $10^3 < k_{ex} < 10^5$ s^{-1} range for weak-affinity ligands (K_D in the μM range). Since ligand-based NMR screening methods are primarily ^1H based, k_{ex} would consequently exceed most differences in intrinsic ^1H relaxation rates and rotating frame precession frequencies, thus supporting the validity of the fast exchange assumption.

2.2.2 Transferred-NOESY experiment (tr-NOESY)

The theory of transferred NOE has been described in numerous publications[173] and several excellent reviews have been written[174].

Since the kinetics of the majority of glycan-protein interactions is fast and with dissociation constants within the μM to mM range, they represent ideal systems for the observation of transferred NOE (tr-NOE), and so, many studies of this type have been reported on carbohydrate-protein systems[175]. On the contrary, too tight binding places the systems outside the range of fast exchange on the NMR time scale and no transferred NOEs are observed.

[3] The *fast exchange limit* implies fast exchange in both the chemical shift and the relaxation time scales. This means that the considered event occurs much faster than both NMR observables. In particular, three conditions must be fulfilled: $k_{ex} \gg \Delta\omega$, $k_{ex} \gg \Delta J$, and $k_{ex} \gg \Delta T_2^{-1}$.

Importantly, to observed transferred-NOEs the following inequality has to be fulfilled

$$|N_b \cdot \sigma_b| \gg |N_f \cdot \sigma_f| \qquad \textit{Eq. 10}$$

where σ is the cross-relaxation rate and N the number of molecules in the free (N_f, σ_f) and bound (N_b, σ_b) states. Thus, the intensity of transferred-NOEs strongly depends on the excess of ligand over protein, being maximum for an optimum ligand-to-protein ratio ($[L]_T/[P]_T$) that is usually between 10 and 30 to 1. Therefore, the preparation of the sample is of key importance in this type of ligand-observed NMR experiment. The reason is that tr-NOESY depends on both fractions of free and bound ligand, with the former and the latter giving rise to positive and negative NOE peaks, respectively. Thus, if the ligand-to-protein ratio is too high, there will be an elevated fraction of free ligand in solution and, as a result, its corresponding positive NOE may lead to a significant reduction or even cancelation of the tr-NOESY enhancements coming from the negative NOE developed by the small fraction of bound ligand.

The observation of tr-NOE relies on the existence of rather different correlation times τ_c for the free and bound ligand. Small molecules (free ligands) are usually low-to-medium molecular weight molecules and therefore have short correlation times, slow NOE build-ups and no spin diffusion, and, as a consequence, exhibit small positive NOEs. On the other hand, when a ligand is bound to a protein receptor, it acquires the motional properties of the macromolecule during the residence time in the bound state, thus exhibiting large correlation times τ_c, rapid NOE build-ups, extensive spin diffusion, and strong negative NOEs (i.e. transferred NOEs)[174a, b]. Thus, binding of a ligand to a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs. Furthermore, the discrimination between tr-NOEs originating from the bound state and NOEs of the ligand in solution can also be achieved by the build-up rate, that is, the time required to achieve maximum intensity, which for tr-NOEs is in the range of 50 to 100 ms, whereas for small, non-binding molecules it is four- to ten-times longer (**figure 6**). Therefore, the maximum enhancement for tr-NOEs is observed at significantly shorter mixing times τ_m than for isolated small molecules in solution. Various experimental implementations have been explored in the last two decades, ranging from 1D selective steady-state experiments to 1D and 2D transient NOE experiments[57a, 174a]. Transferred NOESY is a very useful and widely used experiment to quickly determine binding activity of single ligands and ligand libraries, and, at the same time, it provides conformational information on the bound ligand (intra-molecular tr-NOEs), as well as on the orientation of this in the binding pocket (inter-molecular tr-NOEs), with the advantage that the information is measured from the easily observed and assigned free ligand signals. However, there are several precautions that need to be taken to ensure that the data are realistic.

One of the major drawbacks of this experiment is the possible existence of spin diffusion effects, which are typical for large molecules. In this case, apart from direct enhancements between protons close in space, other spins (including those of the receptor) may mediate the exchange of magnetization. These indirect (protein-mediated) tr-NOE effects can give rise to negative cross peaks between protons that are far apart, which may lead to interpretation errors in the analysis of the ligand bound conformation.

To avoid it, one possibility is to use short mixing times. In addition, the tr-ROESY experiment has been proposed[176] to distinguish direct from indirect NOE cross peaks.

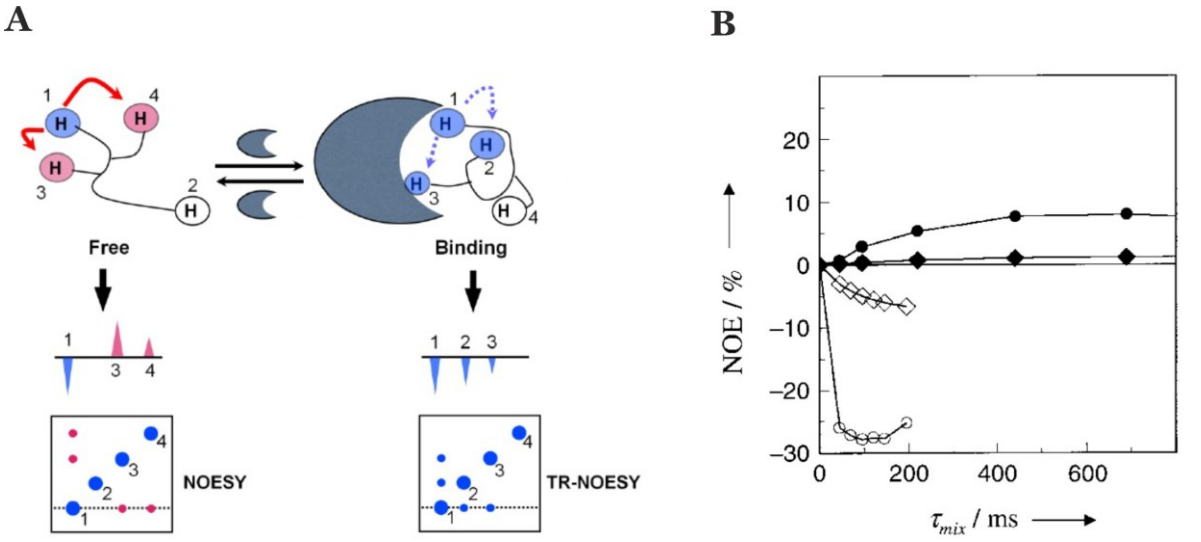

Figure 6. (**A**) Shematic representation of a NOESY (left) and tr-NOESY (right) spectra. Cross-peaks are of the opposite sign to the diagonal peaks (positive NOEs) for a small molecule in the free state. Upon addition of the receptor, a sign change of cross-peaks takes place (same sign as the diagonal peaks; negative NOEs). (**B**) Nuclear Overhauser enhancements (NOEs) and tr-NOEs for α-L-Fuc-(1→6)-β-D-GlcNAc-OMe in the absence (filled symbols) and presence (open symbols) of *Aleuria aurantia* agglutinin, measured at 600 MHz as a function of the mixing time τ_m. Circles and diamonds refer to proton pairs H6proR^{GlcNAc}-H6proS^{GlcNAc} and H1^{Fuc}-H6proS^{GlcNAc}, respectively. *Sources: Doctoral Thesis of Cinzia Guzzi[177] (A) and Claridge 2009[169] (B).*

The setup of transferred NOE experiments is identical to the setup of "normal" NOE experiments. The only difference is the preparation of the sample since the intensity of transferred NOEs strongly depends on the excess of ligand over protein. Applied to glycan systems, depending on the size of the carbohydrate ligand, three regimes may be distinguished:

(a) The molecular weight of the carbohydrate ligand leads to correlation times ranging in the order of tens to hundreds of picoseconds, and therefore NOEs of the free ligand are positive. At 500 MHz, this is usually the case up to the size of trisaccharides. If charges are present as, for example, in sialic acid residues, the tumbling of the molecule is slower and one may observe negative NOEs already for a trisaccharide.

(b) If the molecular weight is such that the correlation time approaches *zero crossing* conditions, no NOEs will be observable. For uncharged carbohydrates at 500MHz, this is usually the case for tetra- and pentasaccharides.

(c) Larger carbohydrates have correlation times of several nanoseconds and, therefore, display negative NOEs at frequencies of 500 MHz and higher.

In cases (a) and (b), the discrimination of transferred NOEs from free ligand NOEs is straightforward because at carbohydrate-to-protein ratios in which the e*quation 10* is fulfilled, the sign of the NOE changes upon binding from positive to negative. At the same time, the mixing time at which a maximum NOE is observed is reduced and in the range of 200 ms, as compared to 600-1000 ms for the free ligand. Because of this change in sign, the experiment has also been used to identify binding in mixtures of low molecular weight compounds[178]. In case (c), discrimination is less straightforward and usually requires the acquisition of NOESY experiments with different mixing times.

2.2.3 Saturation Transfer Difference spectroscopy (STD)

The STD NMR experiment[179] is another spectroscopic technique to study the interactions, in solution, between a large molecule (receptor) and a medium-small sized molecule (ligand), and, alike tr-NOESY, it is based on the Nuclear Overhauser effect and the observation and analysis of the resonances of the ligand protons.

The experiment is carried out by first registering a spectrum under conditions of thermal equilibrium with the irradiation frequency set at a value that is far from any ligand or protein signal (e.g. 40 ppm), i.e, the so-called *off-resonance* spectrum (**figure 7, top**), which is used as reference with signal intensities I_o. A second experiment is then recorded, in which the protein is selectively saturated (*on-resonance* spectrum; see **figure 7, middle**), giving rise to ligand signals with I_{sat} intensities. In general, the selective irradiation consists of a cascade of Gaussian-shaped pulses (low power) that saturate only a region of the spectrum that contains a few protein resonances (but not ligand signals), e.g., the aliphatic (from 0 to -1 ppm) or aromatic region (around 7 ppm), for a specific period of time (saturation time; typically from 0.5 to 5-6 seconds). The selective saturation is transferred to the whole protein via spin diffusion through the vast network of intra-molecular ^1H-^1H cross-relaxation pathways (intra-molecular NOE; see **figure 7, middle**), being a quite efficient processes due to the typical large molecular weight of the receptor. Also, saturation is transferred from the protein to the bound ligand via spin diffusion through inter-molecular NOEs. The dissociation of the ligand will then transfer this saturation into the bulk solution where it accumulates during the saturation time of the experiment, as a result of the much slower relaxation in the unbound that the bound state. In particular, as in fast-exchanging protein-ligand systems the enthalpic relaxation (R1) of fast-tumbling molecules (small) in the free state is much slower than the kinetic off-rate constant of binding ($k_{off} >> R_1$), the accumulation of ligands molecules containing some of their resonances perturbed (NOE of large molecule) results in the macroscopic detection of transferred saturation on the ligand signals in the saturated STD NMR spectrum (I_{sat}; **figure 7, middle**).

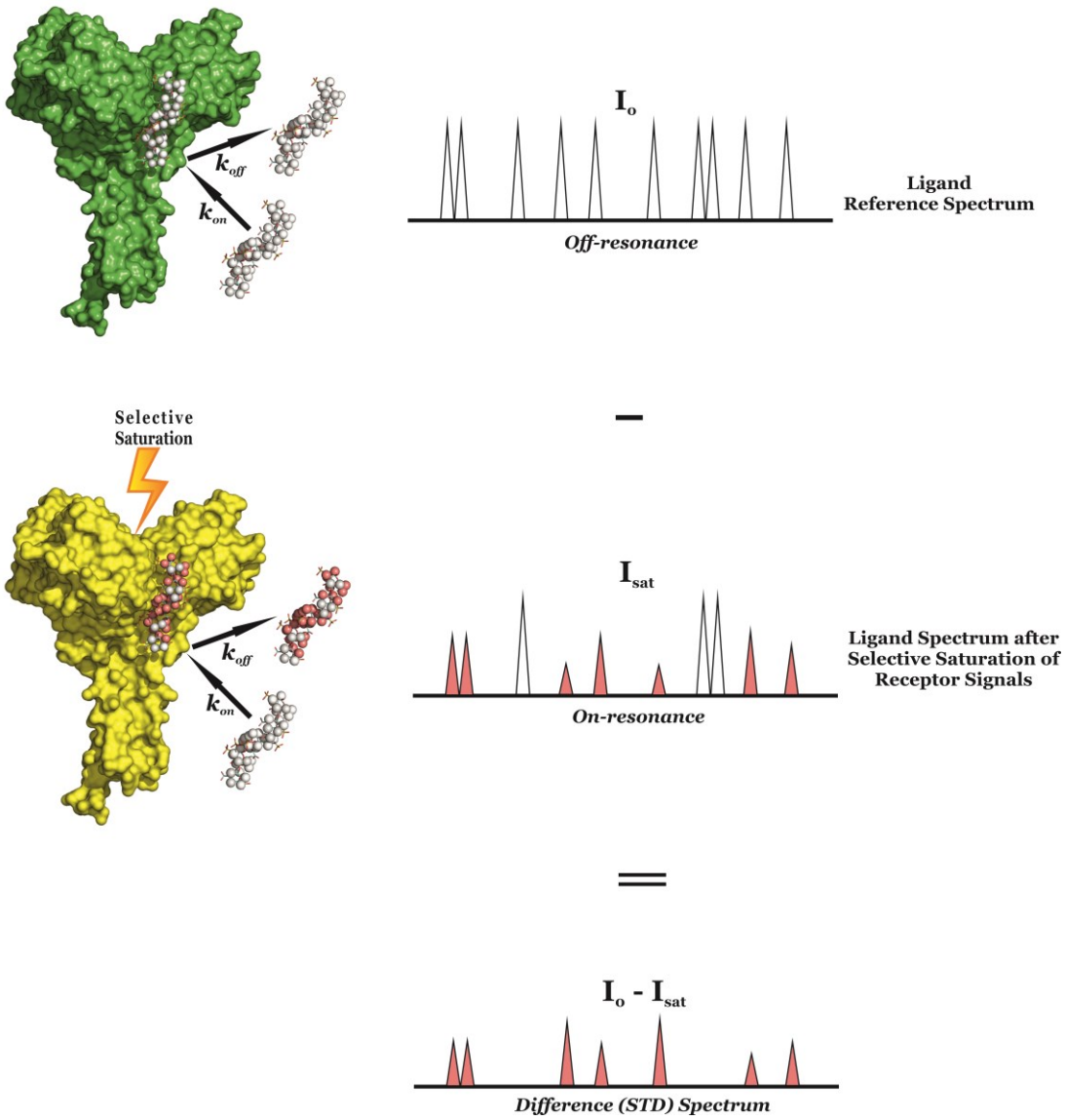

Figure 7. Scheme of the STD-NMR experiment showing the protein in surface representation and the non-exchangeable protons of the ligand as spheres. (Top) A 1D standard NMR experiments show only equilibrium intensities of the ligand in the free state (I_o). (Middle) Upon selective saturation of some receptor signals, this is efficiently spread throughout the protein (yellow surface) by spin diffusion (intra-molecular NOEs). The fast exchange (transient binding) between the free and bound ligand states allows the transfer of magnetization (inter-molecular NOEs) from the receptor to the ligand protons in contact with the protein surface (salmon spheres, *on-resonance* spectrum). (Bottom) The difference spectrum (I_o-I_{sat}) only contains the ligand signals perturbed upon binding, whose intensities reflect the proximity of each proton to the protein surface.

Furthermore, for those hydrogen atoms of the ligand establishing close contacts to the protein surface (4-5 Å) these I_{sat} values will be lower than the I_o intensities, i.e, negative inter-molecular NOE, due to the transfer of the relaxation properties of the macromolecule to the small ligand in the bound state (**figure 7, middle**). By subtracting the *off-resonance* from the *on-resonance* spectrum (I_o-I_{sat}) the difference or STD spectrum is obtained, which will just contain the proton signals of the ligand in close contact to the protein surface (**figure 7, bottom**), and where any signal coming from non-binding

compounds is cancelled out. So, if a non-binder is present in solution its resonances will not appear in the STD spectrum. The signal intensities exclusively coming from the saturation transfer can be quantified ($I_{STD} = I_o - I_{sat}$), showing the proximity (high I_{STD}) or distance (low I_{STD}) of the proton to the receptor surface. Also, a blank experiment must be carried out to assure the absence of direct irradiation of the ligand.

A sample containing the receptor at low concentration and a large molar excess of the ligand (1:50 up to 1:1000) is usually employed in STD NMR experiments. This precludes the perturbations of absolute STD intensities due to rebinding effects (i.e, a ligand already saturated experiences another association process, without previous full relaxation), which would impede to correctly determine the group epitope mapping (**figure 8**).

The theory of NOE indicates that the magnetization transferred from receptor to ligand protons by intermolecular NOE depends on the inverse sixth power of their distances in the bound state. Thus, the shorter the protein-ligand proton-proton distance (bound state), the stronger the intensity of the corresponding STD signal. So, by normalizing all the measured STD intensities ($I_o - I_{sat}/I_o$) against the most intense signal (which is arbitrarily assigned a value of 100%), the so-called "group epitope mapping" is obtained (expressed as percentages; see **figure 8**). This represents the fingerprint of protein-ligand contacts in the bound state, such that it illustrates which chemical moieties of the ligand are key for molecular recognition in the binding site.

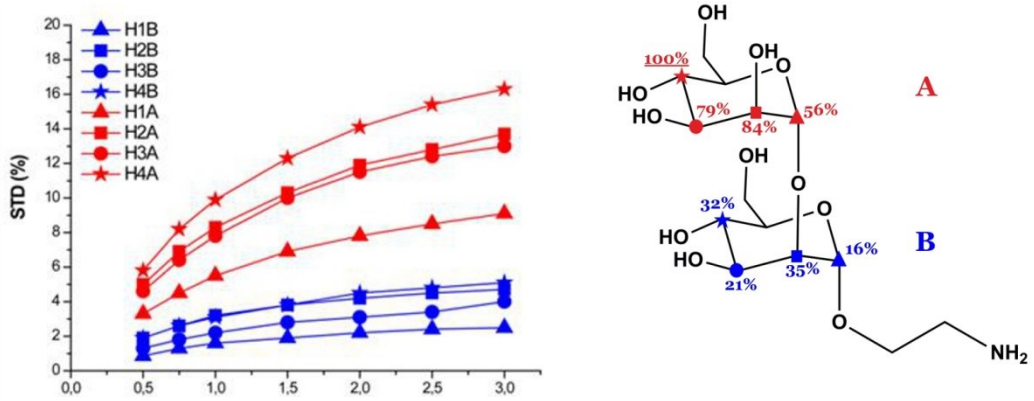

Figure 8. (Left) STD growth curves (absolute values) for the recognition of Man-α(1,2)-Man-α-O(CH_2)NH_2 by the anti-HIV-1 human antibody 2G12. (Right) Ligand group epitope mapping, or binding epitope for Man-α(1,2)-Man-α-O(CH_2)NH_2 compound. Note that proton H4A receives the highest saturation (thus the 100% is arbitrarily assigned) and is used as reference to calculated the STD percentages of the other protons. Residue of mannose-A makes the main contacts with the protein in the bound state. *Sources: Doctoral Thesis of Pedro M. Enríquez-Navas[180], and Enríquez-Navas et al. 2011[175c].*

Binding epitopes were commonly obtained for a given saturation time, assuming that the resulting ligand group epitope mapping did not depend on the chosen saturation time. However, significantly different R_1 relaxation rates of the ligand protons can produce artefacts in the epitope definition[181]. In

particular, protons with slower R_1 relaxation enable a more efficient accumulation of saturation in solution such that their relative STD intensities may be significantly overestimated at long saturation times, thus also overrating the proximity of those protons to the protein surface. Indeed, the structual information that the binding epitope provides can be affected by 1) differences in R_1 relaxation rates of ligand protons, 2) the extent of saturation received in the first place, and 3) the kinetics of binding. Since these sources of distorsions are the consequences of differences in the ability to accumulate saturation in the free state, we can cancel them by deriving STD intensities close to zero saturation time, this is, when virtually no accumulation of saturated ligand molecules is taking place. This is usually referred to as initial slopes of the build-up curves (**figure 9**). In addition, under this approximation, possible artefacts comming from intra-molecular spin diffusion (bound state) can also be minimized. To calculate the initial slopes, Mayer and James proposed fitting the experimental build-up curves to the mono-exponential function[182].

$$STD(t_{sat}) = STD^{max}\left(1 - exp(-k_{sat} \cdot t_{sat})\right) \qquad Eq.\ 11$$

where $STD(t_{sat})$ is the observed STD intensity, STD^{max} is the asymptotic maximum of the build up curve, t_{sat} the saturation time, and k_{sat} the rate constant related to the relaxation properties of a given proton that measures the speed of the STD build-up. k_{sat} and STD^{max} are derived by least-squares fit, and the initial slope (STD_0) of the curve is obtained as

$$STD_0 = \left.\frac{\partial STD(t_{sat})}{\partial t_{sat}}\right|_{t_{sat}\to 0} = STD^{max} \cdot k_{sat} \qquad Eq.\ 12$$

These STD_0 values are then used to characterize the binding epitope independently of T_1 and rebinding effects. Note that for $t_{sat} = {}^{1}/_{k_{sat}}$

$$STD(t_{sat}) = STD^{max}(1 - e^{-1}) = 0.63 \cdot STD^{max} \qquad Eq.\ 13$$

Thus, if we plot the normalized STD factor versus the saturation time (t_{sat}), k_{sat} can be approximated as the inverse of the interpolated t_{sat} value from a STD factor of 0.63 (**figure 9**).

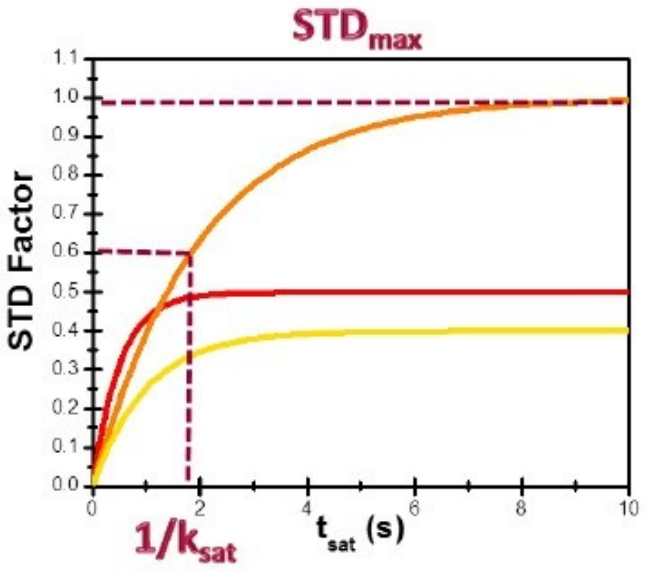

Figure 9. Normalized STD build-up curves fitted to a mono-exponential function to obtain the initial slope (STD_0) from the multiplication of STD^{max} and k_{sat}. By interpolating a value of 0.63 for the STD^{max} factor, the k_{sat} is obtained as the inverse of t_{sat}.

To get protein-ligand association curves from STD NMR experiments, Mayer and Meyer introduced the conversion from observed experimental intensities (I_0-I_{sat}/I_0), which depend on the fraction of bound ligand f_{LB}

$$f_{LB} = \frac{[PL]}{[L]_T} = \frac{[PL]}{[L]+[PL]} = \frac{[P]}{K_D + [P]} \qquad Eq.\ 14$$

, to STD amplification factors (STD-AF) by multiplying the observed STD by the molar excess of ligand over protein[183]

$$STD - AF = STD \cdot \frac{[L]_T}{[P]_T} \qquad Eq.\ 15$$

so that the STD amplification factor (STD-AF) depends on the fraction of bound protein (f_{PB}).

$$f_{PB} = \frac{[PL]}{[P]_T} = \frac{[PL]}{[P]+[PL]} = \frac{[L]}{K_D + [L]} \qquad Eq.\ 16$$

Therefore, a plot of STD-AF values at increasing ligand concentrations will give rise to the protein-ligand binding isotherm, from which the dissociation constant K_D can be derived. However, in *eq. 16* the concentration of free ligand [L] is not known. This can be solved by employing a very low total concentration of protein, such that $[P]_T \ll K_D$, and an excess of ligand. Under these conditions $[L] \approx [L]_T$, so that

$$f_{PB} = \frac{[L]_T}{[L]_T + K_D} \qquad \textit{Eq. 17}$$

and thus the dissociation constant K_D can be easily determined by plotting the normalized STD-AF values at increasing total ligand concentrations, i.e, the binding isotherm (**figure 10**). This can be done graphically by interpolation as the K_D coincides with the amount of free ligand $[L]_T$ that is necessary to reach a 50% of f_{PB} ($f_{PB} = 0.5$)

$$0.5 = \frac{[L]_T}{[L]_T + K_D} \Longrightarrow K_D = [L]_T \qquad \textit{Eq. 18}$$

Therefore, the weaker the interaction, the higher the ligand concentration required to saturate half of the protein molecules.

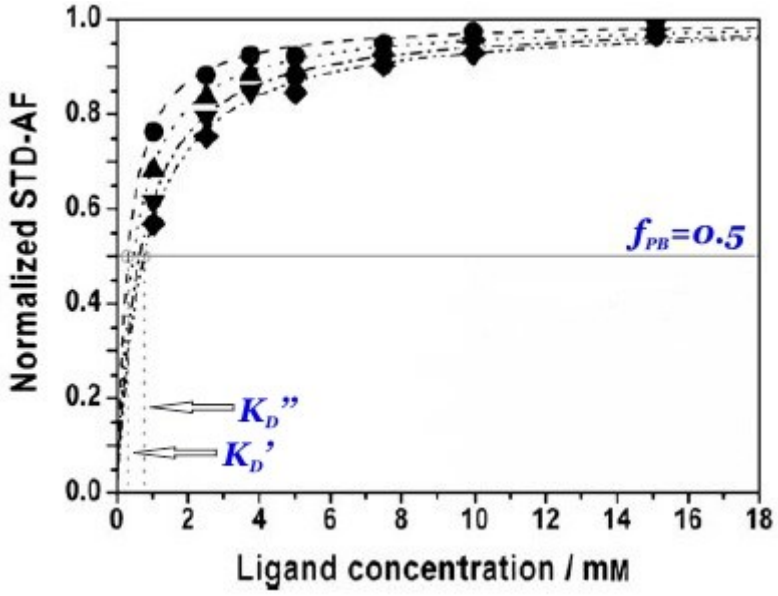

Figure 10. Example of STD binding isotherms, at different saturation times, of the WGA protein (46 μM) titrated with chitobiose. STD-AF values appear normalized against their corresponding plateau values. Note that different K_D values (K_D', K_D'') are obtained at different saturation times. *Adapted from Angulo et al. 2010*[184].

The determination of K_D from STD-NMR experiments is affected by different experimental parameters, as it has been thoroughly studied in our group[184]. In particular, the saturation time (t_{sat}) employed (**figures 10 and 11**), the STD intensity of the signal (**figure 11**) and the fraction of bound ligand are key factors that must be chosen wisely for the accurate determination of dissociation constants. The apparent K_D increases monotonically with t_{sat}, thus understimating the protein-ligand affinity. Similarly, the higher the STD intensity of a resonance the more overstimated K_D value will be obtained. Also, if the fraction of bound ligand is modified by increasing the receptor concentration, the apparent dissociation constant will be larger.

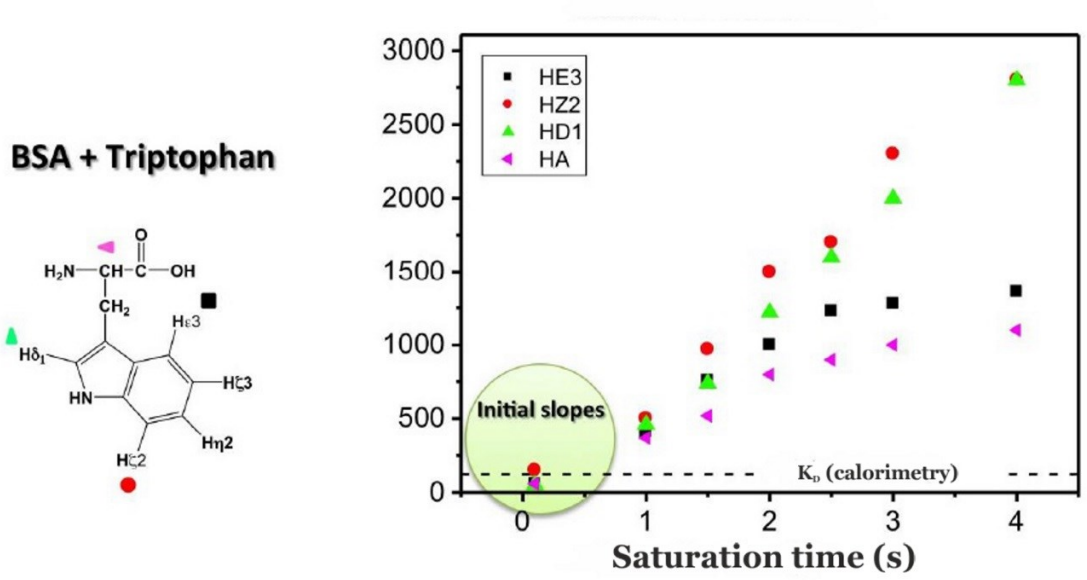

Figure 11. Effect of the experimental factors (saturation time and monitored proton) on the determination of the apparent binding constant for the BSA-L-tryptophan system on a sample of 20 μM of BSA.

2.3 Molecular modelling

A force field consists of the combination of a mathematical formula and associated parameters that are used to describe the overall potential energy of a molecular system as a function of its atomic coordinates. Currently, it is widely accepted that force fields are critical to molecular simulation in many aspects of life sciences research. Understanding, analysing, and predicting 3D structural models of molecular systems, including their conformations, binding affinities and related properties, all depend on accurate atomic force fields. For this reason, there has been a great deal of effort devoted to the development and improvement of potential energy functions and their parameters, which are the two features that define a force field.

Energy minimizations and MD simulations are often limited by inadequate description of the various force field parameters for the systems of interest. For instance, if a crystal structure is minimized without including penalty terms for the structure factor, then the deviations from the experimental structure are often much larger (0.5 to 1.5 Å) than the expected error[185], reducing confidence in molecular mechanics analyses. Thus, the key factor affecting the quality of MD simulations is the force field accuracy, which determines the goodness of the conformational sampling and dynamics performance in reproducing the experimental observables.

Computational methods such as docking and molecular dynamics (MD) simulations (also homology modelling and computational mutagenesis) provide complementary tools, indispensable in many cases, to fully understand both X-ray and NMR data. Importantly in the context of the present thesis,

they perform particularly well (specially MD simulations) in characterizing the structure and dynamics of glycans and glycoconjugates[186] (see next section).

The force fields commonly employed consist of a combination of bonded and non-bonded energy terms[187]. Given atomic positions and velocities, forces are calculated on the fly, as derivatives of the potential energy (V (r_1, . . ., r_N)), at specific time steps. The overall potential energy for a molecular system can be written in classical mechanics as

$$V_{(r_1,\ldots,r_N)} = \sum_{Bonds\ i,j} K^B_{i,j}(r - r_{eq})^2 + \sum_{Angles\ i,j,k} K^A_{i,j}(\theta - \theta_{eq})^2$$

$$+ \sum_{Dihedrals\ i,j,k,l} \sum_{n} \frac{V_n}{2}[1 + cos(n\varphi - \gamma_n)]$$

$$+ \sum_{Coul\ i,j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{LJ\ i,j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right]$$

Eq. 19[63]

where the parameters shown in red must be known from experiments or derived from QM calculations, and included in the force field. of Σ_{bonds}, Σ_{angles} and $\Sigma_{dihedral}$ terms refer to the potential energy associated with bond-stretching, angle-bending, and proper (and improper) dihedral angle rotations, respectively, whereas Σ_{Coul} and Σ_{LJ} represent the pairwise electrostatic interaction and the Lennard-Jones (LJ) repulsion-dispersion potential energy terms, respectively. Thus, classical force fields are defined by both the functional form of the different terms contributing to the global potential energy and by the set of parameters that each term requires. The different energy terms are given by empirical formulae and/or harmonic functions penalizing deviations from ideal values, these being determined from high-resolution crystallographic or spectroscopic data and/or from calibration to QM calculations[187a, b, 187d, 188]. The accurate determination of these empirical parameters, such that when introduced in *equation 19* lead to the correct potential energy landscape of the molecular system, is a crucial, meticulous and challenging task in force field development. Furthermore, due to the coupling between many of the force field terms (e.g. torsions and electrostatics), the parameterization process inevitably requires testing multiple sets of calculations for optimization. In this regard, the better the force field refinement protocol, the more probably the resultant parameters will be broadly applicable[189].

It is well known that hydrogen bonds formation is the driving force to many phenomena, including the generation and stabilization of secondary structures[190], protein folding and stability[191], molecular recognition[192], and drug binding and enzymatic reactions that involve transfer of protons[193]. Therefore, the detailed understanding of hydrogen bonds geometry and their incorporation into accurate potential functions is of fundamental importance, although many efforts in this direction still have to been done.

The resolution of X-ray crystallography data for proteins rarely covers beyond 1.0 Å. For this reason, studies of hydrogen bonds in proteins have been mostly limited to the coordinates of non-hydrogen atoms. Thus, the hydrogen bond selection criteria used in some of these studies are based on the distance between the potential donor and acceptor atoms[194]. However, the hydrogen bond geometry can be better understood in terms of the angle and distances involving the positions of hydrogen[195], even if the hydrogen positions are modelled only implicitly from their heavy atom neighbours. For these reasons, several strategies have been used to account for hydrogen bonding in crystallographic refinement and molecular simulations. The hydrogen bond potential is often implicitly parameterized as a combination of Lennard-Jones (L-J) and electrostatic terms. In force fields that use an explicit hydrogen bonding term, this is typically included as a distance-dependent function without any directional component. The functional form may be a L-J 6–12[196], a L-J 10–12[187a, b, 197], a L-J 6–9[196a, 198], or a Morse type potential[199]. On the other hand, CHARMM[187b] and MM3[200] force fields now include a cosine directional term. In these implementations, the hydrogen bond energy is minimized when the hydrogen bond N-H•••O is linear, with an angle of 180°. However, these do not reflect the non-linear directional preferences of hydrogen bonds at the acceptor molecule, conferred by their covalent component[194a, 195, 201]. In any case, it has to be noted that displacement of water molecules competing for hydrogen bonds is not accounted for in any force field.

2.3.1 Modelling of glycans

As it has been commented in *Chapter 1*, carbohydrates encode an amount of potential information that is several orders of magnitude higher than in the case of any other biological macromolecule, and this complex encoding capacity arise from their enormously diverse (sequence), complex (ramifications) and flexible (local and/or global conformation) structures. This structural complexity makes the characterization of the 3D structure of oligosaccharides, their conjugates, and analogues, particularly challenging for traditional experimental methods. Thus, computational methods provide a basis for interpreting sparse experimental data and for independently predicting conformational and dynamic properties of glycans, which eventually can contribute unique insights into the relationship between oligosaccharide structure and biological function (**figure 12**).

In the early years of biomolecular modelling, MD simulations were technically limited to small biological systems, e.g., small proteins[202], short DNA helices[203] and mono-[204] or disaccharides[205], and for short simulation times. Nevertheless, the growing evidences of the dynamical nature and fundamental role on biological functions of biomolecules acted as a catalyser for the development of computer modelling tools applied to structural biology[206]. Today, advances in computer technology and software algorithms enable us to sample the conformational space and dynamics of biomolecular systems for simulation times that go from hundreds of nanoseconds (e.g. most internal motions in glycans[6a, 207] to several microseconds (e.g. timeframe for the conformational equilibrium of the iduronate ring in GAGs[17]). However, it should be noted that very long timescales are not always necessary to obtain useful information. For instance, not too long MD simulations can be very

effectively employed in the refinement and rescoring of ligand–protein complexes generated from ligand docking[208], providing some level of ensemble averaging to aid in generating robust affinity and specificity predictions.

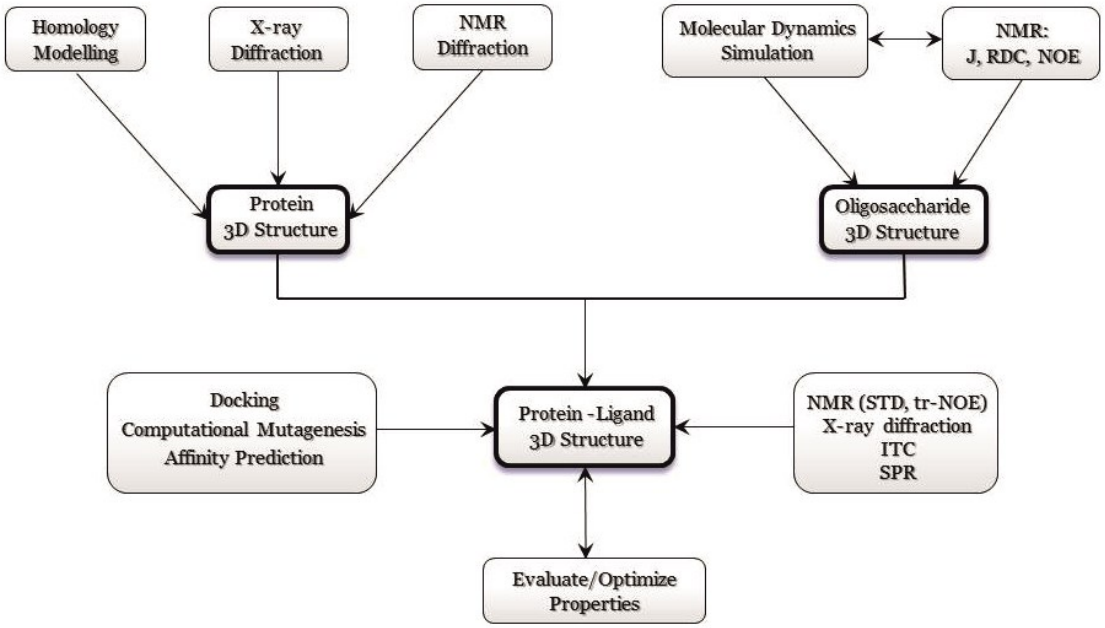

Figure 12. Scheme of the interplay between biophysical and biochemical tools and computational methods (bold outlines). The appropriate combination of techniques may eventually lead to solid structure-function relationships for glycans and carbohydrate-protein complexes. *Adapted from Woods and Yongye 2012*[209].

Because of the pivotal role of carbohydrate–protein interactions in human biology, during the last years a considerable interest has grown up in employing computational simulations to help the structural characterization of these systems and aid in the rational design of new therapeutics[210] and vaccines[211]. Therefore, the potential of this field has promoted in the past decade the development of several force fields specific for carbohydrates, enabling accurate computer simulations. In particular, the compatibility of these glycan-specific force fields (e.g. GLYCAM) with large biomolecular force fields (e.g. AMBER) enables us today to study complex biomolecular systems, such as glycolipids and glycoproteins. For instance, large-scale computational studies of carbohydrate interactions at the cell surface are now feasible. Moreover, given a sufficiently accurate force field, computational methods could provide a fundamental aid in both carbohydrate-based lead discovery and optimization, thus avoiding the synthesis and screening of large numbers of compounds. Interesting examples of automated docking in combination with carbohydrate-specific parameters, and with free-energy functions designed specifically for carbohydrate–protein interactions, are emerging[212].

2.3.2 Carbohydrates force fields

From the standpoint of molecular modelling techniques, the accurate study of carbohydrate structures and physical properties requires glycan-specific parameters and partial charges, and molecular mechanics potential energy functions, this is, a force field specific for glycans. The accurate reproduction of the electrostatic properties of a molecule is essential to model intermolecular interactions of polar molecules such as GAGs; however, the low number of charge derivation schemes reported in literature to date puts of evidence the complexity of this requirement. Furthermore, it should be noted that though the concept of partial charge is artificial and somehow weak, and that there is no definite approach to their assignment, it is very convenient in practice.

The parameter development protocol is specific to each force field and small differences in the parameter sets can lead to significant differences in the energy landscape, such as location and depth of minima[213]. Mixing parameters (this include the partial charges) from different force fields, can therefore result in a loss of internal consistency and, consequently, in erroneous simulations. This is profoundly important for flexible molecules such as oligosaccharides, the state populations of which are particularly sensitive. Thus, the development of a robust carbohydrate force field is a particularly challenging task because of the need to consider the influence of the inherent flexibility of glycans on the approach to parameter development and validation. In addition, unlike proteins and oligonucleotides, glycans present frequently branched structures. Furthermore, the majority of carbon centres in a monosaccharide are chiral and bear a hydroxyl group, which can form a glycosidic link to another carbohydrate unit. Thus, while only one dipeptide can be generated from the same two amino acids, 20 chemically distinct disaccharides can be formed from the same two hexopyranose monosaccharides. This gives an idea of the particularly important difficulties that parameter development for carbohydrate force fields entails.

Force fields well suited for glycans systems have been designed aiming to reproduce the particular physicochemical factors that determine their global structural properties in solution. In this regard, the accurate treatment of the 1–4 non-bonded interactions (i.e., those between the two outermost atoms in a dihedral angle) in glycans is crucial, since the so called “particular flexibility of carbohydrates” constantly mentioned is nothing but the consequence of the high conformational freedom that some dihedral angles within these molecules show (glycosidic and exocyclic torsions). Furthermore, when explicitly included, the treatment of the 1-4 interactions is often scaled by a factor to prevent the Van der Waals term to overcome the torsional interaction (e.g. OPLS-AA[214] and AMBER[196c]) within the force field potential energy. Indeed, it has been reported that 1–4 scaling hinders the correct parameterization of the exocyclic ω torsion in hexopyranoses[6a]. In fact, the weakening of 1–4 (O6-O5) with respect to 1–5 (O6-O4) interactions precludes from the accurate fitting of the rotational properties for this linkage[213b, 215]. In addition, 1–4 non-bonded interactions are not treated in the same manner in all force fields (**figure 13**), and this might be a problem when simulating complex systems (e.g., protein-glycan complex) that require the simultaneous use of different force fields. On the contrary, choosing not to use 1–4 scaling factors may cause a conflict with

the treatment of non-bonded interactions in protein force fields. Therefore, it might be necessary to employ a particular 1–4 scaling for each class of molecule, mainly in the simulation of glycoproteins. However, it should be considered that the potential impact of choosing the 1–4 scaling factors often becomes irrelevant when modelling protein-glycan complexes since the oligosaccharide mobility is generally reduced in the bound state, so that the default scaling appropriate for the protein may be employed.

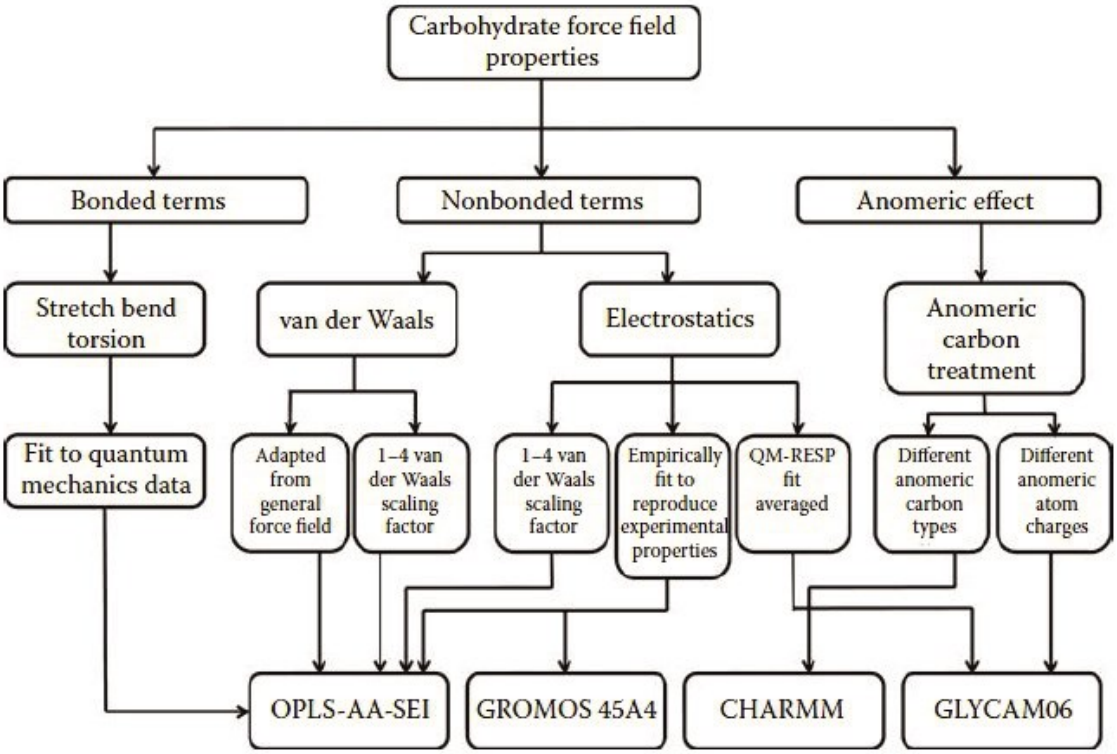

Figure 13. Comparison of the parameterization protocol employed by different commonly used carbohydrate force fields. *Source: Sarkar and Pérez 2012*[216].

In contrast to the analysis based solely on steric effects, it has been observed that the preference of the exocyclic hydroxymethyl group of glucopyranosides displays a strong preference for rotamers in which O6 and O5 are in a gauche orientation, i.e., *gg* and *gt,* and that force fields can reproduce this behaviour by introducing, on one side, explicit water molecules and, on the other side, a suitable treatment of the 1–4 non-bonded interactions[6a]. Different from the anomeric effect, the *gauche* effect is principally caused by solvation and electrostatic interactions rather than stereoelectronic effects[6a]. The primary role of explicit water molecules seems to be to disrupt the hydrogen bonding within the carbohydrate, thereby allowing the rotamer populations to be determined by internal electronic and steric repulsions between the oxygen atoms. The correct evaluation of the conformational preferences for the ω dihedral angle is of great importance because the preferred rotamer states can deeply impact the conformational properties of oligosaccharides containing 1–6 linkages, which are common in mammalian and bacterial cell-surface glycans[62, 217]. However, this is still a notoriously challenging

task in the research area of carbohydrate force fields[218] due to the need for a subtle balance between the inter- and intramolecular interactions and to the relatively long lifetimes of some states[6a, 207].

The performance of different carbohydrate force fields has been extensively reviewed[56a, 63, 213a, 219], but we will just comment the latest versions (parameterizations) of the most widely employed ones in computational chemistry, such as GLYCAM, AMBER, GROMOS, CHARMM, OPLS or TRIPOS.

GLYCAM06*[220]*

GLYCAM06 is a widely used force field, highly consistent and transferable, for modelling carbohydrates, glycoproteins, and glycolipids, as well as for protein–carbohydrate complexes[220-221]. It can be used for describing the physicochemical properties of complex glycans and glycoconjugates and, additionally, it is fully compatible with the AMBER force field. Also, GLYCAM06 may be used in simulation packages other than AMBER through the employment of appropriate file conversion tools. To facilitate the parameter transferability, all atomic sequences have an explicitly defined set of torsion terms (no generic terms), and PARM94 parameters[196c], the same used in AMBER, are employed for modelling the carbohydrate van der Waals components.

GLYCAM06 parameters were developed taking into account a test set of 100 molecules from the chemical families of hydrocarbons, alcohols, ethers, amides, esters, carboxylates, molecules of mixed functional groups, and simple ring systems related to cyclic carbohydrates and fit to quantum mechanical data. Based on a previous study indicating that the use of 1–4 scaling unbalanced the O6-O5 and O6-O4 interactions and, as a result of this, not being possible to correctly reproduce the rotamers populations of the ω dihedral angle[6a], no scaling factors for treating the non-bonded 1–4 interactions were introduced in GLYCAM06. In this carbohydrates parameterization, the stereoelectronic effects that influence bond and angle variations at the anomeric carbon atom are included in a unique anomeric atom type for both α- and β-glycosides. This feature permits to mimic the ring flipping observed in glycosidic monomers that occur, for example, during catalytic events[222] or conformational changes, leading to equilibrium between conformers with axial and equatorial substituents at the anomeric centre. This equilibrium is similar to an interconversion between anomers (but without bonds cleveage), which is not straightforward to simulate with force fields that employ unique torsion terms for each anomer.

A feature that is unique in GLYCAM06 is the use of solvated MD simulations to generate the ensemble of conformations employed in charge averaging. In this force field, the atomic partial charges were calculated residue by residue, so that a MD simulation of 50–100 ns was run for each, 100–200 snapshots were extracted, and the partial charges calculated by fitting to the averaging quantum mechanics molecular electrostatic potential (ESP), thus including charge dependence on the molecular conformation. In addition, restraints were employed in the ESP-fitting procedure (RESP) to ensure that the charges on all aliphatic hydrogen atoms were zero since C–H aliphatic hydrogen atoms are not

significant for reproducing dipole moments[223]. An optimal RESP charge restraint weight of 0.01 was applied, based on simulations of carbohydrate crystal lattices[224].

Comparison with experimental data confirmed that GLYCAM06 is able to reproduce rotational energies and carbohydrate features quite well if combined with an appropriate charge set, except for highly polar molecules (expectable because atomic polarizabilities were not included). For these cases, the dihedral terms were adjusted empirically to achieve an accurate behaviour in aqueous solution[220].

Regarding the modelling of highly polar molecules, specifically GAGs, with GLYCAM06, the recent release of 1) GLYCAM partial charges for the SO_3^- group to be used together with GLYCAM06 parameters (*http://glycam.ccrc.uga.edu/ccrc/pages/parameters.html*), and 2) the protocol of charge adjustment for the O- and N-sulphate moieties (*http://128.192.9.183/pmwiki/pmwiki.php/Main/DerivativesTutorial*), has finally turned this force field into a consistent modelling tool of high quality to simulate GAGs. In this regard, our experience simulating heparin oligosaccharides, as exposed in some parts of the present Thesis, have confirmed this issue.

GROMOS-45A4[225]**, *CHARMM*[226] and OPLS-AA**[227]

The GROMOS force field was earlier developed for MD simulations of proteins, nucleotides, or sugars in aqueous or apolar solutions or in crystalline form, but it was later modified to include the anomeric effects for mono- and oligopyranoses[225]. As in GLYCAM06, quantum mechanics methods were used for calculating bond and angle force constants, whereas dihedral parameters derivation and van der Waals terms were directly taken from previous GROMOS versions[228]. An ESP-fitting procedure, with restraints on aliphatic hydrogen atoms and averaging over atom types, was chosen for reproducing the electrostatic potential, using a trisaccharide as a model for charge development[225]. No distinction was done between α and β monomers in terms of charges and anomeric atom type, and electrostatic–van der Waals 1–4 scaling factors were not introduced so as to correctly reproduce the *gauche* effects on ω angles. A 20-ns-long MD simulation in explicit water[229] was used for validating the force field, showing the capability to correctly predict the stereoelectronic effects and the most stable ring conformations but sometimes failing to reproduce their correct energies. Recently, it was reported that GROMOS96 performed significantly better that GLYCAM06 in mimicking the transitions from the $^1C_4$ chair to the ^2S_O skew boat conformation for a L-IdoA2S monosaccharide [230]. Indeed, in this study GLYCAM06 predicted unrealistic $^1C_4$ to ^2S_O (and vice versa) transitions in the picosecond time scale. However, the authors combined GLYCAM06 parameters with Altona´s partial charges for sulphates[231], both sets developed under very distinct philosophies. This inconsistent use of parameters and partial charges is certainly possible to be the reason for the bad performance showed by GLYCAM06, which put on evidence the huge importance for the carbohydrate community to have a consistent and specific force field for glycans.

The CHARMM force field was recently extended to glucopyranose and its diastereoisomers[226b], assigning different atom types for the α and β anomers, and with the same hierarchical parameterization procedure and treatment of 1–4 nonbonded interactions used for other CHARMM biomolecular force fields[196d, 232] to ensure a full compatibility. In CHARMM, preliminary parameter sets are created using small-molecule models corresponding to fragments of pyranose rings and then successively applied to complete pyranose monosaccharide structures. Missing dihedral parameters are developed by fitting over 1800 quantum mechanical hexopyranose conformational energies. Both partial atomic charges and Lennard–Jones parameter values, taken from previous CHARMM versions, are adjusted to reproduce scaled quantum mechanical carbohydrate–water interaction energies and distances and further refined to reproduce experimental heats of vaporization and molecular volumes for liquids. This force field was validated as it reproduced QM-derived and experimental properties using MD simulations with TIP3P water molecules.

The OPLS force field was expanded to include carbohydrates[227]. In OPLS-AASEI (scaling electrostatic interactions) force field, 1–4, 1–5, and 1–6 scaling factors are introduced to improve the prediction of Φ and Ψ distributions, as well as anomeric effects and relative energies[227]. Unique charge sets and atom types for α and β anomers are used. All non-bonded parameters are imported directly from the parent force-field OPLSAA[233]. Charges are derived, as done for previous force-field versions[214, 233], from standard alcohols and acetals to simply reproduce consistent energetic properties, and then transferred them to carbohydrates.

Other force fields are employed to understand carbohydrate properties in silico. In particular, MM3, a force field initially meant for hydrocarbons, is now applicable to a wide range of compounds. The MM3 force field for amides, polypeptides, and proteins[234] is widely used for the construction of adiabatic maps of disaccharides.

TRIPOS molecular mechanics force field was designed to simulate both peptides and small organic molecules[235], but parameter extension for oligosaccharides includes sulphated glycosaminoglycan (GAG) fragments and glycopeptide-carbohydrate interactions[236]. The TRIPOS force field is implemented in the molecular package Sybyl and it is commonly used for geometry optimizations.

2.3.3 Molecular Dynamics simulations

Briefly, MD simulations consist of applying the laws of motion to the atoms that form the molecules to get an ensemble of molecular configurations. The concept behind MD simulation involves calculating the displacement coordinates in time (trajectory) of a molecular system at a given temperature. Finding positions and velocities of a set of particles as a function of time is done classically by integrating Newton's equation of motion in time. Molecular simulations are usually carried out as a microcanonical (NVE) or canonical (NVT) ensemble. As a consequence, all other thermodynamic quantities must be determined by ensemble averaging. In a classical system, Newton's equations of

motion conserve energy and thus provide a suitable scheme for calculating a microcanonical ensemble. However, of more practical application is the canonical ensemble since it can readily be performed by coupling the molecular system to a constant-temperature bath, which rescales the atomic velocities according to the desired temperature. In a similar manner, constant pressure simulations can be performed. The common procedure for running stable MD simulations of complex molecular systems (inclusion of explicit solvent, counterions, etc) is shown in **figure 14**.

Several algorithms have been developed for MD simulations that predict the time evolution of a system for a limited time. Thus, since physically observed properties are computed as the corresponding time averages on the individual microstates ensemble, for the results to be meaningful, the simulations must be sufficiently long so that the important motions are statistically well sampled. However, it must be considered that the longer the simulation the higher will also be the possibility of force field deviations to appear[6a]. Thus, the time scale of the phenomenon aimed to investigate should be considered in each particular case to decide the simulation time that interests us.

Experimentally accessible spectroscopic and thermodynamic quantities can be computed, compared, and related to microscopic interactions. It should be noted that MD is severely limited by the available computer power. With currently available clusters and computing algorithms, it is feasible to perform a simulation with several thousand explicit atoms for a total time that goes from hundreds of nanoseconds to several microseconds (the use of GPUs is providing cutting-edge velocities). However, it may be possible that the carbohydrate molecules undergo dynamical events on longer timescales and/or that the accessible computational resources are not updated, so that the time scale of interest cannot be investigated with standard MD techniques. For those cases, another way is to use high temperature dynamics to allow the molecule to assume high-energy conformations. However, this approach has to be used with caution since it can force the molecules to adopt unrealistic conformations (artefacts).

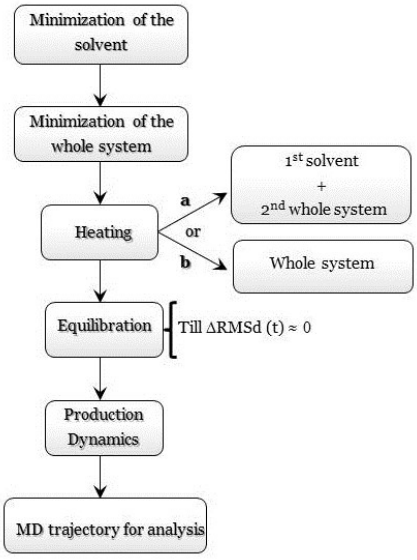

Figure 14. General scheme of the MD simulation protocol commonly followed.

Time-averaged restrained molecular dynamics simulations (tar-MD)

As the inter-conversion between the L-IdoA2S conformers in GAGs is rapid on the NMR time scale, the observed NMR resonances reflect an average of both conformations. Similarly, NOE_S from both conformations are observed simultaneously. Thus, the use of a MD methodology involving "instantaneous" experimental constraints[237] would generate structures that simultaneously satisfy both sets of experimental data. However, there may be no single conformation in agreement with the whole set of NOEs, and even if one is generated, it may be highly strained and physically unrealistic (a so-called virtual conformer[238]). Under these considerations, it is not correct to treat the NOE data as affording a fixed distance boundary. Instead, NOE distance information should be used to enforce an average distance limit through time. This can be achieved by imparting particles with a memory of their history with respect to internuclear distances. At the same time, to truly model the physical nature of the NOE, it is necessary to account for the nonlinear dependence of the measured NOE intensity on the internuclear distance. Based on these grounds, the methodology *time-averaged restrained* molecular dynamics (*tar*-MD)[239] includes the presence of a penalty term, $E_{penalty}$, in the total potential energy equation. For each experimental restraint (distance or coupling constant), six keywords (r_1 to r_4 distances or couplings constants, and r_{k2} and r_{k3} force constants) are defined. These parameters delimit the shape of the restraining potential as follows (**figure 15**)

$$R < r_1 \rightarrow E_{penalty} = r_{k2}(r_1 - r_2)^2 + 2r_{k2}(r_1 - r_2)(r_1 - R)$$

$$r_1 < R < r_2 \rightarrow E_{penalty} = r_{k2}(R - r_2)^2$$

$$r_2 < R < r_3 \rightarrow E_{penalty} = 0$$

$$r_3 < R < r_4 \rightarrow E_{penalty} = r_{k3}(R - r_3)^2$$

$$R > r_4 \rightarrow E_{penalty} = r_{k3}(r_3 - r_4)^2 + 2r_{k3}(r_3 - r_4)(r_4 - R)$$

Eq. 20

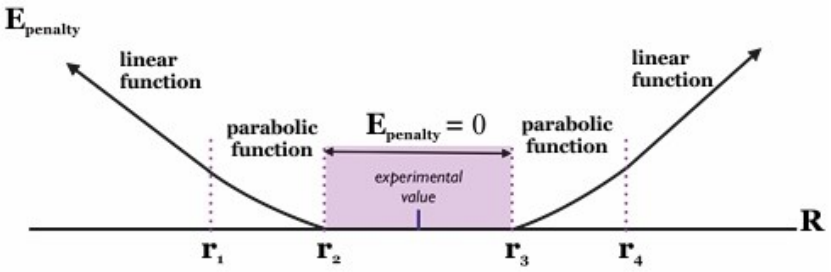

Figure 15. Representation of the penalty energy ($E_{penalty}$) applied in *tar-MD* simulations as a function of the value of the experimental parameter.

where R is the time-averaged value between atoms and r_1 to r_4 are the different limits defined below (r_1 and r_2) and beyond (r_3 and r_4) the experimental distance. The values of r_1 to r_4 have to be specified in Å or Hz, depending on the restraint type. r_{k2} and r_{k3} are defined in kcal·mol^{-1}·Å^{-2} for distance restraints and in kcal·mol^{-1}·rad^{-2} for dihedral restraints. The experimental key proton-pair distances, as well as coupling constants, are usually implemented as structural restraints with a 10% margin, using a flat well potential.

Since the NOE arises from dipolar interactions between nuclei, the intensity of a NOE signal grows as r^{-6} (as long as the simulation time is higher than the correlation time for overall molecular tumbling; otherwise, it has been shown that r^{-3} averaging is necessary)[240]. Thus, for distance constraints the expression for R included in MD force fields is

$$R = \langle r^{-6} \rangle^{-1/6} = \frac{\left(\sum_0^t r(t)^{-6} \cdot e^{(t-t')/\tau}\right)^{-1/6}}{N} \qquad \textit{Eq. 21}$$

with τ being the characteristic time for the exponential decay or exponential decay constant, and t' the total simulation time. The exponential decay constant "helps" the calculation to converge and it is commonly set to a value 10 times smaller than t'. This methodology has been extensively and successfully applied to the study of different molecular systems[241], iduronate containing carbohydrates included[78].

2.3.4 Docking

Molecular docking is a computational procedure that aims at predicting the preferred orientation and conformation of a ligand bound to its target protein. In order to perform computational protein-ligand docking calculations, the 3D structure of the receptor must be known. Each docking program operates slightly differently, but they share common features that enable them to (1) search for locations on the protein surface that lead to favourable interactions with the ligand, (2) sample the conformational space of the ligand, and (3) compute the interaction energy between the protein and ligand (score of "binding affinity" or scoring function). For instance, *glide* (*grid-based ligand docking with energetics*) uses a series of hierarchical filters to search for possible locations of the ligand in the active-site region of the receptor (**figure 16**).

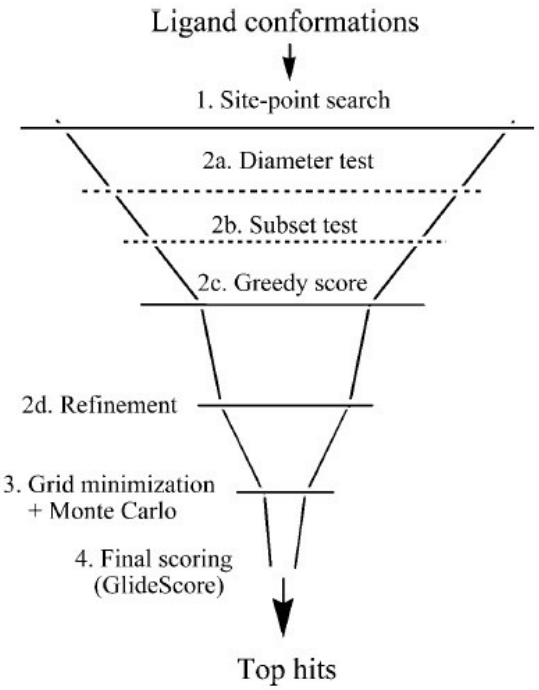

Figure 16. *Glide* docking "funnel" showing the protocol followed to generate docked poses.

Glide docking algorithm approximates a complete systematic search over ligand positions, orientations, and conformations in the receptor site, with increasingly demanding tests applied as the search space is reduced. *Glide* generates and docks many core conformations, but treats the rotamer groups sequentially, rather than combinatorially, which speeds up the calculation.

The hierarchical protocol followed by *glide*, shown in **figure 16**, can be briefly described as follows:

1. Site-point search
   - Generate a 2-Å grid of site points in the active site.
   - Pre-compute histograms of distances between site point and receptor surface in grid setup.
   - Compare site point – receptor surface histograms with the ligand centre–ligand surface histogram.
   - Reject mismatched site points.

2. Dimensional tests and rough scoring
   - Diameter test: check steric clashes of atoms near ligand diameter for ~300 pre-specified orientations of the ligand diameter (**figure 17**).
   - Subset test: rotate about ligand diameter in 15º increments, and score atoms capable of establishing hydrogen bonds or ligand-metal interactions.
   - Greedy scoring: score all atom positions ±1 Å in x,y,z directions and use best score.
   - Refinement: move the whole ligand ±1 Å in x,y,z directions, re-score and reduce ~5000 poses to ~400 for energy minimization.

3. Energy minimization

   - Use pre-computed OPLS-AA electrostatic and van der Waals grids.

   - Anneal from soft-to-hard potential: smoothing reduces large initial energy/gradient terms from close contacts, permits freer movement.

   - Also optimize torsional angles when doing flexible docking.

   - Use Monte Carlo moves to explore nearby torsional minima for a small number of low-energy poses.

4. Final scoring.

   - Choose best pose(s) based on *Emodel*, which is a combination of the Coulomb-vdW energy, the *GlideScore* (enhanced version of *ChemScore*) and internal strain energy.

   - Final scoring based on *GlideScore*, consisting of *ChemScore* terms, the Coulomb-vdW energy and terms that penalize non-physical interactions

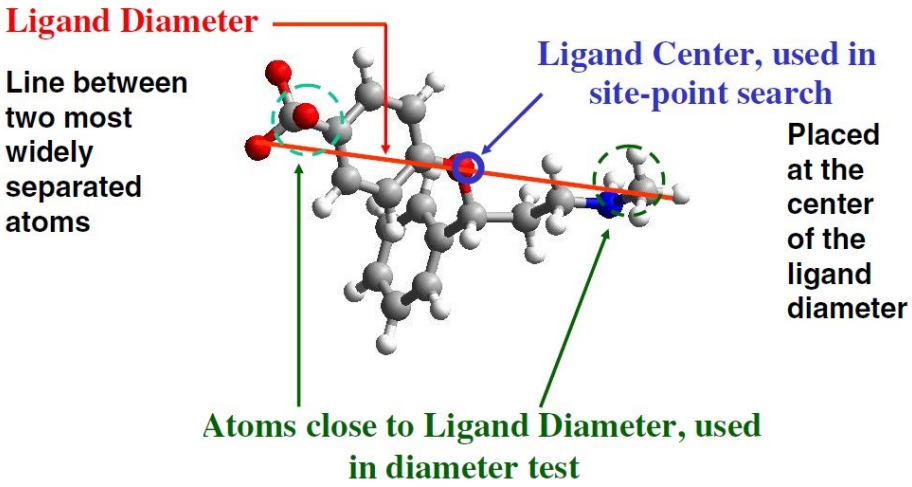

Figure 17. Definition of ligand diameter and ligand centre parameters according to *glide* searching algorithm. *Source: Thomas A. Halgren, Schrödinger.*

Protein interaction with the ligand relies on both the protein backbone fold and the orientation of the side chains in the binding site region. One of the most significant limitations in docking is that it is generally performed while keeping the protein surface rigid, which prevents the consideration of the effects of induced fit within the binding site. These difficulties are mostly due to the high number of degrees of freedom characterizing a protein–ligand system, which increases the computational cost of docking calculations. Thus, several approximations about the flexibility states may be introduced in molecular docking. The simplest approximation (rigid docking) considers only the three translational and three rotational degrees of freedom of the protein and those of the ligand, treating them as two distinct rigid bodies. However, the most widely used algorithms at present enable the ligand to fully explore its conformational degree of freedom in a rigid-body receptor[242].

Induced Fit Docking (IFD)

As we have introduced above, in standard docking studies ligands are docked into the binding site of a receptor where the latter is held rigid and the ligand is free to move. While this approximation present the advantage of reducing the computational cost, it however may give rise to misleading results, since in reality many proteins undergo side-chain or back- bone movements, or both, upon ligand binding. These changes allow the receptor to alter its binding site so that it better adapts to the shape and binding mode of the ligand. This is often referred to as Induced Fit and is one of the most challenging features to model in structure-based drug design. Thus, a good Induced Fit Docking (IFD) protocol should both generate an accurate complex structure for a ligand known to be active but that cannot be docked in an existing (rigid) structure of the receptor and also rescue false negatives (poorly scored true binders) in virtual screening experiments, where instead of screening against a single conformation of the receptor, additional conformations obtained with the IFD protocol are used. Since glycans interactions present dissociation constants (K_D) in the weak binding regime, i.e., poorly scored true binders from the point of view of docking scoring functions, the IFD method may result specially convenient to obtain an accurate description of their interactions.

Furthermore, poor binders exhibit a significant dependence on the initial input conformation during docking due to the use of complex grid-based potentials and the practical limitations of thoroughly sampling the docked poses. Also, docking algorithms usually generates new ligand conformations through torsional variations only, so any differences in bond lengths and angles in the input ligand structures will persist through the docked poses, resulting in scoring and pose differences. For these reasons, docking several conformations of each ligand with variations in bond lengths and bond angles is a reasonable strategy to reduce input dependence.

Grid generation

Docking algorithms represent the shape and properties of the receptor on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand poses. *Glide* allows to define the receptor structure by excluding any co-crystallized ligand that may be present, determine the position and the extent of the region for which receptor grids will be calculated (30Å-sided cube maximum) and set up constraints. In any docking job using these receptor grids, ligands are confined to the enclosing box (**figure 18**).

Also, the ligand centre can be set during grid generation in *glide*. The ligand centre of a ligand is defined, in *glide*, as the midpoint of the longest line segment that can be constructed between any two atoms in the ligand. Furthermore, the ligand diameter midpoint box is the region in which the diameter midpoint of each docked ligand must remain. Each dimension of this box can be modified from its default value of 10 Å to the 6-14 Å range. When doing so, the enclosing box also changes its size to make the distance between faces of the enclosing box and the ligand diameter midpoint box alike.

A larger ligand diameter midpoint box can be useful to allow ligands to find unusual or asymmetric binding modes in the active site. Conversely, if the default ligand diameter midpoint box allows ligands to stray into regions you know to be unfruitful, you can confine their midpoints to a smaller box, eliminating some of the less useful poses and saving calculation time.

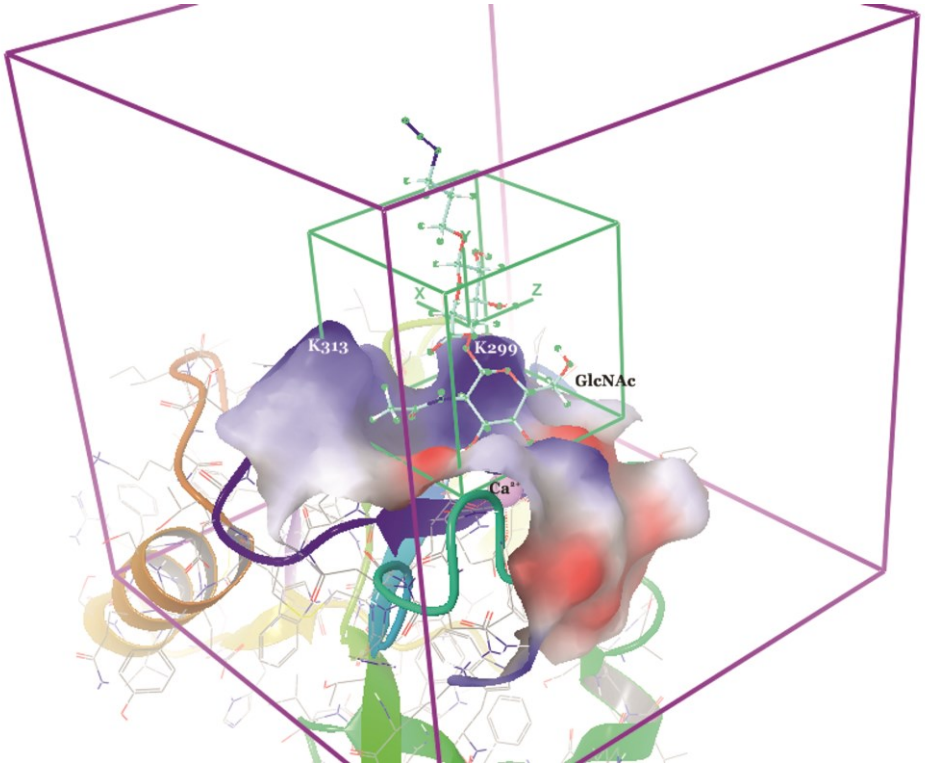

Figure 18. 3D image showing the grid box (purple-lined cube) and the *ligand diameter midpoint box* (green-lined cube) in *glide*.

Docking Algorithms

The docking algorithms can be grouped into deterministic and stochastic approaches. While deterministic algorithms are reproducible, stochastic algorithms include random factors that do not allow the full reproducibility. The most widely used algorithms in docking simulations are described below.

Incremental Construction Algorithms

These algorithms consist of the division of a ligand into rigid fragments. One of the fragments is selected and placed in the protein binding site. The reconstruction of the ligand is then carried out in situ, adding the remaining ligand fragments. For example, DOCK[243] uses incremental construction algorithm to treat ligand flexibility. It generates points (sphere centres) that fill the binding site and try to capture the binding site shape properties for identifying favourable regions in which the ligand atoms may be located. The ligand is divided along each flexible bond to generate rigid segments. An

anchor fragment is then selected from all the rigid pieces and oriented in the active site by matching ligand atoms with sphere centres. After, fragments are added and all possible placements are scored on the basis of their interactions with the protein using the energetic scoring function. Then, the best anchor fragments are used for completing the construction of the ligand in the protein-binding site. Finally, the best scored poses of the complete ligand are selected.

Genetic Algorithms

Genetic algorithms are stochastic searching approaches that use techniques inspired by evolutionary biology to find reliable results. It mimics the process of evolution by manipulating a collection of data structures called chromosomes. AutoDock[244] uses this algorithm for obtaining reliable docking results. First, the protein is placed inside a cube with a predefined size, characterized by a defined number of points (grid points). In the second step, probes corresponding to the different atom types of the ligand are then moved through the cube and, in particular, at each point, protein–probe interaction energies are calculated and stored in affinity maps. Thirdly, a conformational search of the ligand is performed by applying the Lamarckian genetic algorithm[245]. At this stage, a minimization or local search is performed, and the new conformation is then considered as input for a new iteration of the genetic algorithm cycle.

Hierarchical Algorithms

It uses an exhaustive systematic search for discovering the most favoured ligand conformations in the protein active site, with a screening based on progressively restricted energetic cutoffs. A grid and a molecular surface containing information of the protein receptor properties are calculated before the algorithm search. Then a set of initial ligand conformations is produced and screens are performed over the whole phase space available to the ligand to locate promising ligand poses in the respective receptor fields. Afterwards, ligands are minimized in the field of the receptor using a standard molecular mechanics energy function[227, 233]. Finally, the lowest-energy poses are subjected to a Monte Carlo procedure that examines torsional minima and a composite scoring function is then used to select the correct docked poses. The algorithm used in *glide*[246] can be defined as a hierarchical algorithm.

Scoring Functions

Energy scoring functions are necessary to evaluate the free energy of binding ΔG (affinity) of ligand-receptor interactions. The Gibbs free energy equation describes the ligand-receptor free energy of binding as

$$\Delta G = \Delta H - T\Delta S \qquad Eq.\ 22$$

where ΔG represents the energetic changes between the bound and unbound states of both ligand and receptor, ΔH is the enthalpy, T the temperature expressed in Kelvin, and ΔS the entropy of the system.

Furthermore, ΔG is related to the binding association and dissociation constants (K_a and K_d, respectively) as

$$\Delta G^0 = -RTlnK_a = RTlnK_d \qquad Eq.\ 23$$

(where R is the constant for ideal gases), which allows to obtain an estimate of binding affinity.

Some sophisticated techniques for predicting binding free energies are currently too slow to be used in molecular docking of large sets of compounds[247]. Thus, fast scoring functions have been developed. Empirical scoring functions use a set of parameterized terms describing properties known to be important in protein–ligand binding to construct an equation for predicting binding affinities. Multilinear regression is used to optimize these terms using a set of known protein–ligand complexes. These terms usually describe polar–apolar interactions, loss of ligand flexibility (entropy), and desolvation effects. For instance, *GlideScore* 2.5 scoring function[246a] is a regression based empirical scoring function of the form

$$\begin{aligned}\Delta G = {} & C_{lipo-lipo}\sum f(r_{lr}) + C_{hbond-neut-neut}\sum g(\Delta r)h(\Delta\alpha) \\ & + C_{hbond-neut-charged}\sum g(\Delta r)h(\Delta\alpha) + C_{hbond-charged-charged}\sum g(\Delta r)h(\Delta\alpha) \\ & + C_{max-metal-ion}\sum f(r_{lm}) + C_{rotb}H_{rotb} + C_{polar-phob}V_{polar-phob} + C_{coul}E_{coul} \\ & + C_{VdW}E_{VdW} + solvation\ terms\end{aligned} \qquad Eq.\ 24$$

The first term describes the lipophilic and aromatic interactions, whereas the polar terms are included in the second, third and fourth terms (hydrogen bonds separated into differently weighted components that depend on the electrostatic properties of donor and acceptor atoms), and the $C_{max\text{-}metal\text{-}ion}$ term, which includes the anionic(ligand)-metal(receptor) interactions. The seventh term rewards instances in which a polar but non-hydrogen bonding atom is found in a hydrophobic region. Also, Coulomb and van der Waals interaction energies between the ligand and the receptor are evaluated as well as the solvation effect. *GlideScore*[246] has been optimized for docking accuracy, database enrichment and binding affinity prediction, and can be used as an empirical scoring function that approximates the ligand binding free energy. *GlideScore* should be used to rank poses of different ligands, for example in virtual screening.

Other scoring functions different than *GlideScore* are also used in *glide*. Therefore, *Emodel* scoring function has a more significant weighting of the force field components (electrostatic and van der Waals energies), which makes it well-suited for comparing conformers but much less so for comparing chemically-distinct species. *Glide* uses *Emodel* to pick the "best" pose of a ligand (pose selection) and then ranks these best poses against one another with *GlideScore*. This means that if you save multiple poses per ligand, the apparent ranking of poses for a given ligand (by *GlideScore*) will not reflect the

actual ranking that *glide* used for pose selection. So, the value of *Emodel* scoring has to be considered to determine the highest ranked pose for a ligand.

On the other hand, force-field-based scoring functions (e.g. those used by AutoDock or DOCK) are based on the non-bonded terms of the classical molecular mechanics force fields. In AutoDock[244, 248], the implemented scoring function presents five terms with coefficients empirically determined using linear regression analysis from a set of protein–ligand complexes with known binding constants. A 12-6 Lennard-Jones potential and a Coulomb term taken from the AMBER force field[196c] describe the van der Waals and electrostatic interactions, respectively. In addition, hydrogen bonding is described with a 12-10 Lennard-Jones term with Goodford directionality[249]. Also, a desolvation energy potential and an empirical measure of the unfavourable entropy of ligand binding due to the restriction of conformational degrees of freedom are included.

Other docking approaches use knowledge-based scoring functions based on statistical observations of intermolecular close contacts in protein–ligand X-ray databases, which are used to derive potentials of mean force. This methodology assumes that the frequency of close intermolecular interactions between certain ligand and protein atoms contribute favourably to the binding affinity. In this approach, no fitting to experimental affinities is required and solvation and entropic terms are treated implicitly[250].

Chapter 3

Structural studies of heparin-like oligosaccharides by NMR and MD techniques

3.1 Library of sulphated trisaccharides

3.1.1 Background

As it was mentioned in *Chapter 1*, heparin is a glycosaminoglycan (GAG) composed by a disaccharide repeating unit built of a 2-O-sulphated iduronic acid (L-IdoA2S) and a glucosamine (GlcN) residue 1,4-linked and strongly substituted by sulphate groups. GAGs interactions with proteins regulate the activity of hundreds of them[71, 251], and thus allowing proteins to exert their function. Many proteins require the interaction with cell surface GAGs to exert their biological activity[47b, 252]. The effect of GAG binding on protein function ranges from essential roles in development, cell growth, cell adhesion, inflammation and tumorigenesis to interactions with pathogens[39]. Some authors have suggested a continuous range of affinities in function of the substitution pattern, the population of a particular iduronate conformation or the presence of a kink in the carbohydrate chain[3, 253].

From the study of several well-known proteins (e.g. FGF-1), it has been described that heparin interactions (also heparan sulphate) with proteins, do not apparently induce large conformational changes in the structure of the carbohydrate[254]. Therefore, considering heparin 3D structure, a rigid helix with a complete turn every four residues[65] (or a 180º turn every two monosaccharides[65]), it can be expected the interactions with the same side of a protein surface to be discontinued, grouping each three contiguous residues (see *figure 14* in *Chapter 1*)[66b, 251b]. Thus, a heparin-like trisaccharide of D-GlcN-L-IdoA-D-GlcN sequence (e.g. **Tri1-Tri8**) clusters three sulphate groups towards the same side of the molecule, giving rise to a minimum recognition site[255], so that an adjacent second site remains either directed towards the opposite direction, in an antiparallel fashion, or two-residue-shifted in a parallel orientation (see *figure 14* in *Chapter 1*). Furthermore, the number and distribution of sulphate groups might play some role in the specificity of their interactions[67]. Also, a structural key aspect of heparin (or HS) is that whereas it is very rigid from the backbone perspective (global conformation), it behaves very flexibly at the local level (iduronate conformational plasticity)[82].

Based on these premises, our group synthesized a small library of eight trisaccharides, **Tri1-Tri8** (**figure 1**). Starting from the trisaccharide of minimum sulfation, of D-GlcNS-L-IdoA2S-D-GlcNAc sequence (**Tri8**), the rest of potential sulphation sites were permutated except the position 3 of the GlcN residues, that has only been reported to participate in the interaction with antithrombin[256]. Thus, in the case of the N-acetyl group of the reducing end GlcN residue, it was substituted by an N-sulphate group (**Tri1**, **Tri3**, **Tri5** and **Tri7**). Also, the 6-OH groups of both GlcN residues were differently sulphated (**Tri3-Tri8**). The carboxylate group of the L-IdoA2S ring was not permutated since it is a key element in the control of the conformational equilibrium of the iduronate ring and, additionally, this replacement has not been reported in the regular region of heparin. Lastly, it is noticeable the choice of the isopropyl group as reducing terminal. While not modifying significantly the biological and structural properties of the synthetic oligosaccharide, the synthesis of this type of compounds in the isopropyl glycoside form presents important advantages, e.g., the high

stereoselectivity for the α anomer and the good yield obtained for the reaction of the isopropyl group introduction in comparison with other alkyl groups such as the methyl group[257]. Furthermore, it is reasonable to think that the conformation of an isopropyl group at the reducing terminal better mimic that of the natural polysaccharide since we are dealing with a secondary alcohol in both cases. This "polysaccharide-oligosaccharide structural similitude" suggests that the **Tri1-Tri8** set of chemically synthesized derivatives are suitable candidates to be employed as models of the polymer of natural heparin from both the structural and spectroscopic perspectives (indeed, we have been able to measure with high precision the coupling constants; see Results and discussion section). Alike, it is important to highlight the interest of having available compounds of perfectly defined chemical structure which mitigate the disadvantages that, from the spectroscopic point of view, the polymeric nature of heparin presents, and, at the same time, conserve the properties of natural products. Natural oligosaccharides (enzymatically biosynthesized) contain an insaturation at the non-reducing end sugar residue and present anomeric mixture at the reducing terminal. These undesirable border effects become especially significant in small oligosaccharides for which any or both terminals are involved in binding, so that they may distort the results observed in comparison to those obtained with chemically pure compounds. Therefore, the chemical purity of **Tri1-Tri8** synthetic molecules and their designed substitution patterns make this set of eight trisaccharides an excellent framework to study the impact of sulphation differences of adjacent residues on the conformational behaviour of the L-IdoA2S ring (with the exception of the N- position of the non-reducing GlcN residue, which is sulphated in all the trisaccharides). On the other hand,

In this chapter, the investigation on the conformation and dynamics of **Tri1-Tri8** compounds by nuclear magnetic resonance and molecular dynamics simulations is described, and some interesting observed effects of temperature and sulfation pattern on the equilibrium of the central iduronate ring are discussed.

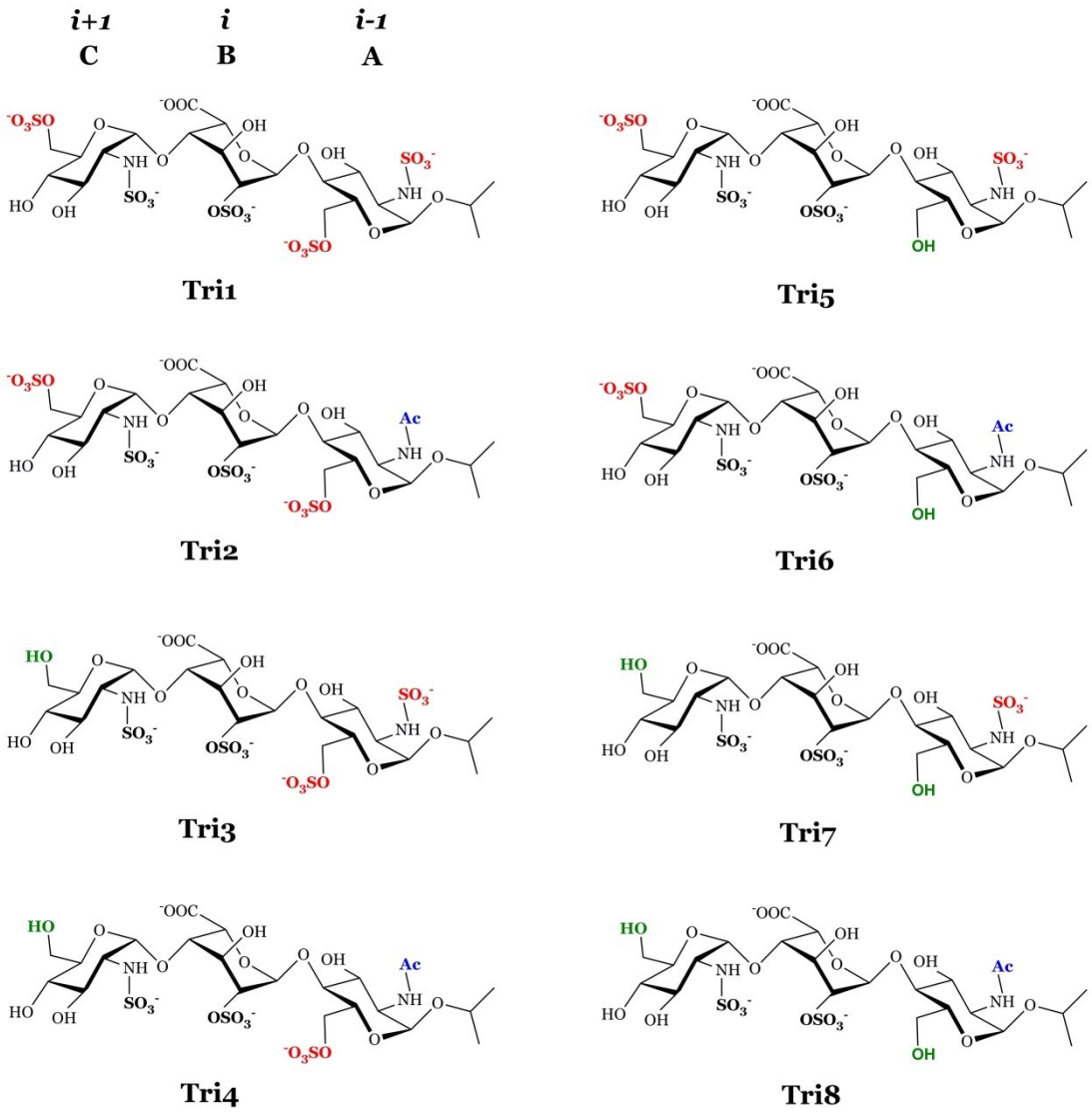

Figure 1. Scheme representation of the library of heparin-like trisaccharides displaying the 3D relative disposition of their substituents up and down the middle plane of the molecules. The rings have been labelled A, B, C (as it is shown for **Tri1**), and this nomenclature is used throughout the text (i, i-1 and i+1 nomenclature is employed when referring to a polysaccharide chain). The non-variable and variable sulphate groups are shown in bold black and bold red, respectively. The acetyl (Ac) and hydroxyl (OH) groups appear in bold blue and bold green, respectively.

3.1.2 Results and discussion

Nuclear Magnetic Resonance

All the compounds **Tri1-Tri8** were ^{1}H and ^{13}C assigned at several temperatures (see *Appendix*). First, the spin systems were assigned from COSY and TOCSY experiments, indicating that the three hexopyranose rings were connected by interglycosidic NOE or ROE contacts between the non-reducing GlcN anomeric proton and protons H3 and H4 of the L-IdoA2S ring (GlcN(A)-IdoA2S(B) linkage) and

between the L-IdoA2S H1 proton and protons H4 and H6(*proR+proS)* of the reducing end GlcN residue (IdoA2S(B)-GlcN(A) linkage). This indicates a preferentially *syn* rearrangement of both glycosidic linkages[86].

On the other hand, the known impact of sulphation on the ^{1}H and ^{13}C chemical shifts was observed, i.e., a downfield shift of both the adjacent protons and the ipso carbon atom with respect to the sulphation site[258] (see *Appendix*). The larger shifts were observed in positions 2 and 6 of the reducing end GlcN residue and 6 of the non-reducing terminal (sulphation sites).

Interestingly, the methylene protons at position 6 of ring A are isochronous, while those of ring C have different chemical shift (with the differences being larger when this is 6-O-sulphated). When the set of **Tri1-Tri8** compounds were compared, all the L-IdoA2S resonances appeared well clustered except for those of H5b protons, that were dispersed without any apparent reason.

NOESY, ROESY, 1D-NOESY, 1D-ROESY and 1D-T-ROESY experiments were register to calculate the experimental distances from the cross-relaxation rates σ_{NOE} (or σ_{ROE}), these derived from the NOE/ROE growth curves[259]. The results at room temperature were consistent with a molecule in a motion regime close to the NOE zero-crossing point, thus implying low intensities and difficulties to integrate. At the same temperature, ROE peaks, although measurable, were too contaminated by strong coupling and other artefacts. For example, TOCSY transfer of magnetization was frequently detected, in particular for the iduronate ring as a result of its smaller chemical shift dispersion. Attempts to solve this problem using 1D-T-ROESY did not give better results. So, we decided to vary the temperature in order to modify the correlation time, obtaining the best results at low temperatures (278 K) using 1D-NOESY. By decreasing temperature to 278K we were able to both increase the global correlation time (τ_o) to obtain negative NOE peaks typical of a macromolecule tumbling in solution (at room temperature the trisaccharides behaved close to zero-crossing) and augment the population of the lowest energy conformations. At higher temperatures (308 and 313 K), the zero-crossing was still a problem in some cases. Thus, due to the comparative nature of this work we decided to use the 1D-NOESY experiments at 278 K since it allowed us to analyse all the compounds in the same conditions. The NOE growth curves for those protons of particular interest were built via mono-dimensional selective analogues of NOESY (dpfgse-1D-NOESY) experiment since they have a larger linearity and better quality than the 2D analogues[260] (**figure 2**). Using the isolated spin-pair approximation (ISPA)[261], the interprotonic distances were calculated (**table 1**), as $\langle r^{-6} \rangle^{-1/6}$, from their respective σ_{NOE} and a reference of known distance (H1c-H2c in all cases), with these σ_{NOE} being obtained from the slope of the linear fit of the NOE normalized intensities at different mixing times (see *Appendix*).

At low temperature (278 K), the NOEs pattern (1D-NOESY) indicated a *syn* rearrangement of both glycosidic linkages (H1c-H3b&H1c-H4b and H1b-H4a for the GlcN-IdoA and IdoA-GlcN linkages, respectively). Nevertheless, the spectral overlapping did not allow us to unequivocally assigned other NOEs observed upon selective irradiation of H5b and H1b protons. This was solved by carrying out

2D-NOESY experiments (278 K), from which cross-peaks corresponding to an *anti*-Ψ geometry around the IdoA-GlcN linkage could be identified (H1b-H3a, H1b-H5a and, in some cases, H5a-H6a+a'), together with larger H1b-H4a and H1b-H6a+a' NOEs (with exceptions), the latter showing a major *syn* conformation (**figure 3**). The *anti*-Ψ exclusive H5b-H6a+a' NOE was only observed, with a very weak intensity, for **Tri1**, **Tri3**, **Tri5** and **Tri6**. However, it is noticeable that the growth of this NOE is surely affected by the loss of magnetization due to a very short longitudinal relaxation time (T1), as a consequence of the efficient (in terms of T1) relative reorientation of the H5b and H6a+a' coupled protons (free rotation of methylene protons and IdoA2S conformational plasticity acting in the nanosecond time scale as relaxation mechanisms). Thus, although the *anti*-ψ conformation is present, the H5b-H6a+a' NOE intensity may not be observed (or very weakly) due to fast relaxation. This is why, to estimate the contribution of the *anti*-Ψ conformers (IdoA-GlcN linkages) it is preferable to consider the H1b-H3a and H1b-H5a NOEs (specially the former). A table containing the normalized H1b-H3a cross-relaxation rates for **Tri1-Tri8** is included in the *Appendix*, showing that the GlcNS-IdoA2S-GlcNAc sequence (**Tri8**) contributes to the largest extent to the presence of *anti*-ψ conformations around the IdoA2S-GlcN linkage.

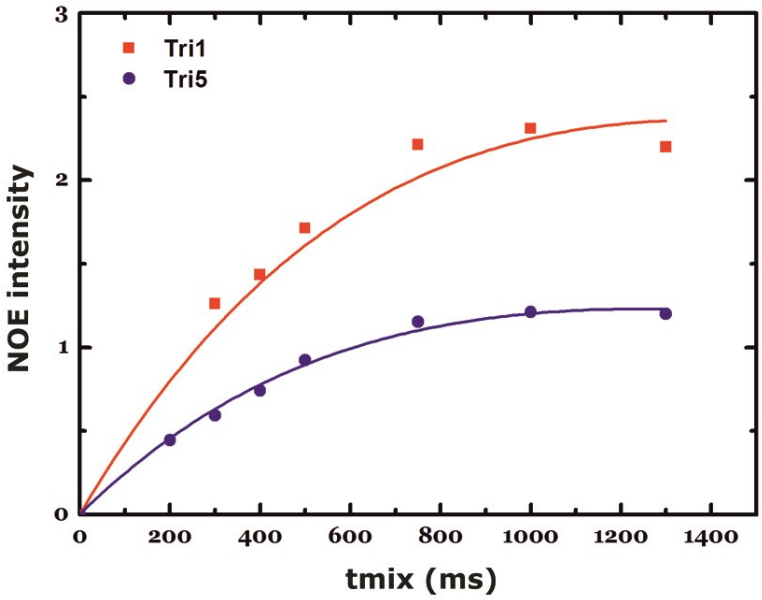

Figure 2. 1D-NOESY growth curves (at 278 K) corresponding to the H2b-H5b distance of trisaccharides **Tri1** (6-OSO_3^- in the reducing terminal) and **Tri5** (6-OH in the reducing terminal). Note the significantly higher NOE intensity (shorter H2b-H5b distance) for the trisaccharide containing the 6-sulphate group (**Tri1**).

Table 1. Comparison between the experimental (exptl) and theoretical (*tar*-MD and free-MD) most relevant distances of the library of trisaccharides. The experimental values were derived from 1D-NOESY experiments at 278 K. The *tar*-MD derived results were calculated as $\langle r^{-6} \rangle^{-1/6}$ over the 8000 frames of *tar*-MD simulation at 278 K. The free-MD calculated were calculated as $\langle r^{-6} \rangle^{-1/6}$ and weighted on the populations of iduronate conformers at 278 K. The *tar*-MD and free-MD H1b-H6a distances represent the r^{-6} average over the H1b-H6a*proR* and H1b-H6a*proS* values.

Comp	Method	Distance (Å)						
		GlcN-IdoA2S linkage		L-IdoA2S		IdoA2S-GlcN linkage		
		H1c-H3b	*H1c-H4b*	*H1b-H3b*	*H2b-H5b*	*H1b-H3a**	*H1b-H4a*	*H1b-H6a**
Tri1	free-MD	2.3	2.8	3.1	2.6	2.7	2.4	3.5
	tar-MD	2.4	2.6	3.3	2.9	4.2	2.3	2.9
	exptl	2.6	2.7	3.0	3.0	3.3	2.5	2.7
Tri2	free-MD	2.3	2.9	3.1	2.6	2.8	2.4	3.4
	tar-MD	2.6	2.5	3.2	2.8	4.2	2.3	2.9
	exptl	2.6	2.6	3.0	2.9	3.0	2.4	2.6
Tri3	free-MD	2.3	2.8	3.1	2.6	3.9	2.3	3.3
	tar-MD	2.5	2.5	3.3	3.0	4.3	2.4	2.8
	exptl	2.6	2.7	3.0	2.9	3.4	2.6	2.7
Tri4	free-MD	2.3	2.8	3.1	2.6	3.8	2.3	3.2
	tar-MD	2.6	2.5	3.3	2.9	4.1	2.3	3.1
	exptl	2.7	2.7	3.0	2.9	3.2	2.5	2.7
Tri5	free-MD	2.3	2.8	3.4	2.9	3.8	2.3	3.5
	tar-MD	2.3	2.6	3.6	3.3	4.1	2.3	2.9
	exptl	2.6	2.6	-	3.2	2.8	2.6	2.5
Tri6	free-MD	2.3	2.6	3.3	2.8	2.9	2.4	3.5
	tar-MD	2.4	2.5	3.6	3.1	4.2	2.3	2.9
	exptl	2.6	2.6	3.2	3.2	3.7	2.6	2.6
Tri7	free-MD	2.3	2.7	3.4	2.8	3.7	2.3	3.3
	tar-MD	2.4	2.5	3.3	3.0	4.3	2.3	2.9
	exptl	2.6	2.6	3.2	3.2	3.1	2.5	2.6
Tri8	free-MD	2.3	2.7	3.2	2.7	3.9	2.3	3.2
	tar-MD	2.5	2.5	3.4	2.9	4.2	2.3	3.1
	exptl	2.6	2.6	3.1	3.0	2.6	2.6	-

*Values derived from 2D-NOESY experiments at 278 K and 600 ms mixing time. The H1c-H2c NOE was used as reference (as in the case of 1D-NOESY derived distances).

Regarding the conformation of the iduronate ring, the analysis of the accurately measured coupling constants (**table 2**), the absence of the H5c-H5b NOE signal (exclusive of the iduronate $^{4}C_{1}$ chair conformation), together with the observation of the $^{2}S_{O}$-exclusive H2b-H5b and H1b-H3b contacts (NOE, ROE, T-ROE; see NOESY spectrum in **figure 3**), indicated the coexistence in solution of the $^{1}C_{4}$ chair and the $^{2}S_{O}$ skew-boat conformers. This is a characteristic feature of the internal iduronate ring in heparins and heparin-derived oligosaccharides[69a].

From the NOE-derived distances obtained (**table 1**), we focused our analysis on those defining the local (distances H2-H5 and H1-H3 of the L-IdoA2S ring) and global (H1c-H3b and H1c-H4b distances of the GlcN-IdoA linkage, and H1b-H3a, H1b-H5a, H1b-H4a and H1b-H6a+a' of the IdoA-GlcN glycosidic linkage) conformations, comparing them along the **Tri1-Tri8** library. First, dealing with the local geometry (L-IdoA2S residue), the calculated values for the H4b-H5b distance (between 2.4 and 2.6 Å) are in agreement with the canonical $^{1}C_{4}$ (2.5 Å) and $^{2}S_{O}$ (2.4 Å) conformations (PDB code 1hpn). Regarding the $^{2}S_{O}$-exclusive H1b-H3b and H2b-H5b distances (**table 1**), the results showed shorter values in the **Tri1-Tri4** ensemble (≅3.0 Å) compared to the **Tri5-Tri8** one (≅3.2 Å). Since this distances, according to the canonical conformations, are much shorter in the $^{2}S_{O}$ conformer (2.9 and 2.4 Å for the H1b-H3b and H2b-H5b distances, respectively; 4.3 and 4.0 Å in the $^{1}C_{4}$ pucker), this observation was in agreement with a higher population of $^{2}S_{O}$ pucker in the **Tri1-Tri4** ensemble, thus when the 6-OSO_{3}^{-} group is present in the reducing end GlcN residue (**figure 4**).

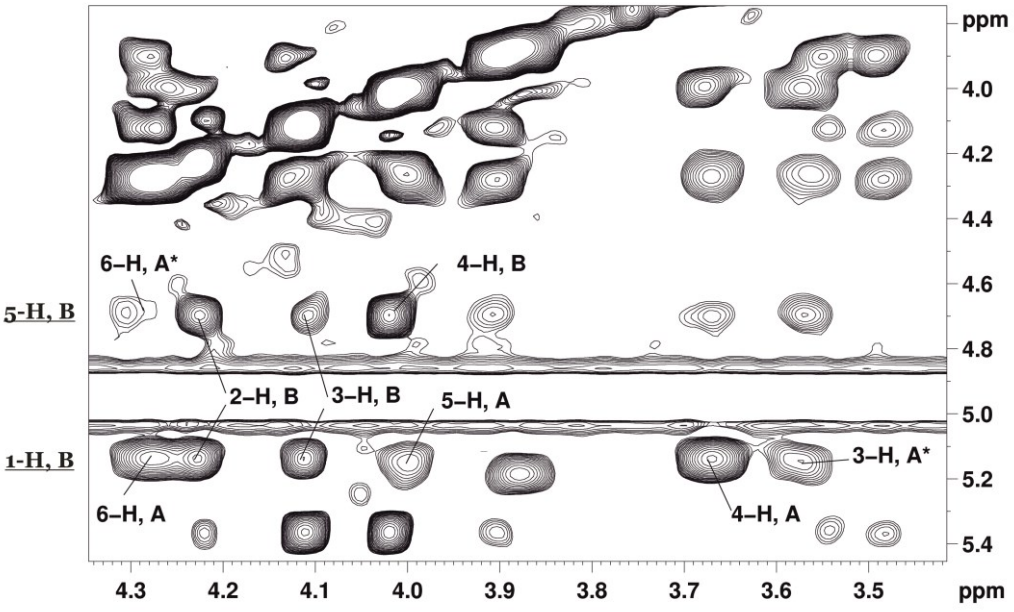

Figure 3. Expansion of a NOESY experiment for **Tri1** showing the signals corresponding to the *anti* and *syn* rearrangements of the Ido(B)-GlcN(A) glycosidic linkage. The most relevant exclusive NOE peaks of the *anti*-Ψ conformation are marked with a star symbol.

Table 2. Comparison of the experimental (exptl) and theoretical (free-MD and *tar*-MD) proton-proton vicinal coupling constants ($^3J_{HH}$) for the L-IdoA2S ring in the **Tri1-Tri8** compounds. The experimental values were measured at 278 K. The *tar*-MD derived results represent the average over the 8000 frames of *tar*-MD simulation at 278 K. The free-MD calculated values are weighted on the populations of iduronate puckers at 278 K. [a] Over 1 Hz of difference with respect to the corresponding experimental value.

Comp	Method	*H1-C1-C2-H2*	*H2-C2-C3-H3*	*H3-C3-C4-H4*	*H4-C4-C5-H5*
	free-MD	3.9	5.6	3.4	2.3
Tri1	tar-MD	3.1	5.1[a]	3.7	2.4
	exptl	3.0	6.2	4.0	3.1
	free-MD	4.1	5.9	3.5	2.3
Tri2	tar-MD	3.5	5.8	3.8	2.5
	exptl	3.2	6.5	3.9	2.9
	free-MD	4.0	5.9	3.6	2.4
Tri3	tar-MD	3.3	6.0	3.9	2.6
	exptl	-	6.3	3.7	3.3
	free-MD	4.1	5.6	3.5	2.2
Tri4	tar-MD	2.9	5.2[a]	4.3	2.8
	exptl	3.1	6.3	3.9	3.4
	free-MD	3.0	4.1	2.9	1.9
Tri5	tar-MD	3.0	3.0[a]	3.1	1.9
	exptl	2.3	4.8	-	2.7
	free-MD	3.3	4.5	2.9	2.0
Tri6	tar-MD	2.2	3.7[a]	4.2	2.6
	exptl	2.5	5.1	-	3.0
	free-MD	3.0	4.3	2.9	2.0
Tri7	tar-MD	3.6[a]	3.8	3.6	2.4
	exptl	2.1	4.8	4.0	2.7
	free-MD	3.6	4.9	3.1	2.0
Tri8	tar-MD	2.8	4.9	3.7	2.4
	exptl	2.8	5.3	3.5	3.1

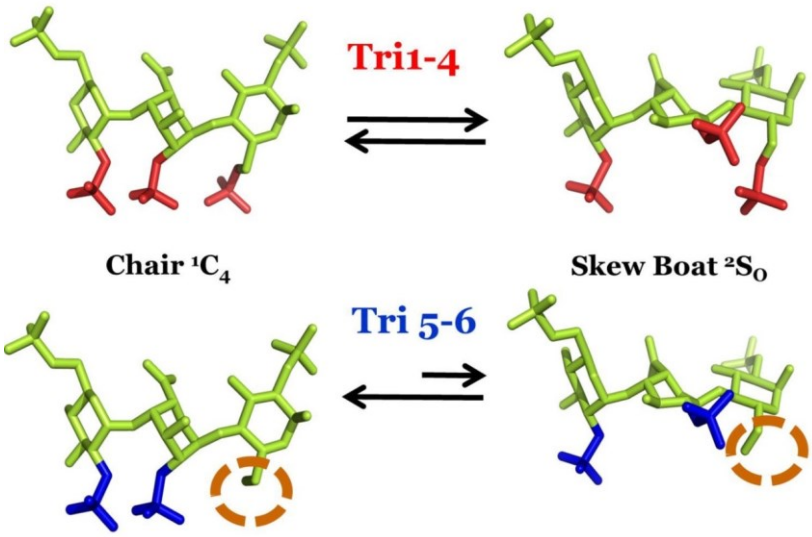

Figure 4. Influence of 6-O-sulphation (reducing end GlcN) on the L-IdoA2S conformational equilibrium. The absence of the 6-O-sulphate group (Glc(A)) is indicated with a dashed line circle.

About the global conformation, the distances defining the geometry around the GlcN-IdoA2S glycosidic linkage (H1c-H3b and H1c-H4b; see **table 1**) are pretty similar (from 2.6 Å to 2.7 Å) for all the trisaccharides, thus indicating a rather rigid *syn* conformation (no correlation with sulphation pattern). This observation is in agreement with the results obtained from MD simulations (see next section). With respect to the IdoA2S-GlcN glycosidic linkage, similar short H1b-H4a and H1b-H6a+a' distances (2.4-2.6 and 2.5-2.7 Å, respectively), corresponding to *syn*-Ψ conformers, have been determined. Differently, the *anti*-Ψ exclusive H1b-H3a distance shows significant variations upon sulphation pattern, going from longer (3.7 Å for **Tri6**) to shorter (2.6 Å for **Tri8**) distances. Although these differences are not clearly correlatable to the substitution pattern, interestingly, for the less sulphated trisaccharide **Tri8**, both the *syn*-Ψ and *anti*-Ψ exclusive H1b-H4a and H1b-H3a distances, respectively, present an equally short value (2.6 Å, intense NOE peaks; see **Tri8** NOESY spectrum in *Appendix*), thus indicating a very similar contribution of both conformers in solution. This result suggests that **Tri8** sulphation pattern, i.e, just N- and 2-O sulphation at the non-reducing glucosamine and the internal iduronate ring, respectively, enhances or facilitates the presence of the *anti*-Ψ conformations around the IdoA-GlcN glycosidic linkage. Probably, this is due to the reduced electrostatic repulsion forces existing in **Tri8** (less sulphated trisaccharide of the library).

Modelling

Previous $^3J_{HH}$-based studies on the conformational equilibrium of the L-IdoA2S ring in heparin derivatives are in agreement with our observations based on NOESY experiments (see NMR section), i.e., provided that the iduronate rings are not present at the terminal positions, the $^4C_1$ chair conformation does not participate in it [68-69] (or its contribution is too low as to be detected in NOESY

experiments). Thus, the conformational sampling of this ring is restricted to the $^1C_4$ chair and the ^2S_O skew-boat conformations, with the latter being involved in the characteristic pseudorotational equilibrium of the hexopyranose ring (see *Chapter 1*, *figure 3*). Furthermore, the balance of the chair to skew-boat equilibrium in internal iduronate residues depends on both their 2-O-sulphation and the substitution pattern of adjacent glucosamine residues[262]. In this regard, our experimental results for **Tri1-Tri8** suggest (see NMR section) that the population of ^2S_O conformer is strongly modulated by the presence or absence of 6-O-sulphation at the reducing terminal (at low temperature).

Molecular dynamics is a powerful and widely employed technique to study the molecular conformation and dynamics, allowing a better interpretation of the experimental observables. Combined to NMR, it is especially useful in estimating distributions of conformers in equilibrium[78, 85c, 263]. For the case of L-IdoA2S, since the characteristic inter-conversion rate between the $^1C_4$ and ^2S_O conformers occurs within the microsecond time scale[17], and therefore, far from being achieved by currently accessible simulation time, two alternative approaches can be used: either consider two different starting geometries for the iduronate residues and run two independent molecular dynamics simulations (unrestrained; free-MD), or use experimental observables as constraints in a single *time-averaged restrained* MD simulation (*tar*-MD).

Based on these grounds, we have studied the conformation and dynamics of **Tri1-Tri8** by both unrestrained and *time-averaged restrained* molecular dynamics simulations (free-MD and *tar*-MD, respectively), with explicit TIP3P[264] water molecules. Thus, we run a 20-nanosecond long free-MD simulation for each trisaccharide and for both L-IdoA2S conformations ($^1C_4$ and ^2S_O) as starting geometries, resulting in 16 independent MD simulations. Furthermore, an 8-nanosecond long *tar*-MD simulation with the iduronate ring adopting an initial $^1C_4$ chair conformation was also accomplished for each trisaccharide, using the NOE-derived H2b-H5b distance (exclusive NOE of the iduronate ^2S_O pucker) as a sole constraint.

It has to be noted that by the time we carried out the unrestrained MD simulations, there were not any force field for carbohydrates which included specific parameters and set of charges for sulphate and/or sulphamate groups. To overcome this technical limitation, the strategy followed was to combine the available parameters for sulphates and sulphamates (Altona´s[265]), which include a explicit hydrogen bond term with a Lennard-Jones 10-12 type potential[197, 266], with other sets of parameters for the carbohydrate moiety, water molecules and counterions developed under the same philosophy for consistency. Thus, the force fields *Parm91*[267] of Amber (for the water molecules and counterions) and *Glycam93*[268] (for the carbohydrate moiety) were used.

On the other hand, the *tar*-MD simulations were carried soon after the parameters and partial charges for sulphate and sulphamate groups had been released in the framework of *Glycam06*[220] force field, the latter entailing a significant improvement compared to *Glycam93*[268] force field (first version of *GLYCAM*). Thus, *Glycam06*[220] together with *Amber99SB*[269] parameters and partial charges were employed (both of them lack the explicit Lennard-Jones 10-12 term for hydrogen bonding).

It is important to highlight that, to simplify, from now on we will use the term *equatorial conformers* (or puckers) to refer to the set of conformers of the equatorial region of the Cremer-Pople sphere of an hexopyranose ring ($^{2}S_{O}$, $^{2,5}B$, $B_{3,O}$, $^{3}S_{1}$, etc.; see Chapter 1). On the other hand, when we say *pure* $^{2}S_{O}$ we will mean just the $^{2}S_{O}$ conformation.

Local conformation: plasticity of the L-IdoA2S ring

Regarding the free-MD approach, to quantitatively determine the populations of $^{1}C_{4}$ and $^{2}S_{O}$ puckers of the iduronate rings, we first monitored their four vicinal proton-proton dihedral angles (H1-C1-C2-H2, H2-C2-C3-H3, H3-C3-C4-H4 and H4-C4-C5-H5) along the 40000 frames of each $^{1}C_{4}$ and $^{2}S_{O}$ trajectory obtained (when a conformational transition was observed, only the previous frames were considered; see Methodology). These values were turned into vicinal proton-proton coupling constants, $^{3}J_{HH}$, by using the Haasnoot-Altona equation[270], which takes into account both the electronegativity and the orientation of the substituents on the H–C–C–H fragment, and averaged for each of the models (L-IdoA2S in $^{1}C_{4}$ or in $^{2}S_{O}$ conformation; see *Appendix*). Next, to obtain the populations of conformers ($^{1}C_{4}$ and $^{2}S_{O}$) of the L-IdoA ring in each trisaccharide, we carried out an iterative fit of the theoretical and experimental $^{3}J_{HH}$ according to 4 equations (one for each J-coupling; *eq. 1*) of the form

$$ {}^{3}J^{\exp}_{m(m+1)} = f_{(^{1}C_{4})} \cdot \langle {}^{3}J^{MD}_{m(m+1)}({}^{1}C_{4}) \rangle + f_{(^{2}S_{O})} \cdot \langle {}^{3}J^{MD}_{m(m+1)}({}^{2}S_{O}) \rangle \qquad \textit{Eq. 1} $$

In this expression, $f(^{1}C_{4})$ and $f(^{2}S_{O})$ are the unknown molar fractions of each conformer, $\langle {}^{3}J^{MD}_{m(m+1)}(\,) \rangle$ are the averages from MD and m is an index that runs from 1 to 4. Therefore, the experimental $^{3}J_{HH}$ coupling constants were considered as averages of the MD-derived ones for each conformer, weighted on the molar fraction of each. As the theoretical values were averages from MD simulations, they implicitly reflected the fluctuations around canonical conformations, which must be considered for this flexible hexopyranose ring[271], particularly to account for the pseudorotational conformational space in the case of the skew-boat conformer ($^{2}S_{O}$). In addition, the experimental measurements of $^{3}J_{HH}$ values at five different temperatures (278, 288, 298, 308 and 318 K) allowed us to monitor the population of conformers as a function of temperature (**figure 5**).

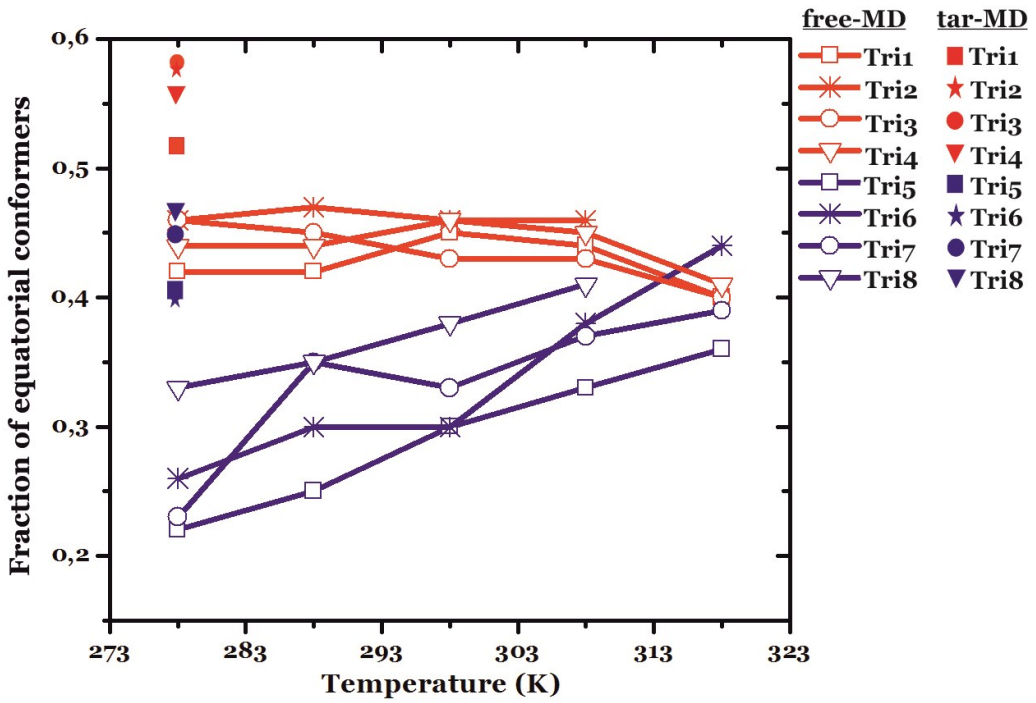

Figure 5. Temperature dependence of the populations of equatorial conformers of the central L-IdoA2S ring in the eight trisaccharides, classified by two chemical series, **Tri1-Tri4** (empty red symbols, red lines), and **Tri5-Tri8** (empty blue symbols, blue lines), and obtained by iterative fit of the NMR-derived and free-MD-calculated $^{3}J_{HH}$. Also, the equatorial puckers populations predicted by *tar*-MD simulations at 278 K are shown (filled symbols). Note that the **Tri1-Tri4** compounds showed significantly higher populations of equatorial conformers with both methods.

From the *tar*-MD approach, the populations of $^{1}C_4$ and equatorial conformers (only at 278 K) were determined by directly tracking the evolution of the Cremer-Pople puckering coordinates θ and ϕ over time, with θ undergoing transitions from the south pole (180º, $^{1}C_4$) to the equator (90º, equatorial puckers) and ϕ fluctuating according to the pseudorotational equilibrium of the iduronate equatorial conformers (ϕ is undefined in the poles). In principle, we thought that this approach was more accurate than the free-MD one as the latter may give rise to important deviations in the calculated values because it is subjected to 1) the experimental uncertainty of the coupling measurements, 2) the higher force field deviations of *Glycam93* force field, and 3) the goodness of the least-squares fit.

The results obtained from both methodologies are shown in **figure 5**, classified by the two series of trisaccharides (**Tri1-Tri4**, in red; **Tri5-Tri8**, in blue) and method (*tar*-MD, filled symbols; free-MD, empty symbols). Thus, the data revealed clear differences in the distribution of populations of conformers of the L-IdoA2S ring between both series, with **Tri5-Tri8** ensemble showing dependence with temperature (free-MD). Explicitly, conformational differences (**figure 5**) that are specific to the presence (**Tri1-Tri4** series) or absence (**Tri5-Tri8** series) of the 6-O-sulphate group on the reducing GlcN ring were observed. Particularly, the substitution of this bulky charged group with a neutral hydroxyl group (**Tri5-Tri8** series) made the populations of equatorial puckers very sensitive to

temperature (**figure 5**). At low temperature (273 K), the $^1C_4$ conformer was favoured in a larger extent in the **Tri5-Tri8** ensemble (65–80 % from free-MD; 55-60 % from *tar*-MD), while the equatorial conformers were promoted in the **Tri1-Tri4** series (47–43 % from free MD), with *tar*-MD simulations predicting their majority presence in solution at 273 K (53-58 %). As the temperature increased, the differences between both ensembles started to vanish (**figure** 5, free-MD).

Focusing on the *tar*-MD data, they indicated that, as long as a $6OSO_3^-$ group was present at the reducing end GlcN residue, the equatorial puckers mostly populated the conformational space sampled by the iduronate ring in heparin derivatives (**figure 5**, **Tri1-Tri4**), with the pure ^2S_O conformer being in all cases the predominant among the other puckers of the equator (**table 3**). Thus, whereas the presence of this functional group enhanced the ^2S_O population above 50%, its lack pushed the equilibrium towards a majority of the $^1C_4$ chair pucker (**figure 5**, **Tri5-Tri8**). It is noticeable that, in all cases, the populations of equatorial conformers obtained were higher than the $^3J_{HH}$ derived ones (free-MD), particularly in the **Tri5-Tri8** series (**figure 5**). Furthermore, when comparing the impact of 6-O-sulphation (Glc(A)) on the equatorial puckers populations of the L-IdoA2S ring (**Tri1-Tri4** versus **Tri5-Tri8** series), *tar*-MD simulations predicted smaller differences in pairs **Tri1:Tri5** (12%) and **Tri3:Tri7** (14%) compared to the results obtained from $^3J_{HH}$ fit (20 and 23 %, respectively; see *Appendix*). Interestingly, **Tri8** tendency towards equatorial conformers compared to its partners **Tri5**, **Tri6** and **Tri7** was in agreement with free-MD data, i.e, **Tri8** presented the highest population of equatorial puckers among the non-6-O-sulphated (at the reducing end GlcN residue) trisaccharides. In addition, and again in agreement with the $^3J_{HH}$ derived results, **Tri1** presented the lowest population of equatorial conformers among the 6-O-sulfated trisaccharides (**Tri1-Tri4**). So, whereas the D-GlcNS,6S-L-IdoA2S-D-GlcNS,6S sequence (**Tri1**) showed the lowest tendency within its group (**Tri1-Tri4**) to populate the equatorial puckers, for the D-GlcNS-L-IdoA2S-D-GlcNAc sequence (**Tri8**) the opposite conformational behavior was observed. These correlations with the sulphation pattern, obtained from both *tar*-MD and unrestrained MD approaches, indicated that as long the reducing end GlcN residue was 6-O-sulphated, the simultaneous presence of N- and 6-O-sulphation in the reducing and non-reducing terminal, respectively, promoted in some extent the $^1C_4$ chair conformer (**figure 5, Tri1**). On the other hand, when the reducing GlcN ring contained a 6-OH group (**Tri5-Tri8**), the population of equatorial puckers was enhanced provided that both the N- and 6- positions at the reducing and non-reducing terminals, respectively, were not sulphated either. Thus, although the reducing end GlcN 6-O-sulphation clearly shifts the conformational equilibrium of the L-IdoA2S ring towards the ^2S_O skew-boat conformer, the other substituted positions are also playing a role in the modulation of this equilibrium. In this regard, it is significant the small difference observed between **Tri1** and **Tri8** (6%), within the same range as with other partners of their **Tri1-Tri4** and **Tri5-Tri8** ensembles, respectively. For these reasons and because of the discrepancies between *tar*-MD and free-MD data regarding specific N- (GlcN(A)) and 6-O-sulphation (GlcN(C)), we suggest that higher level of theory calculations are needed to unambiguously and accurately determine the “subtle” contributions of the N-sulphation and 6-O-sulphation of the reducing end GlcN and the non-reducing end GlcN residues, respectively, on the conformational equilibrium of the L-IdoA2S ring.

From what we have discussed above, it seems reasonable to think that at the origin of the enhancement of the equatorial conformers upon 6-O-sulphation (Glc(A)) the change in the internal dynamics of the L-IdoA2S ring should play a role. Based on this hypothesis, the presence or absence of the 6-O-sulphate (Glc(A)) should be reflected on the reorientation properties of the torsions defining the L-IdoA2S puckering. Thus, we analyzed the internal auto-correlation functions ($C_{int}(t)$) for the vectors between vicinal protons of the L-IdoA2S residue (H1-H2, H2-H3, H3-H4 and H4-H5). We took fragments of the trajectories with the L-IdoA2S ring in $^1C_4$ conformation and just before a conformational transition towards the equatorial puckers (we might expect to observe more evident variations on $C_{int}(t)$ prior to a conformational change). Comparing **Tri2** and **Tri6**, the results (**Figure 6**; see also **Tri1:Tri5** pair in *Appendix*), indicated a significantly higher flexibility (lower value for the plateau) for the H2-H3 vector when the 6-O-sulphate group (Glc(A)) was present (**Tri2**), being the most flexible among the proton-proton vectors for **Tri2**. This correlates to the fact that the H2-C2-C3-H3 torsion participates in the $^1C_4 \rightarrow {}^2S_O$ conformational change to the largest extent. Therefore, 6-O-sulphation at the reducing-end GlcN residue seem to induce some strain on the H2-C2-C3-H3 torsion of the L-IdoA2S ring so that it promotes some additional flexibility on it (**Figure 6**; see also *Appendix*) that might be responsible of the higher tendency of **Tri1-Tri4** to populate the equatorial conformers. Interestingly, the H4-H5 vector is, on the contrary, significantly more rigid when a 6-O-sulphate group (Glc(A)) is present, so that it seems to compensate the higher flexibility of the H2-H3 one, in agreement with the *law of equipartition of energy states*[272].

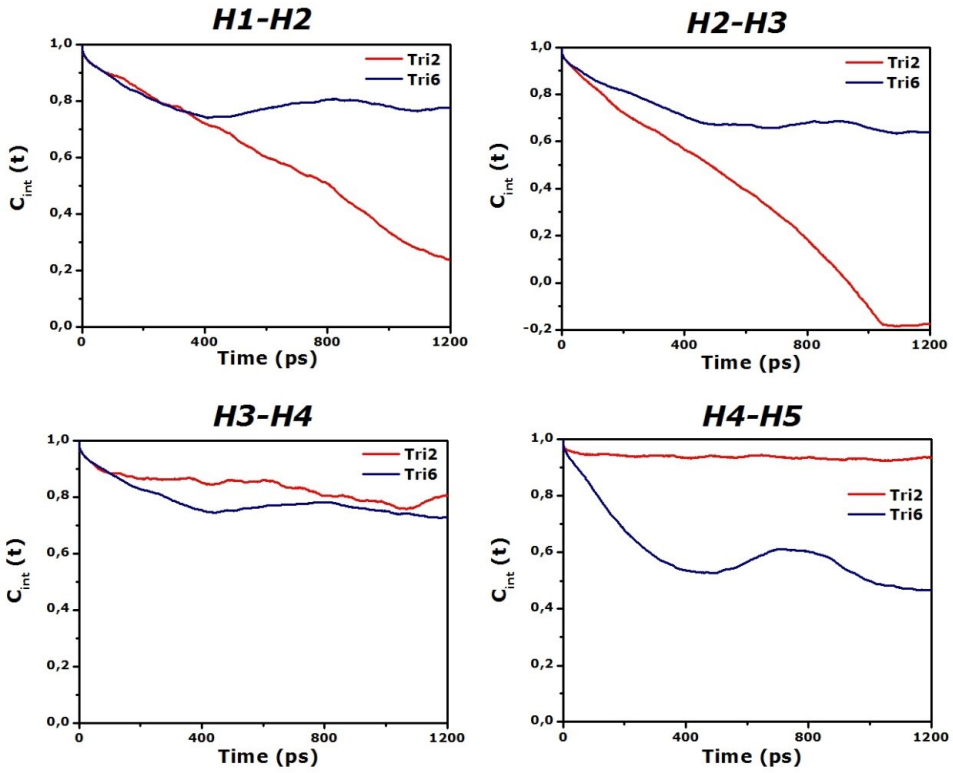

Figure 6. Comparison of the internal correlation function ($C_{int}(t)$) for the vicinal proton-proton vectors of the IdoA2S ring, in **Tri2** (red line) and **Tri6** (dark blue line) compounds. Only parts of the simulations with the iduronate ring in $^1C_4$ chair conformation have been used.

The analysis of the pseudorotational equilibrium of the L-IdoA2S residue predicted by *tar*-MD simulations (**table 3**), indicated that the pure $^{2}S_{O}$ pucker is specially favored in the presence of 6-O-sulphation at the reducing end GlcN residue (**Tri1-Tri4**). The trisaccharides lacking that 6-O-sulphate group (**Tri5-Tri8**) present a significantly lower population of pure $^{2}S_{O}$ skew-boat pucker (below 50 %), with the exception of **Tri8**. As such differences were not observed in the free-MD simulations (populations of the pure $^{2}S_{O}$ conformer of 60-80% in all cases; see *Appendix*), a more detailed discussion of these data is needed for a correct interpretation. First, the *tar*-MD methodology tries to drive the time-dependent averaged constraint (H2b-H5b distance) to eventually satisfy the range of values that we introduce as a constraint. Second, the $^{2}S_{O}$ skew-boat and $^{1}C_{4}$ chair conformations present canonical H2b-H5b distances of 2.4 Å and 4.0 Å, respectively (PDB code 1hpn). Third, the intra-ring H2-H5 distance of the L-IdoA2S residue presents a very high variability among the different puckers that participate in the pseudorotational equilibrium (θ=90º) of the iduronate ring, going from short (e.g. ≈2.5 Å in the $^{2,5}B$ boat; see *Appendix*), to intermediate (e.g. ≈3.5Å in the $B_{1,4}$ and $B_{3,O}$ puckers; see *Appendix*), and long distances (e.g. ≈4 Å in the $^{3}S_{1}$ and $^{1}S_{3}$ conformers, and ≈ 5Å in the $^{1,4}B$ boat; see *Appendix*).

Table 3. Comparing the populations of the different puckers involved in the pseudorotational equilibrium of the IdoA2S residue in the equatorial region of the Cremer-Pople sphere (the range 75º≤Θ≤105º has been taken) obtained from *tar*-MD simulations. The values shown represent the relative distribution frequency of 30º bin size centered at the corresponding value of the ϕ coordinate defining each pucker. A not observed conformer is indicated with a dash symbol (-). [a] Populations below 10%.

Compound	$^{3}S_{1}$ (ϕ=30º)	$B_{1,4}$ (ϕ=60º)	$^{5}S_{1}$ (ϕ=90º)	$^{2,5}B$ (ϕ=120º)	**$^{2}S_{O}$ (ϕ=150º)**	$B_{3,O}$ (ϕ=180º)	$^{1}S_{3}$ (ϕ=210º)	$^{1,4}B$ (ϕ=240º)
Tri1	a	a	a	a	**70%**	a	a	a
Tri2	a	a	a	a	**78%**	a	-	-
Tri3	-	-	-	a	**86%**	a	-	-
Tri4	-	-	a	10%	**79%**	a	-	-
Tri5	37%	10%	a	a	**39%**	a	-	-
Tri6	-	-	-	a	**44%**	10%	18%	13%
Tri7	26%	14%	a	a	**41%**	a	a	-
Tri8	a	a	a	a	**72%**	a	a	a

Thus, we propose that the observation of significant populations of puckers belonging to the latter group (long H2b-H5b distance) in **Tri5**, **Tri6** and **Tri7** (**table 3**) reflects a deviation caused by *tar*-MD calculations, which find energetically more favorable those puckers, compared to the $^{1}C_4$ chair, to satisfy the distance criteria. This is so because the torsional energy barriers involved in conformational changes among equatorial puckers are considerably lower than those in the $^{2}S_O \rightarrow {}^{1}C_4$ inter-conversion pathway[79]; indeed, some puckers in the equator act as bridging points for the inter-conversion between the equatorial $^{2}S_O$ skew-boat and polar ($^{1}C_4$ and $^{4}C_1$ chairs) conformations[79]. In addition, the fact that **Tri5**, **Tri6** and **Tri7** present the longest H2b-H5b experimental distance values (**table 2**) correlate with the observation of significant populations of $^{3}S_1$ (**Tri5** and **Tri7**), $^{1}S_3$ and $^{1,4}B$ (**Tri6**) puckers (long H2b-H5b distance, **table 3**). So, the population of $^{1}C_4$ chair and pure $^{2}S_O$ conformers for **Tri5**, **Tri6** and **Tri7** could be underestimated in favor of the above specified equatorial puckers.

Apart from the Cremer-Pople puckering coordinates, other parameters such as the flexible L-IdoA2S intra-ring proton-proton distances and vicinal coupling constants ($^{3}J_{HH}$) were monitored to estimate the quality of free-MD and *tar*-MD approaches in correctly reproduce the experimental data. Thus, referring to the L-IdoA2S $^{3}J_{HH}$ couplings (**table 2**), a general good agreement with the experimental values was obtained with both sets of MDs (see *Appendix*), with *tar*-MD performing better than free-MD. For the former, the main deviations were observed for the vicinal coupling constants between protons H2b and H3b (in 4 out of the 8 trisaccharides). However, those differences are acceptable because of the characteristic higher flexibility of the H2b-C2b-C3b-H3b torsion (changing from -60º in the $^{1}C_4$ chair to 172º in the pure $^{2}S_O$ skew boat) and the strong dihedral angle dependence of the Haasnoot-Altona curve[270] for the $^{3}J_{H2H3}$ coupling constant of the L-IdoA2S ring (see *Appendix*). Regarding the L-IdoA2S intra-residue distances (**table 1**), a good agreement was found between the NOE and molecular dynamics (both free-MD and *tar*-MD) derived values for the highly conformer-dependent H1-H3b and H2b-H5b distances (note: the H2b-H5b distance was used as convergence criteria in *tar*-MD simulations).

We consider that the results obtained from our *tar*-MD approach are more accurate than those derived from the combination of NMR and unrestrained MD simulations. This was possible due to the use of chemically pure heparin-like trisaccharides differently substituted with perfectly determined structures together with a *tar*-MD setup that allowed us to obtain a reasonable number of $^{1}C_4$ to $^{2}S_O$ (and vice versa) conformational transitions, and thus, to calculate the populations of L-IdoA2S conformers by direct tracking the evolution of the Cremer-Pople puckering coordinates θ and ϕ over time. Furthermore, we used 1) the unique experimental distance that contains information about the proportion of both L-IdoA2S conformers in equilibrium (the intra-ring H2b-H5b distance) as a sole constraint (thus, solely acting with an "external" penalty force at the local level) and 2) a consistent set of optimized parameters and partial charges for sulfated GAGs (instead of the combination of Altona's[231] and glycam93[268] parameters used for the free-MD simulations).[76] These settings facilitated *tar*-MD convergence to a large extent in the first case, and mitigated force field deviations in the second case (*Glycam93* performs worse than *Glycam06*)[220]. Importantly, the excellent agreement

between *tar*-MD-derived geometric parameters and the corresponding experimental observables ($^3J_{HH}$, intra- and inter-residue distances) pointed in this direction. Thus, we find our *tar*-MD approach very suitable to accurately study the conformational landscape of the L-IdoA2S ring, together with the influence of adjacent substitution patterns on it, within different iduronate-containing molecules in optimal computing and real times.

To sum up, the combination of NMR spectroscopy with unrestrained and time-averaged restrained MD simulations has demonstrated a significant influence of the sulphation pattern on the conformational equilibrium of an internal L-IdoA2S residue (**figure 5**). Notably, 6-O-sulphation (Glc(A)) seem to promote some additional flexibility on the H2-C2-C3-H3 torsion of the iduronate ring (see *Appendix*) that might be responsible of the higher tendency of **Tri1-Tri4** compounds to populate the equatorial conformers. Moreover, our *tar*-MD approach has put on evidence that, under the framework of the latest force field parameterizations for sulphated carbohydrates (GLYCAM06)[220] and the *tar*-MD methodology, the conformational plasticity of the iduronate ring together with the possible impact that some chemical modifications in the neighbouring residues may eject on it can be accurately quantified in easily accessible simulation time-scales and using the L-IdoA2S intra-ring distance between protons H2 and H5 as a sole constraint, with the latter facilitating an easier fine-tuning for the restrained simulation to converge. Supported by previously published data describing a similar effect of 6-O-suphation in chemoenzymatically synthesized macromolecular heparin[273], it seems reasonable to extrapolate our results (**figure 5**) for heparin-like trisaccharides (in which the GlcN residues are terminal) to longer chains. Thus, we could reasonably think that internal L-IdoA2S residues (with i positions; see **figure 1**) in heparin polysaccharides will experience, upon 6-O-sulphation of the i-1 GlcN residues (**figure 1**), a very similar impact on its conformational equilibrium than that observed for **Tri1-Tri8** compounds (**figure 5**). In the context of ligand-protein binding, as one major energy component in protein-heparin interactions is the electrostatic term, it is tempting to speculate that the positively charged amino acid side chains of proteins could reduce the electrostatic repulsion around the iduronate 2-O-sulphate group (translated into "internal energy stresses"), and thus affecting the conformational equilibrium of that ring upon binding.

Global geometry: interglycosidic torsions

First, to define the global geometries of the trisaccharides, we analysed the dynamic behaviour of their two glycosidic linkages by monitoring the torsions Φ (H1'-C1'-O-C4) and Ψ (C1'-O-C4-H4) along free-MD and *tar*-MD simulations. In this regard, the relative distribution curves obtained from free-MD (a, b) and *tar*-MD (c, d) are shown in **figure 7**.

For the GlcN-IdoA linkages, free MD simulations predicted distributions of Φ and Ψ dihedral angles centered at -60º and -50º, respectively (see *Appendix*), being narrower when the iduronate ring adopted the ^2S_O conformation and more flexible for the $^1C_4$ pucker (specially the Φ torsion; see *Appendix*). A parallel analysis for the IdoA-GlcN linkages revealed that when the L-IdoA2S residue was in $^1C_4$ conformation, the Φ torsion distributes between 0º and 60º (see *Appendix*), whereas for the

$^{2}S_{O}$ pucker this interval was somewhat extended to 75º (see *Appendix*), thus slightly increasing the accessible conformational space. Regarding the Ψ torsion, it was significantly more flexible (from -80º to 30º) and sensitive to the conformational state of the L-IdoA2S ring (see *Appendix*). In particular, for the L-IdoA2S $^{2}S_{O}$ conformation, important contributions of *anti*-Ψ conformers (Ψ = ±180º), over 30%, were found for **Tri1**, **Tri2** and **Tri6** (see *Appendix*), but not when the iduronate ring adopted the $^{1}C_{4}$ chair conformation. Similarly, significant differences among trisaccharides were observed regarding the relative populations of conformers at the two sub-minima (centred at -60º and 0º; see *Appendix*). Yet, these differences cannot be easily correlated to the sulphation pattern.

Furthermore, a description of the global geometries of the trisaccharides in solution must also take into account the different populations of conformers of the central L-IdoA2S ring. As the timescale for ring interconversion (microseconds)[17] is out of the feasible current molecular dynamics capabilities (in a reasonable real time), we included this effect by using the populations experimentally obtained (free-MD), as described in the previous section (**figure 5**). With these data, we weighted the ensembles of Φ and Ψ values on the populations of $^{1}C_{4}$ and $^{2}S_{O}$ conformers at 278 K (**figure 5** and *Appendix*). The results indicated that for the GlcN-IdoA linkages, the population-weighted distributions of Φ and Ψ were centred at -65º and -50º, respectively, in a narrow fashion (rigid) and without apparent influence of sulphation pattern (**figure 7a**). In *tar*-MD simulations, a slight shift of the frequency distributions was observed, centered at -45º and -30º for the Φ and Ψ torsions, respectively, with the latter showing a slightly wider distribution (**figure 7c**). Again, the sulphation pattern did not seem to eject any noticeable effect.

For the IdoA-GlcN linkages, the population-weighted distributions (free-MD) were centred at 40º for the Φ torsion, and 0º and -45º (two sub-minima) for the Ψ torsion, with the former being quite rigid and the latter behaving more flexibly (-80º to 30º range; **figure 7b**), as it is the common flexible behaviour of the IdoA-GlcN linkage[82]. On the other hand, *tar*-MD simulations predicted a narrow (rigid) distribution around 40º for the Φ torsion, and a flexible Ψ torsion with two maximums centered at 15º and -30º (**figure 7d**). Furthermore, both MD approaches indicated that the sulphation pattern did not seem to notably affect the conformational space sampled around the IdoA-GlcN linkages (**figure 7, b** and **d**). It has to be noted that no *anti*-Ψ conformations have been observed during *tar*-MD simulations. This is not surprising since the simulation time employed (8 ns) is too short for a thorough sampling around the glycosidic torsions. Furthermore, since the focus of the present work is the in-depth study of the conformational equilibrium of an internal iduronate ring, we have not included any IdoA2S-GlcN interglycosidic distance as a constraint. If so, whereas these restraints would have acted as catalyzers of the conformational transitions around IdoA2S-GlcN linkage, their use would have complicated the tuning of the *tar*-MD simulations. With longer *tar*-MD simulations we would possibly observe the *anti*-Ψ arrangements.

Other geometric parameters that define the global conformation of sugars are the interglycosidic distances (**table 1**). The results showed a general excellent agreement between the experimental and theoretical values for both the GlcN-IdoA2S (H1c-H3b and H1c-H4b distances) and IdoA2S-GlcN

(H1b-H4a and H1b-H6a+a') glycosidic linkages, with *tar*-MD better matching the NOE-derived distances, in general (**table 1**); this provided and additional proof of the quality of the *tar*-MD settings employed. An exception was observed for the H1-H3 distance of the IdoA2S-GlcN linkage (H1b-H3a), which presented the bigger deviations (**table 1**), towards higher values in all cases (outside the NOE distance range). A better agreement was observed for the free-MD derived H1b-H3a distance. This is due to the lack of sampling of the *anti*-Ψ conformations during *tar*-MD simulations, above mentioned, for which this distance is within the NOE range. Note that the highest deviations were observed for **Tri8**, which correlates to an important contribution of *anti*-Ψ conformers, as it was indicated in the NMR section.

Also, we have to highlight that, to evaluate the possible negative impact of using local constraints (L-IdoA2S) on the global conformation of the trisaccharides, at any point of their MD trajectories, we superimposed the Φ-Ψ points obtained in each frame on the corresponding relaxed Φ-Ψ maps (see *Appendix*). The results confirmed that the global conformation was not distorted during the restrained molecular dynamics (all the points fell into the energetically allowed regions). Therefore, we obtained "natural" (no artifacts) trajectories, at the global level, for **Tri1-Tri8**.

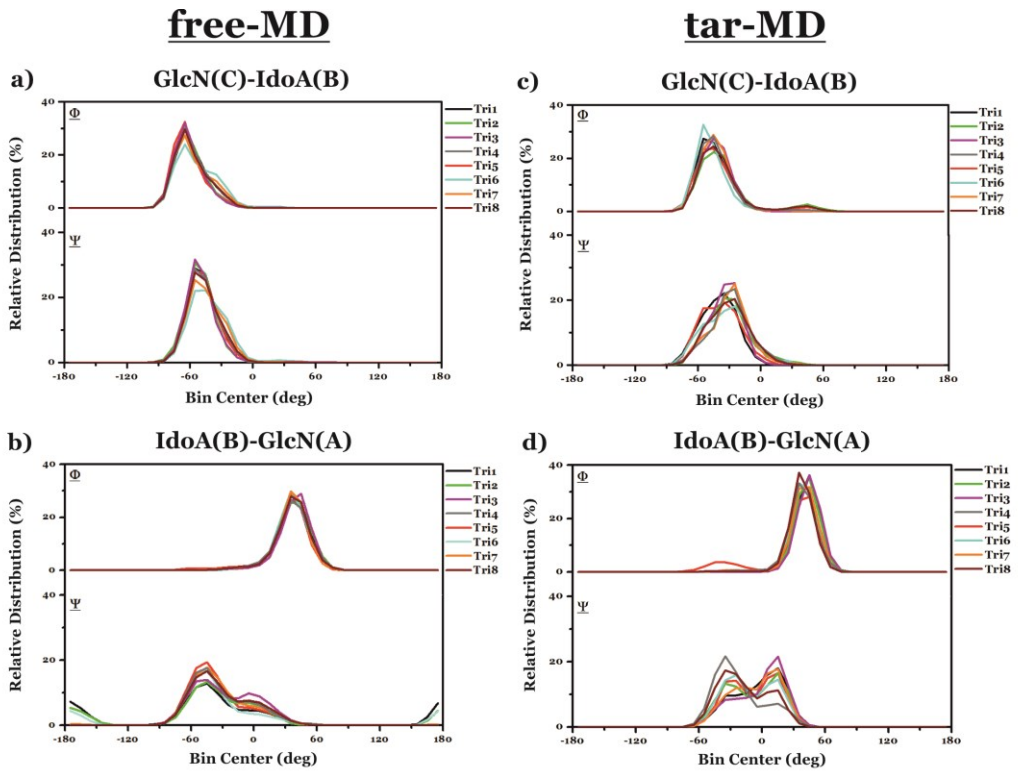

Figure 7. Conformational space sampled by the interglycosidic dihedral angles Φ and Ψ of the GlcN-IdoA (**a** and **c**) and IdoA-GlcN (**b** and **d**) linkages, predicted by both unrestrained (**a** and **b**) and time-averaged restrained (**c** and **d**; at 278 K) molecular dynamics simulations. The distribution curves corresponding to the unrestrained-MD (**a** and **b**) are population-weighted at 278 K. A bin size of 30º was used.

Regarding the orientation of the exocyclic groups (GlcN residues), which can be monitored as a function of the *ω* torsion (O5-C5-C6-O6), we analysed the possible effect of the substitution pattern on the trend to populate the different rotameric states (*gg*, *gt* and *tg*). Thus, for the non-reducing end GlcN ring, the *gg* (+60º) and *gt* (-60º) conformers were observed almost exclusively (the population of *tg* conformer is either null or negligible; see *Appendix*) in both free-MD and *tar*-MD simulations. Moreover, according to the free-MD data, the substitution of the 6-O-sulphate group with a 6-OH group (comparison of pairs **Tri1-Tri5**, **Tri2-Tri6**, **Tri3-Tri7** and **Tri4-Tri8**), which should in principle make the *gt* geometry more stable due to hydrogen bond formation with the intra-ring oxygen, indeed enhanced this rotamer in solution (see *Appendix*). On the other hand, for the reducing end GlcN residue, free-MD indicated that only the *gg* (+60º) and *gt* (-60º) conformers were energetically favoured, with the former generally prevailing over the latter (except for **Tri8;** see *Appendix*). Furthermore, since the substitution of the N-sulphate group with a N-acetyl group increased the population of *gt* conformer (see *Appendix*), a distant effect was hypothesized for this GlcN ring. In particular, we proposed the existence of stabilizing hydrophobic interactions between the acetyl and isopropyl groups that enlarges the conformational space for the *gt* conformer. Thus, by monitoring the N2-CH(isopropyl group) distance along each of the 16 molecular dynamics trajectories obtained, it could be identified that the substitution $NHSO_3^-$/NHAc shifts this distance to shorter values (see *Appendix*), in agreement with our proposed mechanism. On the other hand, the results obtained from *tar*-MD simulations do not correlate to the sulphation pattern. In this case, only **Tri1** and **Tri4** prefer the *gt* conformer. This was probably due to the shorter simulation time employed in *tar*-MD (8 ns) compared to the free-MD (20 ns) approach, which is probably not enough for the conformational sampling of the exocyclic torsion to fully equilibrate.

Solvation shells and hydrogen-bonding

The analysis of the radial distribution functions (rdf) from the independent $^1C_4$ and ^2S_O unrestrained-MD trajectories, permitted us to identify a generalized effect of the L-IdoA2S conformation on solvation. Indeed, the ^2S_O pucker showed the capacity to structure water molecules around the inter-glycosidic oxygen of the IdoA2S-GlcN linkage (O4c), therefore, indicating that the structuration of the molecular solvation shell is sensitive to the conformational state of the iduronate ring (**Figure 8**). Specifically, when the L-IdoA2S residue adopted the ^2S_O conformation, two solvation shells at 2 and 3 Å were identified, being absent for the $^1C_4$ pucker (**Figure 8**). Thus, these two solvation shells are exclusive of the skew-boat ^2S_O conformer, indicating the ability of this particular pucker to induce the structuration of water molecules, a very interesting observation in the context of L-IdoA2S interactions with proteins, in which the desolvation effects can become essential in some cases to understand binding affinity[274].

We also analysed the presence of inter-residual hydrogen bonds from the free-MD trajectories. For that purpose we chose a distance and angle cutoff of 3.0 Å and 120º, respectively (see Methodology). The results allowed us to identify a conserved hydrogen bond involving the oxygen ring of the L-IdoA2S residue (O5b) and the hydrogen H3O of the reducing end GlcN (see *Appendix*). It appeared

with an average occupancy of 57%, an average lifetime of 2.6 ps, an average distance of 2.4 Å and an average angle of 32º (**table 4**). Furthermore, it was not influenced by the conformation of the L-IdoA2S residue or the sulphation pattern. Also, another conserved hydrogen bond was observed between oxygen O5 of the non-reducing end GlcN and the hydrogen H3O of the L-IdoA2S residue. Nevertheless, this hydrogen bond presented, in average, low lifetimes and percentages of occupancy, below 1 ps and 20 %, respectively (**table 4**).

Note that both the rdf and hydrogen bonds analysis have been based on the free-MD results. *Tar*-MD simulations were not consider for the analysis of solvation and hydrogen bonding properties due to the short lifetimes of the L-IdoA2S puckers (fast conformational transitions), which did not permit to reliably study the differential effect of iduronate puckering on them.

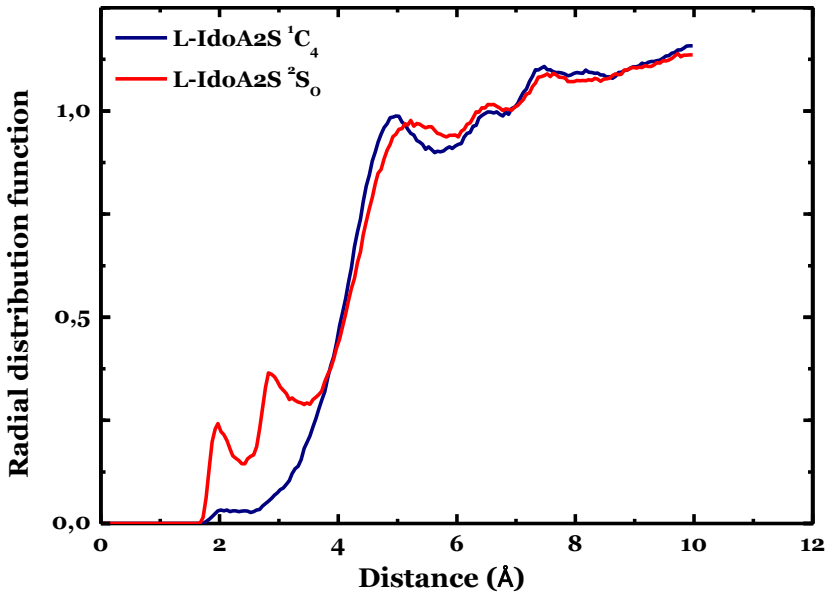

Figure 8. Comparison of the solvation profile of the inter-glycosidic oxygen of the IdoA2S-GlcN linkage (O4c) for both conformations of the L-IdoA2S residue.

Table 4. Set of averaged values for the percentage of occupancy, distance, angle and lifetime of the hydrogen bond O5(L-IdoA2S)-H3O(reducing-end GlcN) for the 16 trisaccharide models. Note: the hydrogen bond distance represents that between heteroatoms. The percentage of occupancy is defined as the number of frames, out of a hundred, of the trajectory for which a hydrogen bond exists under the chosen criteria (distance and angle cutoff of 3.0 Å and 120º, respectively, in this case).

Compound	L-IdoA2S conf.	Occupancy (%)	Distance (Å)	Angle (deg)	Lifetime (ps)
Tri1	$^1C_4$	62.9	2.8 ± 0.2	31.2 ± 12.7	3.5 ± 6.8
	^2S_O	24.6	2.9 ± 0.3	32.5 ± 14.0	2.1 ± 4.1
Tri2	$^1C_4$	64.1	2.8 ± 0.2	31.3 ± 12.9	3.2 ± 6.0
	^2S_O	45.7	2.9 ± 0.2	32.7 ± 13.8	2.3 ± 4.4
Tri3	$^1C_4$	60.7	2.9 ± 0.3	32.3 ± 13.7	2.3 ± 4.8
	^2S_O	64.5	2.9 ± 0.3	33.0 ± 13.8	2.2 ± 4.2
Tri4	$^1C_4$	67.6	2.8 ± 0.2	31.7 ± 12.9	3.2 ± 6.1
	^2S_O	64.4	2.9 ± 0.2	33.0 ± 13.6	2.3 ± 4.4
Tri5	$^1C_4$	59.3	2.9 ± 0.2	32.2 ± 13.8	2.3 ± 4.5
	^2S_O	60.1	2.9 ± 0.3	32.9 ± 14.0	2.1 ± 4.1
Tri6	$^1C_4$	68.1	2.9 ± 0.2	32.0 ± 13.1	2.9 ± 5.4
	^2S_O	24.4	2.9 ± 0.3	33.1 ± 14.0	2.0 ± 3.8
Tri7	$^1C_4$	67.5	2.9 ± 0.2	31.1 ± 13.0	3.1 ± 6.1
	^2S_O	55.3	2.9 ± 0.3	32.5 ± 13.9	2.2 ± 4.4
Tri8	$^1C_4$	68.4	2.8 ± 0.2	31.6 ± 12.8	3.2 ± 6.1
	^2S_O	61.9	2.9 ± 0.2	32.9 ± 14.0	2.1 ± 4.1

Hydrodynamics

For polysaccharides in general and, in particular, for glycosaminoglycans, the molecular motion and flexibility must be considered when discussing the molecular conformation, since the frequency, amplitude and geometry of such motions directly affect the measured NMR spectroscopic parameters[66b]. To formally describe a molecular motion, the use of a reorientational correlation function, or its corresponding density function, is required. These functions represent the time-dependent loss of orientational "memory", for any given vector in the molecular framework, with respect to its initial orientation, as a result of the internal motions and the overall molecular reorientation (the latter typically occurring on the picosecond to nanosecond timescale). While they take a value of 1 (maximum probability) at time zero, over time they decay to a "plateau" which can take a minimum value of zero. If motion is spatially restricted, the correlation function will decay to a value higher than zero, so that values closer to one are associated to rigid motions and those closer to zero are related to flexible reorientations. This is called the parameter of order S^2, which reflects the amplitude of motion (or freedom of reorientation) for the associated vector. Another parameter frequently used to deal with molecular motion is the correlation time, which is defined as the area under the correlation function. On the other hand, the spectral density function represents the frequency spectrum analogue of the reorientational correlation function, so that knowledge of one implies knowledge of the other.

In our research field, it is usual to employ the *model free approach* of Lipari and Szabo[275] to analyse fast internal molecular motions. Within this approach, internal motions give rise to the exponential decay of the correlation functions to a "plateau" value less than one. The validity of this model relies on a much faster timescale for any internal motion than for the overall reorientation ($\tau_{int} << \tau_o$), so that the slower process of overall tumbling in solution subsequently causes a further loss in the orientational correlation.

We determined the overall correlation time (τ_o) for each trisaccharide from the analysis of the internal molecular motions considering the *model free approach*[275] (see Methodology for details). The results (**Figure 9**) showed a very significant increase of τ_o for the **Tri1-Tri4** ensemble compared to the **Tri5-Tri8** one. Thus, the removal of the 6-O-sulphate group on the reducing GlcN ring gave rise to the reduction of the global correlation time and, therefore, a faster reorientation in solution for the trisaccharides belonging to the **Tri5-Tri8** ensemble. Interestingly, we observed that distinct sulphate groups exert a different influence on the overall correlation time. In particular, comparing 6-O-sulphation at the non-reducing and reducing terminal indicated that the former substitution augments τ_o to a smaller extent than the latter provided that the reducing end GlcN residue is N-sulphated (**Tri1: Tri3** and **Tri5:Tri7** pairs). On the contrary, when this is N-acetylated (**Tri2:Tri4** and **Tri6:Tri8** pairs) the overall correlation time is similarly enhanced. Regarding N-sulphation at the reducing terminal, its impact on τ_o was lower in all cases compared to the 6-O-sulphation at the same terminal (**Tri1:Tri2**, **Tri3:Tri4**, **Tri5:Tri6** and **Tri7:Tri8** pairs).

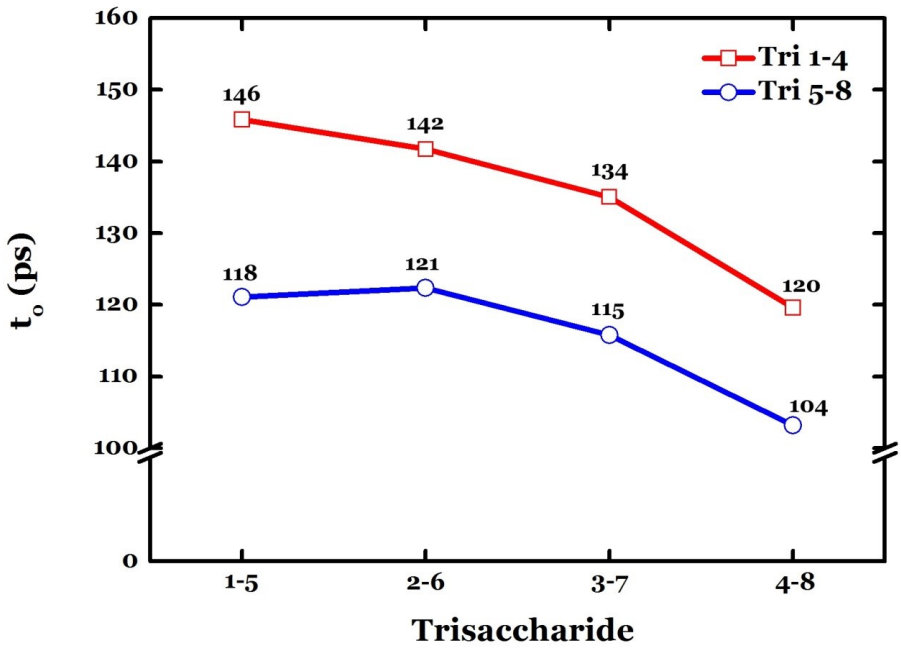

Figure 9. Population-weighted (at 278 K) overall correlation times (τ_o) calculated from unrestrained-MD simulations for the differently substituted **Tri1-Tri8** compounds. The exact τ_o values determined are shown on each point.

3.2 An inactive hexasaccharide sequence for the FGF-1 mitogenic activity

3.2.1 Background

FGF-1 is a member of the Fibroblast Factor family that, by forming a ternary assembly with heparin/heparan sulphate (HEP/HS) and the extracellular domain of membrane receptor FGFR2, triggers a signal that leads to different cellular essential functions (regulation of embryonic development, homeostasis and regenerative disorders). The key step for the activation of the FGF-1 signalling pathway is the formation of such a ternary complex (FGF1-HEP/HS –FGFR2) which gives rise to the dimerization of the receptors and, subsequently, the autophosphorylation that activates a mitogenic response through an enzymatic cascade.

Previously in our group, in the context of a wider research programme about the factors that regulate the activation of FGF - FGFR signalling pathway by glycosaminoglycans, some hexa- and octasaccharides containing the GlcN-IdoA repeating unit of the major sequence of heparin with diverse substitution patterns were prepared, tested and solved their structures[85]. Among them, we will focus our analysis in this chapter on the three hexasaccharides shown in **Figure 10**. While one of them represents the heparin regular region (**Hexa1**), the other two, **Hexa2** and **Hexa3**, display a non-axially symmetric sulphate distribution. On one side, the size of these molecules was chosen as the minimal chain length to be expected to stimulate FGF-1-induced mitogenic activity that could be obtained with a reasonable synthetic effort. On the other hand, the sulphation pattern was varied to obtain distinct distributions of electrostatic potential and assuming that **Hexa1-Hexa3** compounds would adopt a helix-like conformation in solution as found for heparin-like GAGs[66b]. This was later confirmed by NMR spectroscopy and MD simulations for **Hexa1**[85a] and **Hexa2**[276]. However, the conformational analysis of **Hexa3** has not been carried out until now.

Hexa2 was designed based on the X-ray structure published by DiGabriele *et al.*[87], which indicated that for heparin oligosaccharides to interact to FGF-1 the formation of a *trans* dimer was necessary (each heparin side interacting with one FGF-1 molecule). Thus, **Hexa2**, displaying sulphate groups on just one side of the helix, would bind FGF-1 in a much less extent (or would not interact) compared to **Hexa1**, so that the induced mitogenic activity would be greatly diminished. Surprisingly, **Hexa2** activated FGF-1 as effectively as an octasaccharide of the heparin regular region[67]. Furthermore, it was later demonstrated that GAGs induced FGF-1 dimerization either in a *cis* or *trans* disposition with respect to the heparin chain is not an absolute requirement for biological activity[88].

On the other hand, **Hexa3**, which presents a pseudo-palindromic relationship with **Hexa2**, was synthesised[85d] with the aim to interact simultaneously with FGF-1, at sub-site **a**, and the receptor FGFR (see *Chapter 1* and **figure 17**). Thus, the mitogenic activity would be maximized, as it was proposed by Pellegrini from the analysis of several crystallographic structures of FGF-1 and/or

FGFR-heparin oligosaccharides complexes[277]. However, when **Hexa3** was subjected to the biological assay, it resulted to be inactive[85d]. Interestingly in contrast to the highly active **Hexa2** sequence, this result demonstrated that the presence of the recognition sites in the oligosaccharide sequence is not enough to trigger the biological process, putting on evidence the complexity of FGF-1 activation. Furthermore, a previous study employing different synthetic oligosaccharides demonstrated that small variations in the sequence, size or sulphation pattern may dramatically impact their capacity to induce mitogenic activity[67].

Since the original work with **Hexa3**[85d] did not include a detailed structural analysis[85d], we decided to obtain its structure with the highest possible resolution to analyse the potential reasons for its unexpected lack of activity. To do so, the combination of NMR spectroscopy and molecular dynamics calculations was determinant.

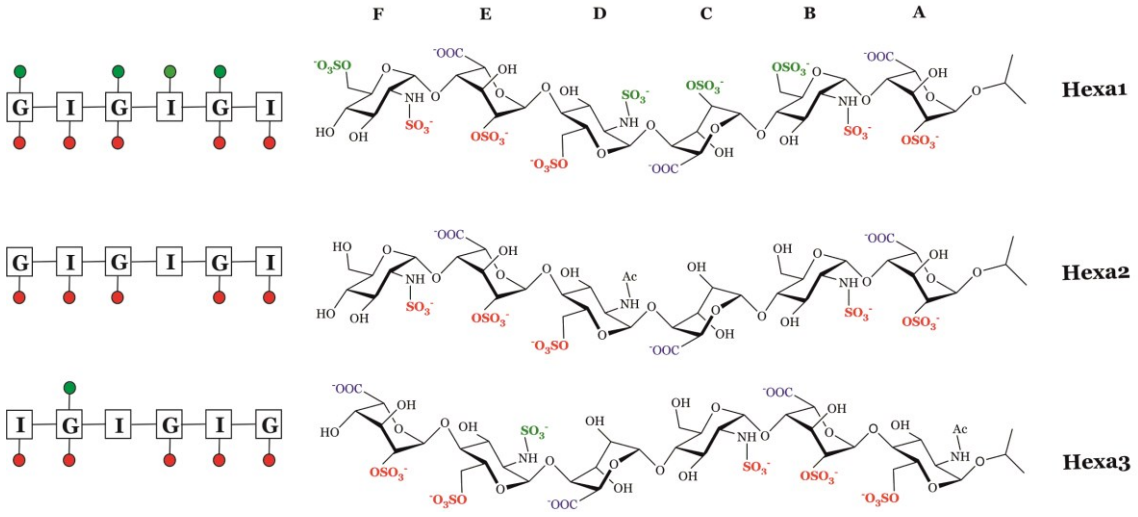

Figure 10. Structure (right) and schematic (left) representation of 3 different hexasaccharides presenting the heparin regular region (**Hexa1**), sulphate groups on just one side of the helix (**Hexa2**), and a pseudo-palindromic relation with the latter (**Hexa3**). The 3D structure of heparin helix (PDB code 1HPN[65]) has been considered in this 2D structure representation (right), showing the relative disposition of the sulphate groups.

3.2.2 Results and discussion

Nuclear Magnetic Resonance

It is well known that heparin oligomers longer than tetrasaccharides are anisotropic[278], exhibiting an hydrodynamic top rotor behaviour. Therefore, they rotate with different correlation times along the transversal (short) or the longitudinal (long) molecular axis[278a, 279]. Both **Hexa1** and **Hexa2** compounds exhibit such hydrodynamic anisotropic behaviour originated in the rigidity of their glycosidic bonds that it could reflect the electrostatic repulsion of the negatively charged groups (sulphates and carboxylates). In previous works, this behaviour was studied in terms of the differences between the perpendicular ($\tau_{\perp}$) and parallel ($\tau_{||}$) correlation times derived from NOE experiments and

complete ^{13}C relaxation analysis for **Hexa1** and **Hexa2**[279]. About **Hexa3**, given its high chemical similarity with those, we should expect it to behave anisotropically as well.

In an anisotropic molecule, the NOESY or ROESY based distances calculated using the *Isolated Spin Pair Approximation* (ISPA)[261] using a single distance (and therefore a single correlation time) as a reference are not accurate because they depend on the angle between the interprotonic vector and the molecular axis, which governs the correlation time of the vector. As an alternative method, it was proposed to use several reference distances selected to cover the range of possible orientations with respect to the rotation axis[85a-c]. However, in this case the accuracy needed to distinguish some of the characteristic features of heparin-derived oligosaccharides was lost. Thus, the *off-resonance* ROESY methodology [172] appeared as a more precise method to calculate distances from dipolar relaxation data without a pre-assigned model of motion (out from the isolated spin pair approximation (ISPA)[261]), allowing to simultaneously obtain the correlation time and distance for each pair of protons. This method relies on the determination of several *off-resonance* ROESY values by varying the tilted angle of the effective ROESY spin-lock field to achieve enough amount of independent data as to extract the correlation time for each vector and, from this, to calculate each interprotonic distance.

For the reasons above mentioned, we employed *off-resonance* ROESY spectroscopy[172] as experimental approach to study the hydrodynamics properties of **Hexa3**. Therefore, we could extract the cross relaxation rates σ_{NOESY} and σ_{ROESY} and the effective correlation times, τ_{eff}, by measuring a linear combination of NOE and ROE effects controlled by the spin lock offset. This way, we were able to accurately derive proton-proton distances for the anisotropic **Hexa3** molecule independently of their relative orientation with respect to the molecular axis.

Hexa3 spectra were assigned by standard procedures, i.e., identifying the spin systems of each ring by scalar coupling and interconnecting them via interglycosidic NOEs. On the other hand, arrays of several series of *off-resonance* ROESY experiments at several mixing times for different tilted angles (6, 10, 20 and 30 kHz) were recorded to determine the distances. All the growth curves for each proton and each tilted angle were linearly fitted. Then, calculating the growth rate of several series of *off-resonance* ROESY experiments corresponding to different spin lock offsets, the σ_{NOESY}, σ_{ROESY} and τ_{eff} parameters were independently obtained for each proton pair (**table 5**; see Methodology, *eq*. 3 and 7), and then used to calculate the experimental interprotonic distances (see Methodology, *eq*. *8*).

The diffusional anisotropy has also been studied for **Hexa3** in terms of the anisotropy factor, i.e., the quotient between the parallel ($\tau_{||}$) and perpendicular ($\tau_{\perp}$) correlation times ($\tau_{\perp}/\tau_{||}$). The most precise method to calculate these parameters ($\tau_{\perp}$ and $\tau_{||}$) is by COmplete Matrix Relaxation Analysis (CORMA)[280] based on ^{13}C T_1 and T_2 and heteronuclear NOE measurements. However, this method is time consuming and, reasonably, it can be replaced by considering the relationship between the correlation times of two orthogonal interprotonic vectors of similar distances, with one of them being aligned with respect to the molecular axis. For **Hexa3**, we have chosen the vectors H1-H2 and H2-H4

of glucosamines C and E because 1) they are roughly orthogonal to each other and 2) vectors H1-H2 are nearly parallel to the molecular axis (**Figure 11**). Therefore, the effective correlation times, τ_{eff}, of H1-H2 and H2-H4 proton pairs (derived from *off-resonance* ROESY experiments; see **table 5**) approximately represent the $\tau_{\perp}$ and $\tau_{||}$ correlation times, respectively, associated to **Hexa3**. Thus, we calculated the anisotropy factor as the $\tau_{eff}^{H1-H2}/\tau_{eff}^{H2-H4}$ ratio.

The average anisotropy factor obtained for **Hexa3** was 1.2 ($\tau_{\perp}/\tau_{||}$ equal to 1.2 considering both C and E GlcN rings). The comparison (of the $\tau_{eff}^{H1-H2}/\tau_{eff}^{H2-H4}$ ratio) to those previously reported for **Hexa1** ($\tau_{\perp}/\tau_{||}$ = 1.5)[85a] and **Hexa2** ($\tau_{\perp}/\tau_{||}$ = 1.5)[276], obtained by a similar methodology, clearly indicated that, although **Hexa3** exhibits an anisotropic behaviour, it is slightly less anisotropic than **Hexa1** and **Hexa2** compounds.

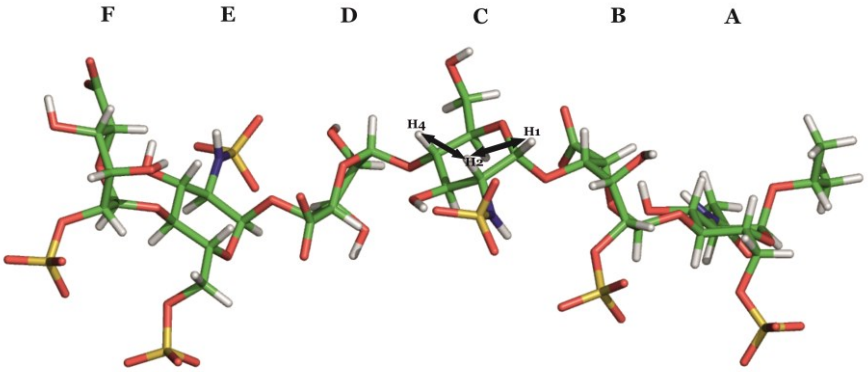

Figure 11. 3D structure of the average conformation (obtained over 500 ns of MD trajectory) of **Hexa3**, indicating with arrows the H1-H2 and H2-H4 vectors for ring C (D-GlcNS), roughly parallel and perpendicular to the molecular axis, respectively.

NOESY experiments were also registered for **Hexa3**. The spectra showed a varied behaviour of the iduronate rings (B, D and F), which depended on their position and substitution along the hexasaccharide chain (**Figure 12**). For instance, the non-reducing end L-IdoA2S residue (ring F) did not give rise to the exclusive H2F-H5F NOE characteristic of the presence of the $^{2}S_{O}$ skew-boat conformation (**Figure 12**). On the other hand, two $^{2}S_{O}$-exclusive H2-H5 NOE cross-peaks of medium-weak and weak intensity were identified, corresponding to the internal L-IdoA2S (B) and L-IdoA2OH (D) rings, respectively (**Figure 12**). Thus, we observed that, in **Hexa3**, the $^{2}S_{O}$ skew-boat conformer only participates in the conformational equilibrium of the non-terminal iduronate residues (B and D), although to a less extent (weaker NOE intensity) in that of the non-sulphated iduronate ring (D, L-IdoA2OH). Also, the NMR-derived (NOESY and *off*-resonance ROESY) H2-H5 and H1-H3 intra-ring distances of the iduronate ring (**table 6**), both in the NOE range for the $^{2}S_{O}$ conformer, showed a shorter distance for the L-IdoA2S ring B than the L-IdoA2OH ring (D), thus indicating the highest contribution of the $^{2}S_{O}$ conformer in the equilibrium of the former. This 2-sulphation effect on the conformation of the iduronate residue has been previously reported[10, 281]. In conclusion, the

NOESY experiments showed that the proportion of iduronate ^2S_O conformer decreases along the **Hexa3** chain from the reducing to the non-reducing end, being larger in B than in D ring, and absent or negligible in F residue. By combining NMR and MD-derived $^3J_{HH}$ couplings, these differences were translated into populations of ^2S_O pucker, which are discussed in the next section (**table 7**).

Finally, regarding the geometry around the glycosidic linkages that define the global conformation in oligosaccharides, the presence of the pairs of intense H1'-H3&H1'-H4 and H1'-H4&H1'-H6 interglycosidic NOEs corresponding to the GlcN-IdoA and IdoA-GlcN linkages, respectively, indicated a major *syn* rearrangement (**figure 13**). Furthermore, the absence of the H5'-H6, H1'-H5 and H1'-H3 NOEs for the IdoA-GlcN linkages, exclusive of the *anti*-Ψ rearrangements, confirmed the unique presence of *syn*-Ψ conformations, and consequently, the rigidity of the **Hexa3** backbone (**figure 13**).

Table 5. σ_{NOESY} and σ_{ROESY} cross relaxation rates and effective correlation times (τ_{eff}) of the intra- and inter-residue proton pairs of **Hexa3**, calculated from *off-resonance* ROESY[172] measurements.

	σ_{NOESY} (s^{-1})	σ_{ROESY} (s^{-1})	τ_{eff} (ns/rad)
H1B-H4A	-0,09	0,61	0,50
H1B-H6A*	-0,03	0,17	0,51
H1C-H3B	-0,05	0,20	0,59
H1C-H4B	-0,07	0,40	0,53
H1D-H4C	-0,03	0,18	0,50
H1D-H6C	-0,07	0,41	0,51
H1D-H6'C	-0,05	0,16	0,75
H1F-H4E	-0,08	0,69	0,46
H1F-H6E	-0,13	0,83	0,50
H1F-H6'E	-0,04	0,12	0,85
H1B-H3B	-0,04	0,17	0,57
H4B-H5B	-0,11	0,54	0,57
H1C-H2C	-0,02	0,07	0,65
H2C-H4C	-0,01	0,07	0,54
H1D-H2D	-0,03	0,19	0,51
H1E-H2E	-0,01	0,06	0,62
H2E-H4E	-0,01	0,07	0,50
H4F-H5F	-0,05	0,47	0,45

* Values assigned to each diastereotopic hydrogen (*proR* and *proS*) under the approximation that the most populated conformer around the GlcN(C) ω torsion is *gg* (in agreement to MD results, see below).

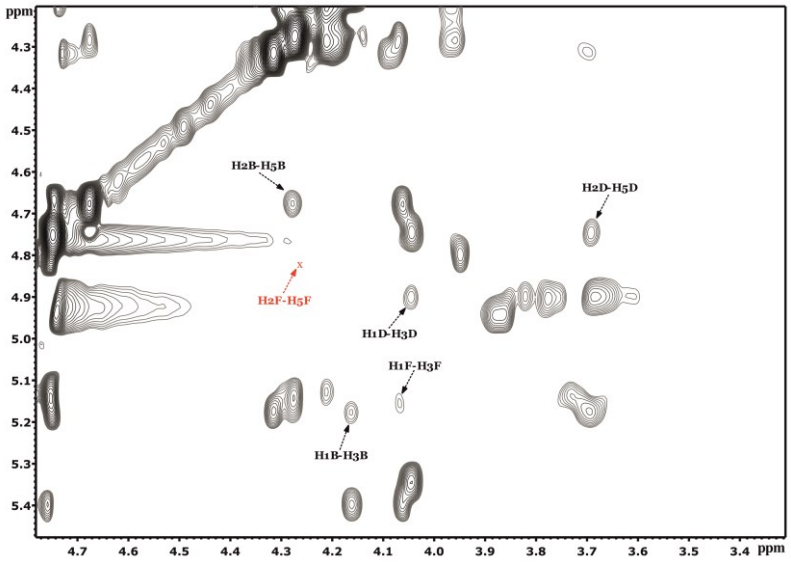

Figure 12. Expansion of a NOESY experiment registered at 400 ms mixing time for **Hexa3** showing the signals corresponding to the $^{2}S_{O}$-exclusive NOE cross-peaks of the iduronate residues (B, D and F rings). The absence of the H2-H5 cross-peak for the non-reducing end L-IdoA2S residue (H2F-H5F) is indicated with a cross symbol, in red. Also note the very low intensity of the $^{2}S_{O}$-exclusive H1-H3 NOE for ring F compared to those of the internal iduronate rings B and D.

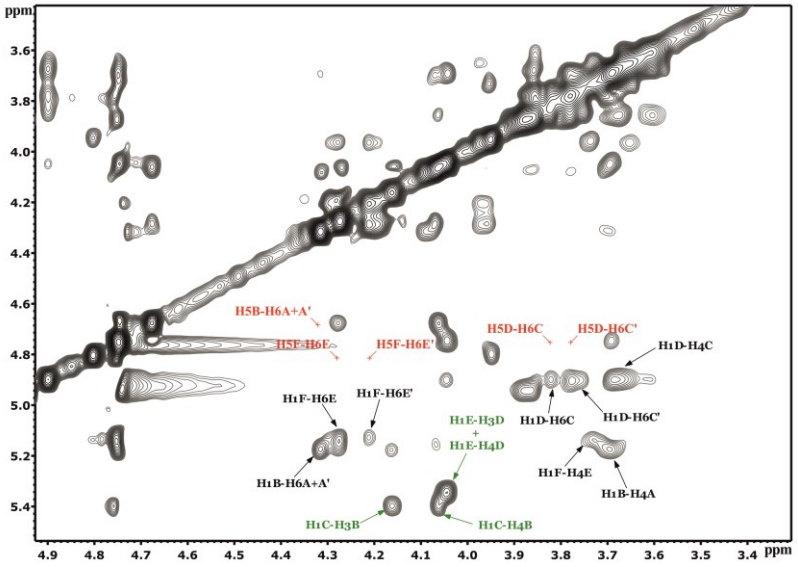

Figure 13. Expansion of a NOESY experiment registered at 400 ms mixing time for **Hexa3** showing the cross-peaks corresponding to a *syn* rearrangement around the GlcN-IdoA (H1'-H3 and H1'-H4 NOEs, in green) and IdoA-GlcN glycosidic linkages (H1'-H4 and H1'-H6 NOEs, in bold black), respectively. The non-observed H5'-H6 NOEs, exclusive of the *anti*-Ψ rearrangements around the IdoA-GlcN linkages, are labelled in red.

Modelling

With the aim to get a deeper insight into the structural properties of **Hexa3**, an intensive molecular dynamics simulation (500 ns) was performed in explicit water, with Na^+ as counterions for electroneutrality. To do so, the last GLYCAM parameters together with GLYCAM partial charges (GLYCAM06)[220], which provide a consistent set of simulation conditions suitable for the modelling of glycosaminoglycans, were employed. AMBER 12 was used as a platform for MD. After the appropriate equilibration steps we run a 500 ns MD simulation in explicit TIP3P water molecules, Periodic Boundary Conditions (PBC) and using the Particle Mesh Ewald method[282] (PME). The results were consistent with iduronate residues adopting a stable $^1C_4$ chair puckering with scarce transitions towards the ^2S_O skew-boat conformer (see *Appendix*), and narrow *syn*-Φ and *syn*-ψ distributions around the glycosidic linkages (**figure 14**).

Starting with the analysis of the particular conformational flexibility of iduronic acid, although the simulation time was not enough to thoroughly sampling the conformational equilibrium of the iduronate rings (only a few short $^1C_4$ to ^2S_O transitions, see *Appendix*), the MD-derived H2-H5 distance was slightly shorter for ring B, in agreement with the decrease of ^2S_O skew-boat population from B to F ring observed by NMR (**tables 7** and **8**).

We also analysed the conformational equilibrium of the L-IdoA residues assuming a mixture the $^1C_4$ and ^2S_o puckers in fast equilibrium in the chemical shift timescale (**table 7**). To determine the populations of conformers in equilibrium we performed an iterative fit using the canonical and NMR-derived proton-proton vicinal coupling constants, following the procedure described for the trisaccharides **Tri1-Tri8** (*eq. 1*). Two different sets of canonical $^3J_{HH}$ values for the $^1C_4$ and ^2S_O conformers were used (corresponding to a α-d-IdoA2S-OH monosaccharide in Hricovíni 2006[283] and a α-D-GlcNS,6S-(1→4)-α-D-IdoA2S-OMe disaccharide in Hricovíni 2011[284]; see **table 7**). The calculated populations followed the same tendency for both sets of canonical $^3J_{HH}$ couplings, with the ^2S_O pucker population decreasing from the reducing (B) to the non-reducing end (F). Interestingly, when ring F was subjected to this analysis, an appreciable population of ^2S_O conformer (17 - 19%) was obtained. With these data, the H2-H5 distances were weighted on the populations of conformers, obtaining a good correlation and agreement with the experimental ones (**table 8**). However, the population-weighted H2-H5 distance obtained for ring F was within the NOE range (3.1 Å), which contradicts the experimental evidence (absence of the corresponding NOE cross-peak). Furthermore, for the other two rings, the population-weighted distances were shorter (0.2-0.5 Å) than the experimental ones. This indicated that the ^2S_O-pucker populations derived from $^3J_{HH}$ fit were probably overestimated for the 3 iduronate rings, thus giving rise to shorter population-weighted H2-H5 distances.

Table 6. Intra- and inter-residue distances of **Hexa3**. [a]Distances calculated as $<r^{-3}>^{-1/3}$ and $<r^{-6}>^{-1/6}$ from 500 ns of MD simulation. [b]Distances derived from ^{1}H-NMR *off-resonance* 2D-ROESY experiments. [c]2D-NOESY derived distances using the Isolated Spin Pair Approximation (ISPA)[261]; the H1-H2 constant distance of the GlcN (E) residue, 2.5 Å, was used as reference.

Proton pair	Theoretical distances, MD (Å)[a]		Experimental distances (Å)[b]
	$<r^{-3}>^{-1/3}$	$<r^{-6}>^{-1/6}$	
H1B-H4A	2.3	2.3	2.3
H1B-H6A*	3.7	3.3	2.8
H1C-H3B	2.5	2.4	2.8
H1C-H4B	2.4	2.4	2.4
H1D-H4C	2.3	2.3	2.8
H1D-H6*proR***	2.9	2.7	2.4
H1D-H6*proS***	3.3	3.0	3.0
H1E-H3D	2.3	2.3	-
H1E-H4D	2.5	2.5	-
H1F-H4E	2.3	2.3	2.2
H1F-H6*proR*	3.2	2.9	2.2
H1F-H6*proS*	4.0	3.6	3.2
H1B-H2B	2.6	2.6	-
H1B-H3B	4.2	4.2	2.9
H2B-H5B	3.9	3.9	3.0[c]
H4B-H5B	2.4	2.4	2.3
H1D-H2D	2.6	2.6	2.8
H1D-H3D	4.2	4.2	-
H2D-H5D	4.0	3.9	3.3[c]
H4D-H5D	2.4	2.4	-
H1F-H2F	2.6	2.5	-
H1F-H3F	4.2	4.2	4.0[c]
H2F-H5F	4.0	4.0	-
H4F-H5F	2.4	2.4	2.3

*Theoretical distances obtained from r^{-3} and r^{-6} average, respectively, over the *proR* and *proS* values. **Values assigned to each diastereotopic hydrogen (*proR* and *proS*) under the approximation that the most populated conformer around the GlcN(C) ω torsion is *gg* (in agreement to MD results, see below).

It hast to be noted that the $^3J_{HH}$ –derived populations of ^2S_O pucker for **Hexa3** (**table 7**) were lower than those previously reported for **Hexa1**[85a] and **Hexa2**[82]. As an exception, for ring B (**Hexa3**) similar conformers distributions as hexasacharides **1** and **2** have been observed. On the contraty, a different behaviour was obtained for the internal non-sulphated iduronate D (**Hexa3**), showing a lower tendency to populate the ^2S_O state (12-13%)[82] than **Hexa2**. Regarding ring F (**Hexa3**), compared to **Hexa1** and **Hexa2** the largest discrepancies observed are associated to its position at the non-reducing terminal of the chain, previously reported to present significantly lower ^2S_O populations[69a]. It should be also kept in mind that the differences observed could also be partially due to the uncertainty associated to goodness of the $^3J_{HH}$ fit (**table 7**).

Table 7. Population of the skew-boat ^2S_O pucker of the L-IdoA2S rings calculated by multi-parametric fit of the experimental and canonical (of the $^1C_4$ and ^2S_O conformers) proton-proton vicinal coupling constants ($^3J_{HH}$). In brackets, the reduced chi-square , χ_{red}^2, of the fit is shown. [a] Fit with canonical $^3J_{HH}$ values calculated by Hricovíni[284], at the B3LYP level of theory and including the solvent and Na^+ ions effect. [b] Fit with canonical $^3J_{HH}$ values calculated by Hricovíni[283], at the B3LYP level, for a IdoA2S monosaccharide.

	IdoA2S (B)	**IdoA2OH (D)**	**IdoA2S (F)**
Hricovíni 2006[a]	40% (0.20)	33% (0.41)	17% (0.09)
Hricovíni 2011[b]	37% (0.27)	32% (0.44)	19% (0.37)

Table 8. Population-weighted H2-H5 distance of each iduronate ring compared with the experimental values determined from 2D-NOESY experiments. The $^1C_4$ and ^2S_O puckers populations shown in **table 7** have been used. The canonical H2-H5 distance for both the $^1C_4$ and ^2S_O conformers, 4 Å and 2.4 Å, respectively, have been taken from the NMR resolved structure of heparin with PDB code 1HPN.

	IdoA2S (B)	**IdoA2OH (D)**	**IdoA2S (F)**
Hricovíni 2006[283]	2.8	2.8	3.1
Hricovíni 2011[284]	2.8	2.9	3.1
Experimental	3.0	3.3	undetected

Related to the overall molecular shape of **Hexa3**, the agreement between the experimental and MD-derived interglycosidic distances was very good in general (**table 6**). The experimental distances fit well with the NOE calculated distances from both r^{-3} or r^{-6} averages, indicating that **Hexa3** either performs an intermediate hydrodynamic behaviour so that it fits to both models of motion, or that the distances do not fluctuate significantly (**table 6**).

An additional characteristic of the conformation of L-IdoA containing GAGs is the rigidity of its backbone in spite of the conformational equilibrium of the iduronate residues. A possible explanation for this behaviour could be the electrostatic repulsion between the charges of the iduronate 2-O-sulphate groups and those present on the same side in adjacent GlcN residues. The distributions of the values of glycosidic linkage dihedral angles Φ and Ψ along the MD simulation are shown in the **figure 14**. The results were consistent with the known behaviour of the GlcN–IdoA glycosidic linkages being more rigid than the IdoA–GlcN ones, and with the Ψ torsion being more flexible than Φ for the latter linkage (**figure 14**). Comparing the overall shapes (backbone) of **Hexa3** and **Hexa2** (**figure 15**) demonstrated they both present very similar conformations at the global level, i.e, the extended helix-like shape of heparin. Furthermore, the effect of the glycosidic linkages of the central iduronate ring (D) on the stiffness of the backbone was smaller than for the other linkages (more flexible D-C linkage; see **figure 14**), reflecting the absence of the sulphate group in position 2 (L-IdoA2OH). In addition, the MD simulation predicted the increase in the degree of flexibility from the centre to the ends of the hexasaccharide chain (**figure 16**), as it has been previously described for heparin-like oligosaccharides[273].

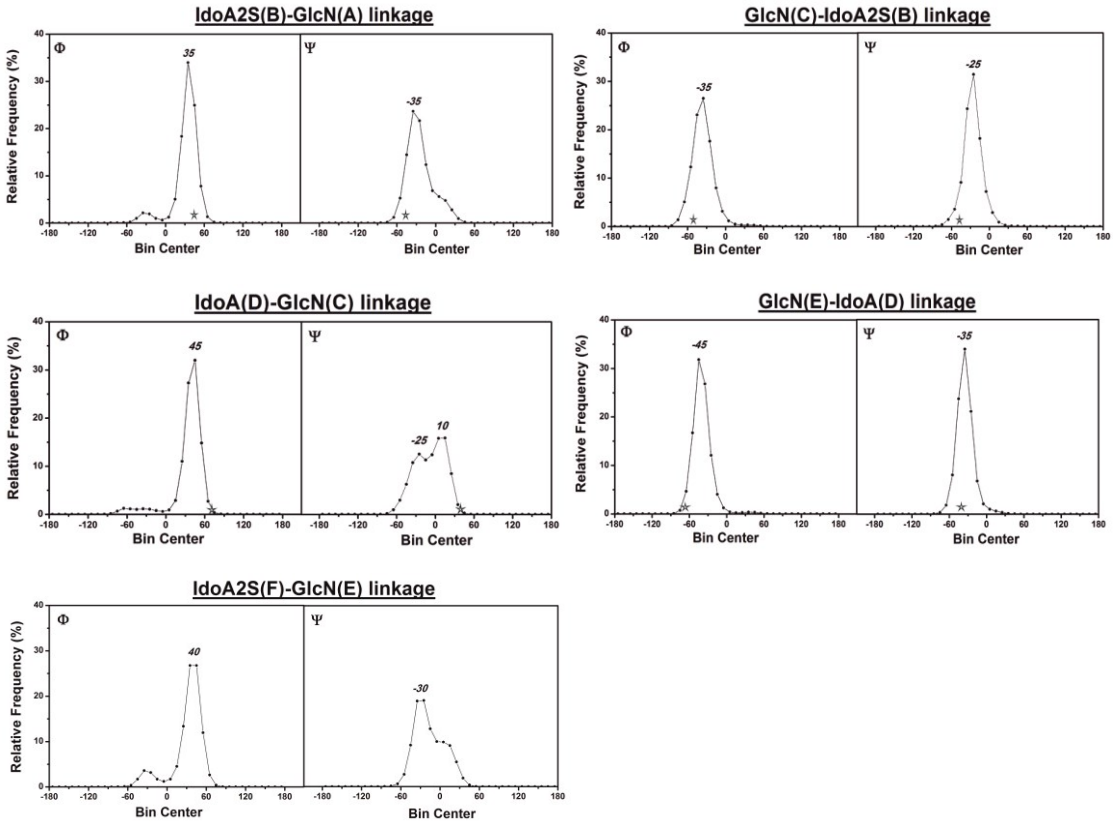

Figure 14. Frequency distribution curves of the Φ and Ψ interglycosidic torsions for each linkage of **Hexa3**, obtained from 500 ns MD simulation. The Φ-Ψ values corresponding to the conformation of **Hexa2** bound to FGF-1 (PDB code 2ERM) are indicated with a star symbol. A bin size of 10º has been used.

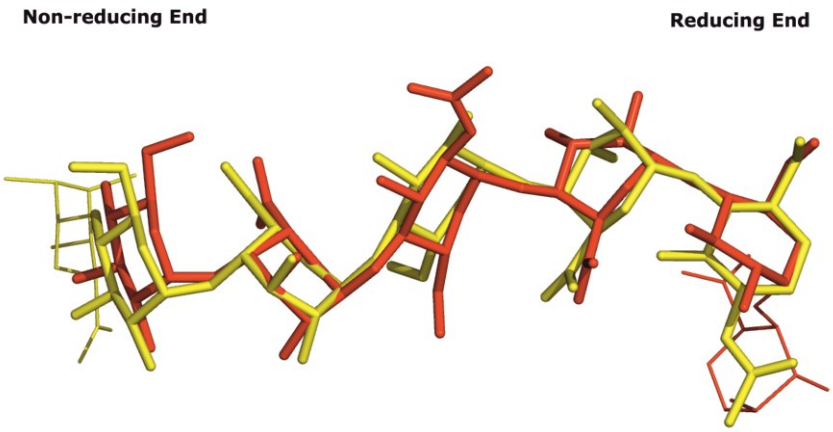

Figure 15. Backbone superposition of **Hexa2** (red) and **Hexa3** (yellow) compounds, representing the conformation in the bound state with FGF-1 (PDB code 2ERM) and the average conformation over 500 ns of MD trajectory, respectively. Note that both present very similar global conformations.

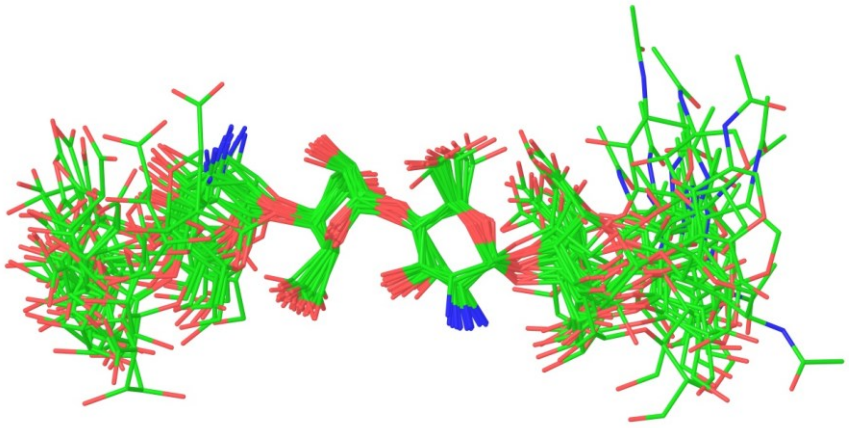

Figure 16. Superimposition of 20 conformers randomly taken from the MD simulation. The backbone of the two central residues (IdoA(D) and GlcN(C)) have been considered for the superimposition so that the increase of flexibility from the center to the ending residues of the polysaccharide chain is clearly observed. The hydrogen atoms and the sulfate groups have been removed for clarity.

Finally, regarding the orientation of the exociclyc torsion of the GlcN rings (ω), the MD simulation indicated a majority of *gg* conformers for the three glucosamines, with important contributions of *gt* (rings A and C) and *tg* (rings A and E; null for C) rotamers (see *Appendix*).

In conclusion, we have analysed the 3D structure of the hexasaccharide **3** (**Hexa3**), which is inactive as inductor of the FGF1 – FGFR signalling pathway[85d]. We have demonstrated that the structural characteristics of **Hexa3**, i.e., linear overall shape and backbone rigidity due to the geometry of the glycosidic linkages driven by the electrostatic repulsion between the charges of the sulphate groups[67, 85], rotational anisotropy, and iduronate conformational equilibrium, are very similar to those observed for previously studied heparin-like oligosaccharides, e.g. **Hexa1**[85a] and **Hexa2**[276]. In

particular, the overall conformation of **Hexa3** is very close to that of **Hexa2,** the latter presenting a remarkable mitogenic activity mediated by FGF-1 and sharing many of its structural characteristics, as it is the non-symmetrical distribution of the sulphate groups at both sides of the molecule[67].

Some discrepancies, although not too significant, have been found regarding the conformational equilibrium of the L-IdoA residues in **Hexa3** compared to the previously analysed compounds with the same length but different sulphation pattern and sequence, probably due to the different positions occupied in the oligosaccharide chain and the uncertainty introduced by the $^3J_{HH}$ mathematical fit. However this is not likely to be the cause of the inactivity as the differences in energy that can be inferred from differences in population are not large enough to justify the differences on binding observed. Furthermore, it has been reported that the conformational equilibrium remains even when the **Hexa2** is within the complex with FGF-1 as a result of the interactions with flexible side chains via electrostatic charges[88, 285].

Origin for the lack of mitogenic-induced capacity of Hexa3

To investigate the ability of the three **Hexa1-Hexa3** compounds to interact with FGF-1, IC50 values were determined from SPR competition experiments carried out at Prof. Lortat-Jacob´s Group (Grenoble, France). The results were consistent with **Hexa1** (heparin regular region) interacting with the highest affinity (IC_{50} = 83 nM), followed by **Hexa2** (IC_{50} = 460 nM) and its pseudo-palindromic sequence **Hexa3** (IC_{50} = 1600 nM). Interestingly, this affinity order is different from the previously reported data for the mitogenic activity, which pointed at **Hexa2** as the most active followed by the much more inactive **Hexa1** (heparin regular region), with **Hexa3** being almost inactive[67, 85d]. In addition, it has been reported that an octasaccharide is the minimum length for the sequence of heparin regular region to stimulate FGF-1- induced mitogenesis[67]. A potential reason for the discrepancies between the IC_{50} measurements is that while **Hexa1** may bridge two molecules of FGF-1 due to its axially symmetric sulphate distribution, **Hexa2** and **Hexa3** cannot (asymmetric sulphate distribution on one side of the molecule; see **figure 10**). Indeed, it has been reported that FGF-1 does not dimerize in the presence of **Hexa2[67]** and, additionally, that this compound maximizes the favourable interactions with the binding site of FGF-1 (**Hexa2** sulphates groups interact at both **a** and **b** sub-sites of FGF-1; see **figure 17**)[88].

Focusing on the different induced mitogenic activity and IC_{50} values obtained for **Hexa2** and **Hexa3**, it has to be noted that, if the directionality of the chain is disregarded, the distribution of sulphate groups in **Hexa2** and **Hexa3** can be considered to be pseudo-palindromic (if we disregard the sulphamate group at glucosamine E in the latter), and therefore analogue for both compounds. Since both the interaction of **Hexa3** with FGF-1 (IC_{50} value) and its capacity to induce mitogenic activity are much lower than in the case of **Hexa2** (also **Hexa3**), it seems reasonable to think that the only possible origin for its lack of activity should be the different geometry around the glycosidic linkages for both directionalities of the chain (non-reducing to reducing end and vice versa), as the distinct conformational behaviour of the GlcN-IdoA and IdoA-GlcN linkages suggests (**figure 14**). To prove

this hypothesis, the most representative conformation of **Hexa3** obtained from MD was manually superimposed (backbone) to the NMR-resolved structure of **Hexa2** bound to FGF-1 (PDB code 2ERM). All the different possible superimpositions were performed, starting with **Hexa3** trying to fit its five sulphate groups located at the same side of the helix in both sub-sites of FGF-1 by inverting the directionality of the hexasaccharide chain (see *Appendix*). However, the resulting structures of inverted directionality showed unavoidable steric clashes, indicating that **Hexa3** global conformation does not allow to maximize the interactions with FGF-1 (see *Appendix*), as **Hexa2** does. Also, we superimposed the heavy atoms of the GlcNS-IdoA2S-GlcNAc,6S triad, contained in both molecules, in the two possible directionalities for **Hexa3**, this is, from the non-reducing to the reducing end (shown in **figure 17**) and its reversed mode (see *Appendix*), i.e., exchanging the positions of the GlcNAc,6S and GlcNS residues within the triad. Interestingly, this can be done, in principle, because the distances between the sulphate groups are similar. However, the reverse mode (GlcNAc,6S-IdoA2S-GlcNS superimposition) did not fit in sub-site **a** (steric hindrance) and, even more, the other part of the chain fell away from the binding site (see *Appendix*). Differently, in the non-reducing to reducing end orientation the GlcNAc,6S-IdoA2S-GlcNS triads of both **Hexa2** and **Hexa3** compounds presented the same geometry for their glycosidic linkages (because of the same directionality of the linkages), thus allowing the 3 sulphate groups of **Hexa3** triad to be properly oriented to fully occupy sub-site **a**, while remaining the secondary sub-site (**b**) unoccupied. Therefore, the mode of interaction shown in **figure 17** for **Hexa3** complexed to FGF-1 is the only possible, thus confirming that this compound, whose chemical design aimed to simultaneously interact to FGFR and FGF-1 through its non-reducing and reducing terminal, respectively, was correctly devised. According to this model, the lack of binding site occupancy at sub-site **b** would attenuate the interaction between the GAG chain and FGF-1, explaining the lower IC_{50} value obtained for **hexasaccharide 3**.

On the other hand, it was demonstrated that FGF-1 interaction with **Hexa2** made more rigid the residues involved upon ligand binding (entropic cost) and more flexible those amino acids participating in the interactions with the FGF-1 receptor (entropy increase to compensate the entropic cost of ligand binding). Under these premises, we propose that the number of accessible FGF-1 conformations (or orientations of the side chains) able to interact to a FRFR, or number of "active microstates", would be higher for the FGF-1 in the bound state (FGF-**Hexa2**), so that the enthalpy-entropy balance for the FGF-FGFR interaction would be favoured and, consequently, the biologically relevant FGF-Hexa2-FGFR ternary complex stabilized (high induced mitogenic activity). Following the same reasoning, for the interaction (weaker) of **Hexa3** with just the sub-site **a** of FGF-1 (**figure 17**), we hypothesize that the number of accessible FGF-1 "active microstates" for FGFR recognition would be considerably lower, thus destabilizing FGF-FGFR interaction and, as a result, the formation of the FGF-Hexa3-FGFR ternary complex necessary for the signalling pathway to be triggered. We think this provide a reasonable hypothesis for the negligible induced mitogenic activity observed for **hexasaccharide 3**[85d].

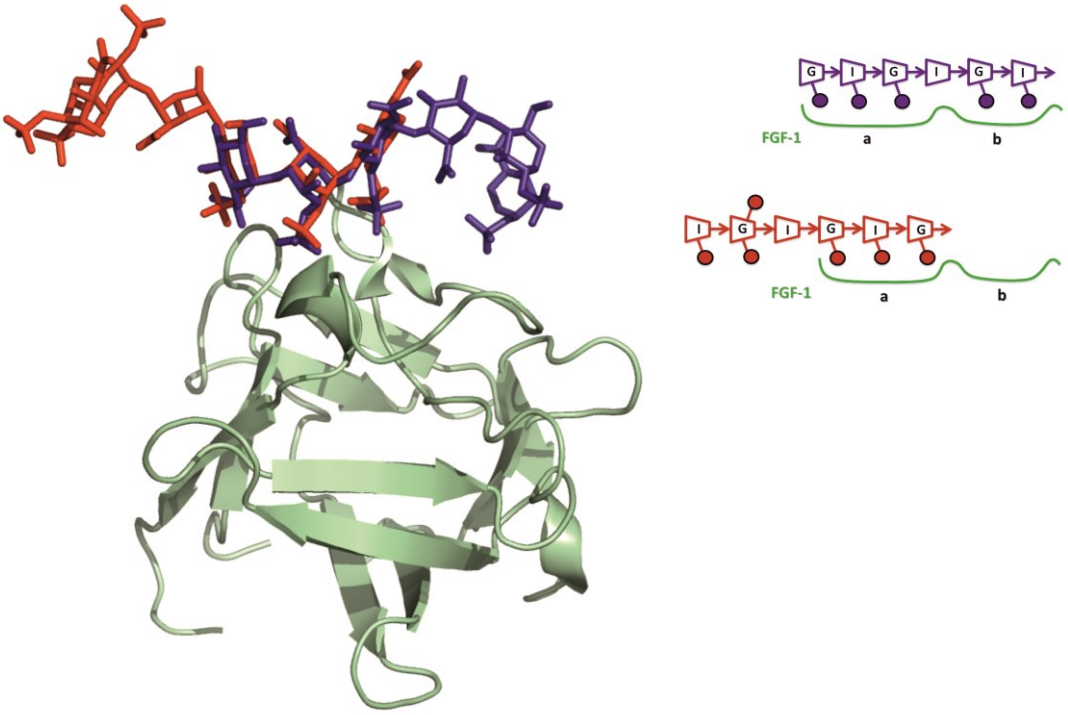

Figure 17. 3D (left) and schematic (right) representation of the complexes of FGF-1 (ribbons) with **Hexa2** (purple sticks; PDB code 2ERM) and **Hexa3** (red sticks), the latter manually superimposed (through the sulphate groups) to **Hexa2** at sub-site **a**. The two sub binding sites of FGF-1 (green) are labelled **a** and **b**.

3.3 Methodology

3.3.1 Nuclear Magnetic Resonance

Library of sulphated trisaccharides (subchapter 3.1)

NMR experiments were performed on Bruker DRX 500 MHz spectrometer equipped with 5 mm inverse triple-resonance probe. NMR samples were prepared at pH*≅7 in 500–600 or 200 mL in 5 mm or 3 mm tubes, at 2 mm and 6 mm, respectively, in 99.9% D2O and at several temperatures varying from 278 to 318 K. Sizes of acquisition matrices were 2 K_512 for COSY-dqf, gradient selected, experiments and 1 K_256 for TOCSY with mixing time of 80 ms. HSQC were recorded in gradient enhanced versions using echo-antiecho detection both with or without decoupling during acquisition. When it was required presaturation was applied by low power irradiation at water frequency.

The preliminary results of the NOESY were unsatisfactory because the molecules were close to the zero crossing point and give very weak peaks. ROESY sequences were also applied but the strong coupling of the protons of the L-IdoA2S residue biased the results. No better results were obtained using T-ROESY. Therefore, we recorded all the NOESY experiments at 278 K to increase the correlation time in order to obtain negative NOE peaks. Additionally, the use of lower temperatures allowed us to exploit the increase of the population of the lowest energy conformations. The build-up experiments were acquired with 1D sequence selected with gradients spin echo (dpfgse)[286] and two spin echoes flanked by bipolar gradients during the mixing time (200, 300, 400, 500, 600, 700, 800, 1000, 1300, 1500 ms).

An inactive hexasaccharide sequence for the FGF-1 mitogenic activity *(subchapter 3.2)*

A NMR sample was prepared dissolving 2.5 mg of **Hexa3** in 600 μL of D_2O 100%, adjusting the pH* to 7.0. 2D experiments were recorded using z-gradients pulses when possible: gs-DQF-COSY,[287] TOCSY[288] using 80ms of spin lock, NOESY,[286, 289] ROESY,[288] HSQC using gradient selected with sensitivity enhanced versions[290], and coupled HSQC.

In order to estimate the longitudinal and transversal cross relaxation rates, σ_{NOESY} and σ_{ROESY}, compensated *off-resonance* ROESY[172] experiments were acquired using different mixing times (100, 200, 300, 400 and 500 ms) and several radiofrequency offsets (6, 10, 20 and 30 kHz) for the spin-lock. The spin locking field was 7.14 kHz and the experiments were carried out at 298 K.

The off resonance pulse locks the spin along the effective field, which makes a ϕ angle with the z axis. This angle is defined as follows:

$$\phi = \arctan(\omega_1 / \omega_{offset}) \qquad \textit{Eq. 2}$$

The pure longitudinal and the pure transverse cross relaxation rates (σ_{NOESY} and σ_{ROESY}) were calculated for each proton pair from the dependence of the NOE cross peaks versus the ϕ angle, according to the expression:

$$\sigma_{obs} = \sigma_{NOESY} \cos^2 \phi + \sigma_{ROESY} sen^2 \phi \qquad Eq.\ 3$$

On the other hand, NOE and ROE cross relaxation rates are related to interproton distances and it can be expressed as[174a]:

$$\sigma_{NOESY} = r_{IS}^{-6}\left(6J(2\omega) - J(0)\right) \qquad Eq.\ 4$$

$$\sigma_{ROESY} = r_{IS}^{-6}\left(2J(0) + 3J(\omega)\right) \qquad Eq.\ 5$$

Transforming the cross relaxation rates in terms of interproton distances (r_{IS}) and effective correlation times for each proton pair (τ_{eff}) requires an assumption on the behavior of the spectral density function $J(n\omega)$, which involves that the motion of two interacting protons can be described by a single exponential. The spectral density can be then written as:

$$J(n\omega) = \left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2 \gamma^4}{10} \frac{\tau_{eff}}{1 + n^2 \omega^2 \tau_{eff}^2} \qquad Eq.6$$

where γ is the magnetogyric ratio of the nuclei, $\hbar$ is the reduced Planck constant, μ_o is the vacuum permeability and ω is the Larmor frequency (the later related with the magnetic field of the spectrometer).

For each proton pair, correlation times and, therefore, proton-proton distances can be calculated from the ratio $\kappa = \sigma_{NOESY} / \sigma_{ROESY}$ by solving the following equations[291]:

$$\tau_{eff} = \frac{1}{2\omega} \sqrt{\frac{1 - 22\kappa + 3\sqrt{9 + 4\kappa + 36\kappa^2}}{2 + 4\kappa}} \qquad Eq.\ 7$$

$$r_{IS} = \left(\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\hbar^2 \gamma^4}{10\sigma_{ROESY}} \left(2 + \frac{3}{1 + \omega^2 \tau_{eff}^2}\right)\right)^{1/6} \qquad Eq.\ 8$$

We have obtained internuclear distances and correlation times disregarding equal mobility between different proton pairs. Thus, no model of motion was assumed a priori. This method reduces the intrinsic error resulting from the use of an internal reference (e.g. in the Isolated Spin Pair Approximation[261]) and is appropriate for anisotropic molecules.

3.3.2 Modelling

Library of sulphated trisaccharides (subchapter 3.1)

A. Unrestrained MD (free-MD)

Input preparation

In all cases, the starting geometries were generated from the available data[65] deposited in the Protein Data Bank (PDB code 1HPN) and modified accordingly. The topologies were built with PREP-LINK-EDIT-PARM module of Amber 5.0, employing the residues and the set of partial charges published by Perez *et al*[236b] (the latter developed under the context of the set of parameters for carbohydrates PIM[292]) and the force fields *parm91*[266b] of Amber and *glycam_93*[268] together with the set of Altona parameters for sulphates[265]. Two independent starting geometries of each heparin-like trisaccharide structure were built, one with the IdoA2S residue in the chair $^1C_4$ conformation and one with the IdoA2S in the ^2S_O skew boat geometry. Each of these models was immersed in a 41Å-sided cube of pre-equilibrated TIP3P water molecules.

Molecular dynamics

MD simulations were run on the *Finis Terrae* cluster belonging to the *Centro de Supercomputación de Galicia* (CESGA), Spain, taking advantage of the prioritized computing time we were awarded (ICTS-2010-ID119).

To equilibrate the system we followed a protocol consisting of 10 steps. Firstly, only the water molecules were minimized, and then heated to 300 K. After, the water box together with the sodium ions were minimized and then followed by a short MD simulation (3 ps). At this point, the whole system is minimized by four consecutive steps imposing positional restraints on the solute, with a force constant decreasing step by step from 20 to 5 kcal/mol. Finally, an unrestrained minimization (100 steps) was carried out.

The production dynamics simulations were accomplished at a constant temperature of 300 K (by applying the Berendsen coupling algorithm[293] for the temperature scaling) and constant pressure (1 bar). The Particle Mesh Ewald Method[267, 282b] (to introduce long-range electrostatic effects) and periodic boundary conditions were also turned on. The SHAKE algorithm for hydrogen atoms, which allows using a 2 fs time step, was also employed. Finally, a 9 Å cutoff was applied for the Lennard-Jones interactions.

MD simulations have been performed with the sander module of Amber 6.0, with explicit treatment of the 10 12 hydrogen bond potential, in agreement with the parameters set for sulphates we use (Altona). A total of 16 MD simulations of 20 ns each were obtained. The trajectory coordinates were saved each 0.5 ps.

The data processing of the 16 generated trajectories were done with the *ptraj* module of Amber 9.0, except for the Cremer-Pople puckering coordinates, which were calculated with the Carnal module of Amber 5.0.

The final theoretical $^3J_{HH}$ values were obtained as averages for each of the models (L-IdoA in $^1C_4$ or in ^2S_O) according to the expression $^3J^{exp}_{m(m+1)} = f_{(^1C_4)} \cdot <^3 J^{MD}_{m(m+1)}(^1C_4)> + f_{(^2S_O)} \cdot <^3 J^{MD}_{m(m+1)}(^2S_O)$ (*Eq. 1)*. Thus, to obtain the populations of conformers ($^1C_4$ and ^2S_O) of the L-IdoA2S ring in each trisaccharide, we performed an iterative fitting of theoretical and experimental J-coupling data (*Eq. 1*; the $^4C_1$ conformation was disregarded as no experimental support was obtained for it, particularly the exclusive H5c-H5b NOE was not observed).

In this equation $f_{(^1C_4)}$ and $f_{(^2S_O)}$ are the molar fractions of each conformer, $<^3 J^{MD}_{m(m+1)}( )>$ are the averages from MD, and m is an index that runs from 1 to 4. Therefore, the experimental $^3J_{HH}$ coupling constants were considered as averages of the MD-derived $^3J_{HH}$ for each conformer weighted on the molar fraction of each one. As the theoretical values were averages from MD simulations, they implicitly reflected the fluctuations around canonical conformations, which must be considered for this flexible hexopyranose ring[271], particularly to account for the pseudorotational conformational space in the case of the skew boat conformer (^2S_O). In addition, the experimental measurements of $^3J_{HH}$ values at five different temperatures (5, 15, 25, 35 and 45 ^oC), allowed us to monitor the population of conformers as a function of temperature.

The overall correlation times (τ_o) for the 16 trisaccharide models (8+8 with the iduronate ring adopting the $^1C_4$ chair and ^2S_O skew-boat conformation, respectively) were calculated according to the *model-free approach* of Lipari and Szabo[275] from the auto-correlation functions for each proton-proton vector (between vicinal hydrogens), which were derived with the *ptraj* module of Amber 9.0. Since both the internal and overall motions act on the correlation function of each H-H vector, we first eliminated the translational and rotational components of the molecular tumbling at the global level to obtain the internal auto-correlation functions, for which only the internal motions contribute. This was done by RMS fit the backbone coordinates of each frame on the starting ones, prior to calculate the auto-correlation functions. Thus, the internal auto-correlation functions ($C^{LS}_{int}(t)$) were fit to the Lipari and Szabo expression

$$C^{LS}_{int}(t) = S^2 + (1 - S^2) \cdot e^{-t/\tau_{int}} \qquad \textit{Eq. 9}$$

This allowed us to obtain the parameters of order (S^2) and the internal correlation times (τ_{int}).

Since the correlation function describing the global motion (assuming an isotropic tumbling), $C_o(t)$, is given by the equation

$$C_0(t) = e^{-t/\tau_0} \qquad \textit{Eq. 10}$$

the correlation function (C(t)) can be broken down by its different contributions (as long as both the global and internal motions are not correlated), so that

$$C(t) = C_0(t) \cdot C_{int}(t) \qquad \textit{Eq. 11}$$

or

$$C(t) = S^2 \cdot e^{-t/\tau_0} + (1 - S^2) \cdot e^{-t/\tau_{int}} \cdot e^{-t/\tau_0} \qquad \textit{Eq. 12}$$

Therefore, the overall correlation times τ_o were determined by iterative fit to *Equation 12*, using the S^2 and τ_{int} values derived from *Equation 9*. Furthermore, since they corresponded to independent values for the $^1C_4$ and ^2S_O models, we calculated a representative τ_o value for each **Tri1-Tri8** compound by weighting on the populations obtained for both puckers at 278 K (**figure 8**).

B. Tar-MD

Input preparation

In all cases, the initial coordinates were based on the NMR-resolved dodecasaccharide structure of natural heparin[65] (PDB code 1HPN), following the protocol we have previously described[76]. The starting conformation of the L-IdoA2S unit of each trisaccharide was the $^1C_4$ chair. The topology and coordinates files of every system were built with the *tLEAP* module of AMBER 11[294] package. All the trisaccharides were neutralized with sodium ions and then immersed in a TIP3P[264] water box, giving rise to systems of about 4000 atoms. The Glycam06g-1 parameters[220] were used to model the sugar moiety, including the sulphate and sulphamate moieties. For the water molecules and sodium ions, the Amber99SB parameters[269] were employed. Furthermore, the partial charges of GLYCAM06[220] were employed for the sugar moiety, adjusting the partial charge on the O- and N- atoms bound to the SO_3^- groups according to GLYCAM philosophy for charge development. For the O-isopropyl group, partial charges were derived from the molecular electrostatics potential (MEP) using the RESP method[295] with a constraint of 0.01, for consistency with the procedure employed in GLYCAM06[220] force field development. The HF/6-31G* level of theory was used for both the structure optimization and the MEP calculation. Detailing the procedure employed, a methyl-O-isopropyl and a D-Glc-OMe were built and charge constraints imposed as follows: the total charge of both molecular models is set to 0, the methyl group in the D-Glc-OMe must have a charge of +0.194 whereas the charge of the oxygen involved in the glycosidic linkage is set to -0.194, and both methyl groups are set to be equivalent and to be removed during the last step of charge derivation, and, thus, being both compounds merged to form a D-Glc-O-Isopropil. Additionally, the partial charges on aliphatic hydrogens and on the O-isopropyl group were constrained to 0 and -0.194, respectively, in agreement with GLYCAM philosophy. The standard error obtained was 0.005. It is noticeable that a similar protocol has been successfully applied for the development of parameters for different sugar derived compounds[296]. The quantum mechanical calculations and the RESP[295] procedure were carried out with ante-R.E.D 2.0 and R.E.D IV of the R.E.D web server[297].

Molecular dynamics

MD simulations run on the *Finis Terrae* cluster belonging to the *Centro de Supercomputación de Galicia* (CESGA), Spain, taking advantage of the prioritized computing time we were awarded (ICTS-2011-ID162).

MD simulations were carried out with AMBER 11[294]. Prior to include the constraints, we performed an equilibration protocol consisting of an initial minimization of the water box (20000 steps), followed by a minimization of the whole system (10000 steps). Then, the TIP3P[264] water box was heated at constant volume until 278 K using a time constant for the heat bath coupling of 1 ps. The equilibration finished with 200 picoseconds of molecular dynamics simulation without restraints, at constant pressure (1bar) and turning on the Langevin temperature scaling with a collision frequency of 1 ps. Furthermore, non-bonded interactions were cut off at 8.0 Å and updated every 25 steps. *Periodic Boundary Conditions* and the P*article Mesh Ewald method*[282] were turned on in every step of the equilibration protocol to evaluate the long-range electrostatic forces (the grid spacing was approximately 1 Å).

The time-averaged restraints molecular dynamics were run with the same settings used in the last step of the equilibration protocol. A sole NOE-derived distance, that between H2 and H5 protons of the IdoA2S residue (table 2), was imposed as time-averaged constraint, applying a r^{-6} averaging. The equilibrium distance range was set to r_{exp}-0.1Å ≤ r_{exp} ≤ r_{exp}+0.1Å. Trajectories were run at 278 K, with a decay constant of 800 ps and a time step of 1 fs. The force constants k_2 and k_3 used in each case go from 25 to 45 kcal·mol^{-1}·A^{-2} (Supporting Information, table S5). The overall simulation length for the simulations was 8 ns. The coordinates were saved each picosecond, thus, obtaining MD trajectories of 8000 frames each. Convergence within the equilibrium distance range was obtained in all cases. The analysis of the *tar*-MD trajectories has been carried out with the *ptraj* module of AMBER 11[294], except for the Cremer-Pople coordinates, which were determined with an in-house script (see acknowledgements). The auto-correlation functions ($C_{int}(t)$) shown (**Figure 9**) were obtained following the procedure described in the previous section (A). In this regard, it has to be noted that the $C_{int}(t)$ functions have been only obtained for the **Tri1:Tri5** and **Tri2**:**Tri6** pairs since the others (**Tri3**:**Tri7** and **Tri4**:**Tri8**) showed too fast conformational transitions as to obtain the decay of the internal auto-correlation functions with the L-IdoA2S ring keeping the $^1C_4$ conformation.

An inactive hexasaccharide sequence for the FGF-1 mitogenic activity *(subchapter 3.2)*

The initial coordinates were based on the NMR-resolved dodecasaccharide structure of natural heparin (PDB code 1HPN). The topology and coordinates files were built with the *tLEAP* module of AMBER 11[294] package. The system was neutralized with sodium ions and then immersed in a TIP3P[264] water box, giving rise to a molecular system of 4526 atoms. The GLYCAM_06h[298]

parameters were used to model the sugar moiety, including the sulphate and sulfamate groups, and the AMBER ff12SB parameters[269] for the water molecules and calcium ions. The partial charges of GLYCAM06 were employed for the sugar moiety, adjusting the partial charge on the O- and N- atoms bound to the SO_3 groups according to GLYCAM philosophy for charge development. For the O-isopropyl group, partial charges were derived from the molecular electrostatics potential (MEP) using the RESP method[295] with a constraint of 0.01, for consistency with the procedure employed in GLYCAM06 force field[220] development. The HF/6-31G* level of theory was used for both the structure optimization and the MEP calculation. Detailing the procedure employed, a methyl-O-isopropyl and a D-Glc-OMe were built and charge constraints imposed as follows: the total charge of both molecular models is set to 0, the methyl group in the D-Glc-OMe must have a charge of +0.194 whereas the charge of the oxygen involved in the glycosidic linkage is set to -0.194, and both methyl groups are set to be equivalent and to be removed during the last step of charge derivation, and, thus, being both compounds merged to form a D-Glc-O-Isopropyl. Additionally, the partial charges on aliphatic hydrogens and on the O-isopropyl group were constrained to 0 and -0.194, respectively, in agreement with GLYCAM philosophy. The standard error and relative root mean square error were, respectively, 0.005 and 0.188. It is noticeable that a similar protocol has been successfully applied for the development of parameters for different sugar derived compounds[296, 299]. The quantum mechanical calculations and the RESP procedure were carried out with ante-R.E.D 2.0 and R.E.D IV of the R.E.D web server[297].

The molecular dynamics simulations have been run on a 4-node AMD Opteron Interlagos cluster (2,3 GHz, 16 cores per node). MD simulations were carried out with AMBER 12[300]. The equilibration protocol consisted of an initial minimization of the water box (20000 steps), followed by a minimization of the whole system (10000 steps); finally, the system was heated (40000 steps) at constant volume until 300 K using a time constant for the heat bath coupling of 1 ps. The production dynamics have been carried out at a constant temperature of 300 K, by applying the Langevin thermostat[301] with a collision frequency of 5 ps^{-1}, and at constant pressure (1bar), applying Periodic Boundary Conditions (PBC) and using the Particle Mesh Ewald Method[282] (PME) to account for the long range electrostatic effect (the grid spacing was approximately 1 Å). The SHAKE algorithm[302] was also employed, thus, allowing a 2 fs time step, and non-bonded interactions were cutoff at 8.0 Å and updated every 25 steps. The equilibration protocol and production dynamics have been performed with the *sander.MPI* and *pmemd.MPI* modules of AMBER 12[300], respectively. One MD simulation of 500 ns have been performed, saving the trajectory coordinates each picosecond. The analysis of the MD trajectory has been done with the *ptraj* module of AMBER 12[300]. The Cremer-Pople coordinates were calculated with a script from R.J. Wood´s Group (see acknowledgements).

Chapter 4

Structural features underlying Lg ECD interactions with GAGs

4.1 Calcium-dependent interactions

4.1.1 Sulphated GAGs: heparin-like trisaccharides

Results and Discussion

Nuclear Magnetic Resonance

Binding of Lg ECD with heparin-like trisaccharides is calcium dependent

The binding features of the GAG-like **Tri1**-**Tri8** trisaccharides, introduced in *Chapter 3*, to Lg ECD were investigated by STD NMR spectroscopy. In addition, we also wanted to assess the role of the divalent cation calcium on binding.

In STD NMR experiments, ^{1}H saturation is selectively created on the large protein that in turn can be transferred to a given small molecule in the solution if (i) there is binding of the molecule to the protein, and (ii) the binding equilibrium occurs within a regime of fast exchange in the relaxation time scale. In the large majority of cases the latter means that the binding must be of medium-to-weak affinity (K_D in the range of tens of μM to tens of mM). In the case of strong affinity, the ligand fully relaxes in the bound state before getting out of the protein binding pocket, so that there is no accumulation of saturation in solution, and no STD signal is observable.

Langerin ECD has been very recently shown to strongly bind to heparin of 6 kDa molecular weight in a Ca^{2+}-independent manner, which sharply contrasts with its binding to neutral carbohydrates, like mannose or N-acetylglucosamine[114]. In that work, an EC_{50} of 150 nM in the absence of Ca^{2+} was determined, whereas in the presence of the divalent cation a non-saturable behavior was observed. Under these premises, we first tested whether the binding of the trisaccharides would be of appropriate kinetics as to be observed by STD NMR experiments.

Thus, in the presence of Ca^{2+}, all the ligands showed clear STD signals in the spectra on samples containing Lg ECD (**figure 1**, and *Appendix*). From this result, we can deduce that in this case binding must be weaker than for the 6 kDa heparin, as for nM binders no STD signals would be expected. In addition, and surprisingly, the addition of an excess (10 mM) of EDTA, to sequester the Ca^{2+} ions, did abolish the binding of the trisaccharides **Tri1**-**Tri8** to Lg ECD, as no STD signals were then observed in the spectra (**figure 1**).

These results differ from the previous observation of Ca^{2+}-independent binding of longer heparin polysaccharides (6 and 15 kDa) to Lg ECD[114]. In that work, heparin was shown to bind Langerin in an extended binding site only present in the trimeric form of the protein, and different from the Ca^{2+}-binding site 2 of C-type lectins[115] (typical specific site for the recognition of neutral carbohydrates). However, in the present work, the molecular recognition of the trisaccharides

Tri1-Tri8 by Lg ECD has been demonstrated to require the presence of the divalent cation, suggesting a different site of binding in which the calcium ion must play an essential role.

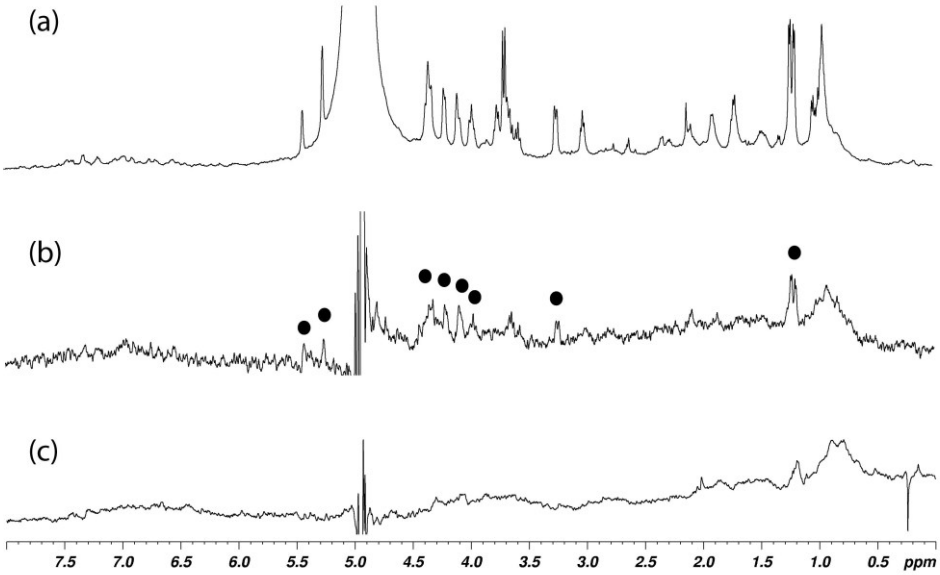

Figure 1. The interaction of GAG-like trisaccharides with Lg ECD is Ca^{2+}-dependent. (a) 1D ^1H NMR reference (*off-resonance*) spectrum of **Tri1** in the presence of Lg ECD. (b) 1D STD NMR spectrum of **Tri1** in the presence of Lg ECD in 4 mM Ca^{2+} Tris-d_{11} buffer. (c) 1D STD NMR spectrum of **Tri1** in the presence of Lg ECD in 10 mM EDTA Tris-d11 buffer. Black circles mark ligand STD signals. In (c) the STD spectrum shows only the proton signal envelope of the large protein. Note: the four NMR spectra were registered at 15 ºC and 500 MHz.

Binding of Lg ECD with heparin-like trisaccharides is independent of their sulfation-pattern

To go in depth into the mode of binding of each trisaccharide, **Tri1-Tri8**, to Lg ECD, series of STD NMR experiments were carried out at very high field (800 MHz) varying the saturation time to record the growth of STD signals. **Figure 2** shows the STD NMR experiment at 2 second saturation time (**figure 2a**), as well as the build-up curves (**figure 2c**), for the three sugar residues of **Tri1** (for **Tri2-Tri8** see *Appendix*). The data show that the residue receiving the largest amount of saturation is the glucosamine ring at the non-reducing end of the ligand (residue C), with a lower and rather homogeneous saturation being transferred to the remaining two other sugar rings (residues B and A). From these STD build-up curves, we mapped out the main contacts of the ligands with Langerin in the bound state (binding epitope), by determining the initial growth rates of the curves and normalizing all the values within a given ligand by the highest one, to which arbitrarily a value of 100% was assigned. The use of initial slopes increases the accuracy of the method by avoiding the detrimental effects of different relaxation properties of the ligand protons on the determination of the binding epitope. **Tri1** binding epitope is shown in **figure 2b** (see *Appendix* for the remaining ligands).

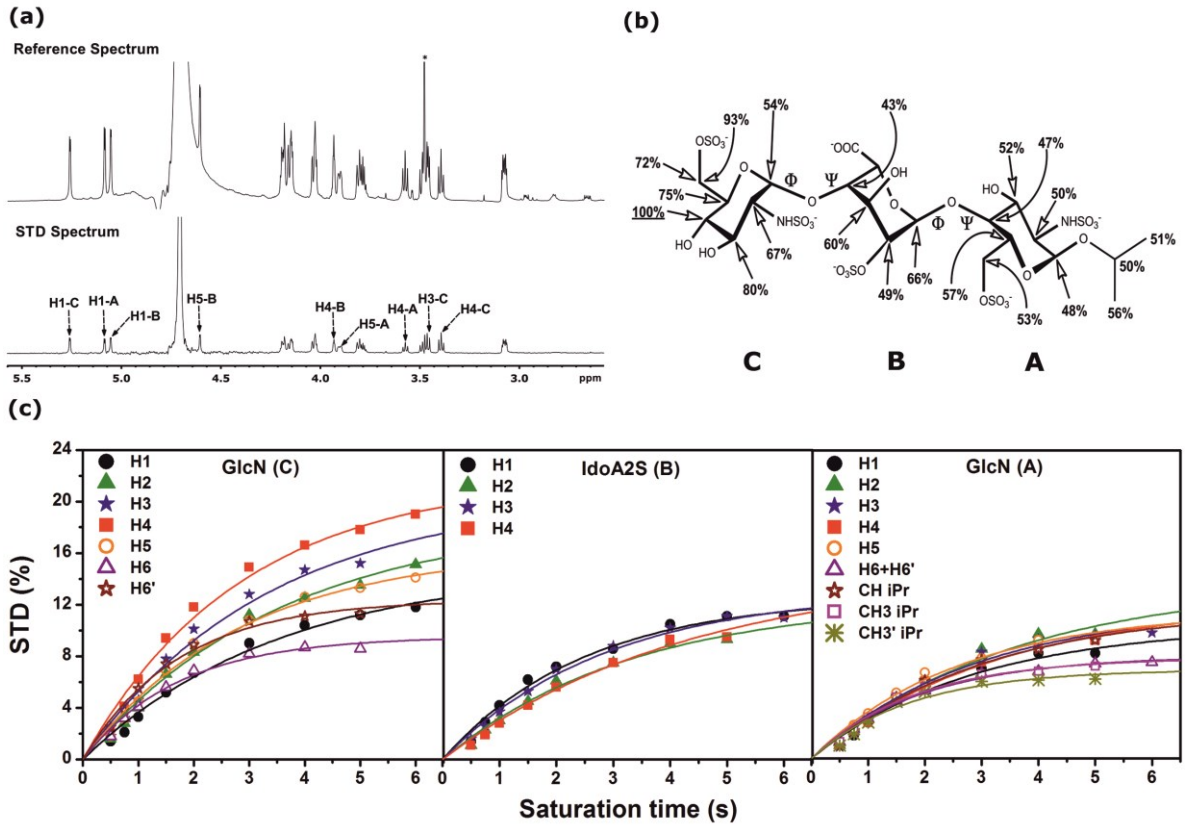

Figure 2. (a) Reference (top) and STD (bottom) spectra of the interaction of **Tri1** with Lg ECD (800 MHz, 15 ºC, 2 s saturation time). (b) Binding epitope of **Tri1** from STD NMR experiments. The numbers denote the fraction of saturation (in %) received from the protein by those ligand protons, relative to the maximum ligand STD signal (H4 of non-reducing GlcN ring; 100 %). Values close to 100 % represent intimate contacts to the protein surface in the bound state. (c) STD build-up curves of ligand **Tri1** in the presence of Lg ECD. The curves are divided by residue. From these curves the STD initial growth rates are obtained, from which the binding epitope of the ligand is determined (see Figure 3). Similar data were obtained for the remaining ligands **Tri2**-**Tri8** (*Appendix*).

Remarkably, all the ligands showed very similar patterns of relative STD distribution (**figure 2** and *Appendix*) within the experimental error. This means that the eight trisaccharides bind Lg ECD with equivalent binding modes, so the ligand sulfation pattern does not play a role on the binding geometry. The STD NMR experiments are hence supporting that the variable sulphate groups along the trisaccharides series (N- and 6-O-sulfates of residue A, and 6-O-sulfate of residue C) are not fundamental for the interaction. Furthermore, the data indicate that in the binding mode shared by the eight ligands, the protons that establish closest contacts with the protein surface are H3, H4, H5, and H6 of the non-reducing glucosamine ring (GlcN C). In all the cases, H4 of this residue showed the largest saturation transfer (100 % relative STD) among the ligand protons, highlighting the importance of this part of the molecule for the molecular recognition by Lg ECD (Figure 2). Thus, the analysis of the binding epitopes, along with the observation of the abolishment of binding to Lg ECD upon removal of the divalent cation, strongly supports that all the trisaccharides bind Langerin at the Ca^{2+}-dependent binding pocket, by a classical coordination of the calcium ion through the

di-equatorial oxygen atoms O3 – O4 of the non-reducing GlcN ring (strong STD intensities for protons H3 and H4; see **figure 2** and *Appendix*).

Lg ECD does no perform a conformational selection upon binding of heparin-like trisaccharides

We also carried out transferred NOESY experiments on the same samples used for the STD NMR experiments, containing the trisaccharide ligands, Lg ECD and Tris d_{11} buffer with Ca^{2+}. Since these samples were prepared with an appropriate ligand excess over protein (1:14 protein-to-ligand ratio) as to be at good conditions for the observation of transferred-NOEs, the observed averaged NOEs come mainly from the bound state, allowing the determination of the bioactive conformation or global geometry of the ligand in the complex. The results were then compared to NOESY experiments on the free ligands, in the absence of protein, to reveal possible conformational changes during the molecular recognition of the ligands by the large receptor. **Figure 3** shows expansions of the NOESY and transferred-NOESY spectra of **Tri1** (see also *Appendix*).

For heparin-like glycosaminoglycans, whereas the global conformation is determined by the rearrangements around the interglycosidic torsions (ϕ and ψ of the GlcN-IdoA2S and IdoA2S-GlcN linkages, see **figure 2b**), the local geometry is featured by the unique conformational equilibrium of the iduronate ring. This sugar residue typically shows a polyconformational behavior of its hexopyranose ring, being usually in chemical equilibrium between the chair conformations $^1C_4$ and $^4C_1$, and the skewed-boat ^2S_O. For the particular case of an internal L-IdoA2S residue, this equilibrium has been reported to only be participated by the $^1C_4$ and ^2S_O puckers[75].

For the library of heparin-like trisaccharides (in the free state), as we have reported in *Chapter 3*, the presence of the ^2S_O conformer in solution was confirmed by the observation of the exclusive NOE between protons H2 and H5 of the iduronate ring (**figure 3**, top; also see *Appendix*), which are far beyond the NOE distance in any of the chair conformations. In addition, the non-observation of the exclusive H5c-H5b NOE (GlcN-IdoA2S linkage) indicated the absence (or negligible presence) of the $^4C_1$ pucker in solution. Furthermore, all the trisaccharides showed an equilibrium between the ^2S_O and $^1C_4$ conformers, with a larger population of the skewed-boat present on the series **Tri1**-**Tri4** in comparison to **Tri5**-**Tri8**, due to the effect of the 6-O-sulfate group on the reducing end GlcN residue, only present in the **Tri1-Tri4** series.

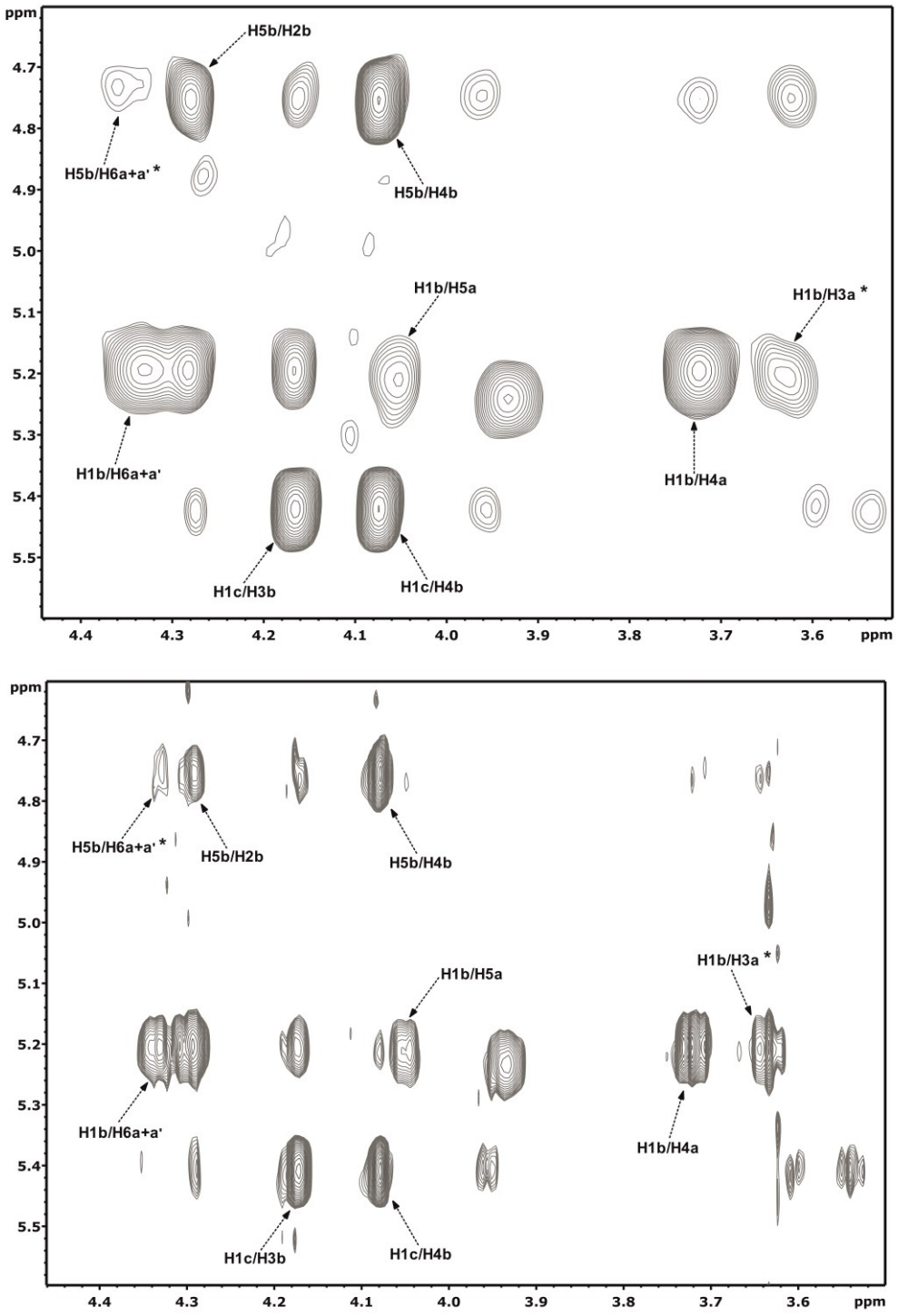

Figure 3. Expansion of the NOESY spectrum, at 600 ms mixing time, of ligand **Tri1** in the free state (top), and equivalent zoom of the transferred-NOESY spectrum, at 200 ms mixing time, of the same ligand in the presence of Lg ECD (bottom, bound state). Note that the NOE signals corresponding to a *syn* rearrangement of the IdoA(B)-GlcN(A) linkage (H1b/H4a and H1b/H6a+a') are, in both spectra, much more intense than those of the *anti* conformation (marked with a star symbol).

On the other hand, in the presence of Lg ECD, all the ligands showed the characteristic H2b/H5b NOE as well (**figure 3**, bottom; also see *Appendix*), indicating that the skewed boat $^{2}S_{O}$ conformation of the central iduronate ring is also present in the bound state of the ligands in complex with Langerin.

Regarding the global conformation of **Tri1-Tri8**, the interglycosidic NOEs H1b/H3a and H1b/H4a for the GlcN(C)-IdoA(B) linkage, and H1b/H4a and H1b/H6a+a' for the IdoA(B)-GlcN(A) linkage, appeared intense in both, NOESY and transferred-NOESY experiments. In qualitative terms, this result substantiates that the *syn*-ψ –like conformation (ψ torsions centered at the ±60º minima) is the major one for both glycosidic linkages, and that this preference does not change appreciably upon binding to Lg ECD. In any case, two weak-to-medium intensity NOEs related to the *anti*-ψ conformation around the IdoA(B)-GlcN(A) linkage have also been observed for all the trisaccharides in the bound state to Lg ECD, alike what was observed in the free state (*Chapter 3*). We are referring to the H1b-H3a and H1b-H5a NOE peaks (**figure 3**). Note that, whereas the H1b-H5a NOEs is not strictly exclusive of the *anti*-ψ conformation (the *syn*-ψ conformer with ψ = 60º would also give rise to a medium-intensity H1b-H5a NOE peak), it is the H1b-H3a NOE, and when both NOE signals appear together the existence of the *anti*-ψ conformer is specially trustworthy. On the other hand, it is fair to mention that the H5b-H6a NOE is also exclusive of the *anti*-ψ conformation. However, the growth of this NOE peak might be highly affected by the lost of magnetization due to a very short longitudinal relaxation time (T1), as a consequence of the efficient (in terms of T1) relative reorientation of the H5b and H6a+a' coupled protons (free rotation of methylene protons and IdoA2S conformational plasticity acting in the nanosecond time scale as relaxation mechanisms). Thus, although the *anti*-ψ conformation is present, the H5b-H6a+a' NOE intensity may not be observed due to fast relaxation. For the compounds **Tri1-Tri8** analyzed herein, in the bound state to Lg ECD, the H5b-H6a+a' NOE has been only clearly identified (with a very low intensity) in some cases, e.g., **Tri1** (**figure 3**, bottom), as previously observed in its free state[76] (**figure 3**, top).

Also, we have carried out a semi-quantitative analysis of the NOESY and transferred-NOESY spectra, with comparative purposes, to clarify any possible difference in the conformations of the ligands upon binding to the protein. Thus, regarding the *anti*-ψ conformation around the IdoA(B)-GlcN(A) linkage, we calculated the ratio between the cross relaxation rates (σ_{NOE}) of the *anti*-ψ exclusive NOE H1b-H3a and the reference NOE distance H1-H2 of the non-reducing end glucosamine (H1c-H2c, which let us to normalize the values). Thus, changes in the H1b-H3a/H1c-H2c ratio can be roughly "translated" into differences in the populations of the *anti*-ψ conformers. The calculated ratios (*Appendix*) are, in general terms, quite similar, in both the bound and free states, although in some cases the *anti*-ψ conformers seem to be "slightly promoted" upon biding (higher H1b-H3a/H1c-H2c ratios). Anyway, considering the experimental error, the differences observed are not big enough as to reliably correlate them to a possible conformational selection.

On the other hand, to semi-quantitatively compare between the conformation of the central iduronate ring in the bound and free states, we calculated the ratio between the cross relaxation rates (σ_{NOE}) of

H2b-H5b and H4b-H5b contacts (L-IdoA2S). It should be noted that the latter is an internal distance of the iduronate ring that does not change between the $^1C_4$ and ^2S_O conformations, and helps to normalize the values. Therefore, changes in the populations of the skewed-boat pucker upon binding would be observed if differences in the H2b-H5b/H4b-H5b ratio were detected. Thus, according to the calculated ratios (**table 1**), a different conformational behavior of L-IdoA2S ring between the trisaccharides series **Tri1**-**Tri4** and **Tri5**-**Tri8** was observed (fairly similar ratios between the free and bound states). Even more, also in the bound state there were larger populations of the ^2S_O conformer for the **Tri1**-**Tri4** series, as previously observed in the free state[76], supporting that the protein indeed does not show any preference for one of the iduronate conformations over the other. It is interesting to highlight that, although the central iduronate ring makes significant contacts with the protein in the bound state (STD signals, see **figure 2**), the tr-NOESY data indicate that Lg ECD can recognize the ligands in any of the two conformations $^1C_4$ or ^2S_O (**figure 3** and **table 1**). This has been previously observed in a larger heparin-like GAG hexasaccharide bound to the acidic fibroblast growth factor FGF-1, and it was proposed to be a mode for alleviating the entropic penalty associated with the restriction on the degrees of freedom of the side chains caused by binding[254]. Like in that case, Lg ECD has flexible residues involved in binding, as its Ca^{2+} binding pocket is flanked by two lysine residues, which can keep flexible enough in the bound state as to accommodate the sugar ring in two different conformations $^1C_4$ and ^2S_O.

Table 1. Iduronate H2-H5 σ_{NOE} (s^{-1})[a] in the free and bound states, together with their ratio with respect to the iduronate H4-H5 σ_{NOE}.

	Free	Bound	Ratio Free	Ratio Bound
Tri1	5.27	7.45	0.32	0.26
Tri2	5.29	13.36	0.33	0.36
Tri3	4.32	13.70	0.30	0.31
Tri4	5.77	16.49	0.33	0.28
Tri5	2.64	5.80	0.17	0.16
Tri6	3.03	4.28	0.20	0.16
Tri7	1.98	5.62	0.17	0.14
Tri8	1.42	6.40	0.15	0.11

[a] Cross relaxation rates (σ_{NOE}) were approximated by the ratio of the normalized NOE volume and the mixing time (see Methodology).

Table 2. Interglycosidic ^{1}H-^{1}H distances defining the global conformation of the trisaccharides in the bound state.

	Proton pair	Docking Distances (Å)[a]	Experimental Distances	
			***Bound*[b]**	***Unbound*[c]**
Tri1	H1c-H3b	2.6	2.5	2.6
	H1c-H4b	2.5	2.5	2.7
	H1b-H4a	2.6	2.5	2.5
	H1b-H6a	2.7	2.7	2.7
Tri2	H1c-H3b	2.6	2.5	2.6
	H1c-H4b	2.5	2.5	2.6
	H1b-H4a	2.7	2.5	2.4
	H1b-H6a	2.6	2.7	2.6
Tri3	H1c-H3b	2.6	2.5	2.6
	H1c-H4b	2.6	2.5	2.7
	H1b-H4a	2.6	2.6	2.6
	H1b-H6a	3.0	2.7	2.7
Tri4	H1c-H3b	2.5	2.5	2.7
	H1c-H4b	2.5	2.5	2.7
	H1b-H4a	2.8	2.6	2.5
	H1b-H6a	2.3	2.7	2.7
Tri5	H1c-H3b	2.3	2.5	2.6
	H1c-H4b	2.6	2.4	2.6
	H1b-H4a	2.8	2.4	2.6
	H1b-H6a	2.4	2.5	2.5
Tri6	H1c-H3b	2.3	2.5	2.6
	H1c-H4b	2.6	2.5	2.6
	H1b-H4a	2.7	2.5	2.6
	H1b-H6a	2.3	2.7	2.6
Tri7	H1c-H3b	2.5	2.5	2.6
	H1c-H4b	2.5	2.4	2.6
	H1b-H4a	2.8	2.4	2.5
	H1b-H6a	2.7	2.5	2.6
Tri8	H1c-H3b	2.5	2.5	2.6
	H1c-H4b	2.5	2.5	2.6
	H1b-H4a	2.6	2.5	2.6
	H1b-H6a	2.8	-	-

[a] Boltzmann r^{-6} averages ($\langle r^{-6}\rangle^{-1/6}$) over the best 20 ranked poses (see *Modelling*) and r^{-6} weighted on the populations of $^{1}C_4$ and $^{2}S_O$ puckers at 15 ºC previously reported[76]. H1b-H6a distances represent the r^{-6} average over the H1b-H6a*proR* and H1b-H6a*proS* values [b] Determined from tr-NOESY experiments, at 15 ºC, considering the Isolated Spin Pair Approximation[261] (see Methodology). [c] NOESY-derived values, at 5 ºC (see table 1 in *Chapter 3*,).

Finally, we experimentally determined key ^{1}H-^{1}H distances that define the conformations around the interglycosidic linkages of the trisaccharides in the bound state, from the transferred-NOESY experiments (**table 2)**. The corresponding ^{1}H-^{1}H distances in the free state were experimentally determined in a previous work[76], which allowed us to quantitatively compare the set of conformation-defining distances between the bound and the free states (**table 2**; the distances obtained from docking calculations are also shown for comparison; see Modelling section). Thus, the high agreement (within 0.1-0.2 Å tolerance) between them (**table 2**) corroborated that the conformation around the interglycosidic linkages, and hence the global geometry of the eight trisaccharides, does not change as a consequence of the binding to Lg ECD.

Modelling

Docking

In order to obtain 3D molecular models for the heparin-like trisaccharides complexed to Lg ECD, we carried out flexible docking (SP) calculations using the module *glide* within the *Schrödinger* suite of programs. The crystal structure of the complex between Langerin carbohydrate recognition domain (CRD) and laminaritriose (glucose trisaccharide of β(1-4) linkages) was used as a starting point (see Methodology).

Prior to introduce the results, it is important to set in a context the general idea of docking calculations (see also *Chapter 1*). Thus, it is important to remember that in standard docking studies, ligands are docked into the binding site of a rigid receptor while the ligand is left free to move. However, we have performed *Induced Fit Docking* (IFD), which takes into account possible movements of amino acid side chains to better adapt to the shape of the ligand (see Methodology), prior to standard flexible docking, as an "*ad-hoc"* refinement step of the receptor conformation. We have chosen IFD methodology for Langerin as it should, in principle, generate more accurate structures of its complexes with the heparin-like GAG trisaccharides, because 1) it does not exist any published 3D structure of Langerin bound to such highly charged ligands in the Ca^{2+} binding site, and 2) these ligands may exert a stronger influence in the conformations of the protein side chains involved in binding than the already studied neutral carbohydrate ligands[148b]. Based on these grounds, we have carried out IFD for the most sulphated trisaccharide (**Tri1**) with the carbohydrate recognition domain (CRD) of Langerin. Moreover, considering the known influence of the input ligand geometry on docking accuracy[246a], different conformers for each ligand have been generated by Monte Carlo Multiple Minimum (MCMM) torsional sampling. Thus, the 10 most representative clustered conformations obtained for **Tri1** were flexibly docked (SP, *glide*) to the receptor conformation (among the different IFD-generated complexes) which better optimized the non-bonding interactions with this ligand (see Methodology).

Happily, the combinatorial approach consisting of sourcing of a set of different input conformations for each ligand together with the receptor structure refinement (IFD), prior to "productive docking", resulted in the generation of a huge ensemble of docking solutions (recovery of false negatives)

qualitatively in agreement with the experimental data. The post-docking clustering of conformers (see Methodology) let us to group the most representative poses for each trisaccharide ligand, not observing among them unrealistic or distorted conformations either at the local (rings puckering) or global (glycosidic linkages) levels (compared to previous structural studies of heparin-like derivatives[76, 78]; see relaxed maps in *Appendix*).

The results show two different binding modes, common for the eight trisaccharide models. From now on, we will be referring them as *main binding mode* (MBM) and *secondary binding mode* (SBM). The MBM is the most populated one (**figure 4**) and contains the lowest energy poses in all cases. In this binding mode every proton of every residue of the ligand (including the isopropyl group) establishes contacts with the protein surface (none of the residues are solvent exposed), with the non-reducing end GlcN residue interacting more closely (**figure 5**). Besides, the superimposition of the best docked pose (e*model*) of **Tri1-Tri8** trisaccharides, both in $^1C_4$ and ^2S_O conformations, revealed that the whole library bind Lg ECD in a very similar binding mode (figure 6). The calculations thus corroborate, as observed in the NMR experiments, that the different pattern of sulphation of these ligands does not influence their mode of binding to Lg ECD. Interestingly, and supporting the reliability of the docking results, the MBM resembles the geometry of binding of Man-α-(1-2)-Man and Gal-6-O-SO_3-bound to Langerin (crystal structures 3P5F and 3P5I, respectively). Thus, while comparing to the former case the non-reducing end GlcN residues of the trisaccharides present an orientation very similar to the mannose residue (slightly re-orientated around the calcium ion, to allow the carboxylate group of the IdoA2S ring to be placed between K313 and K299 residues), when the Gal-6-O-SO_3 ligand is considered the IdoA2S COO^- group occupies the same place in the binding site as the galactose sulfate group (see *Appendix*).

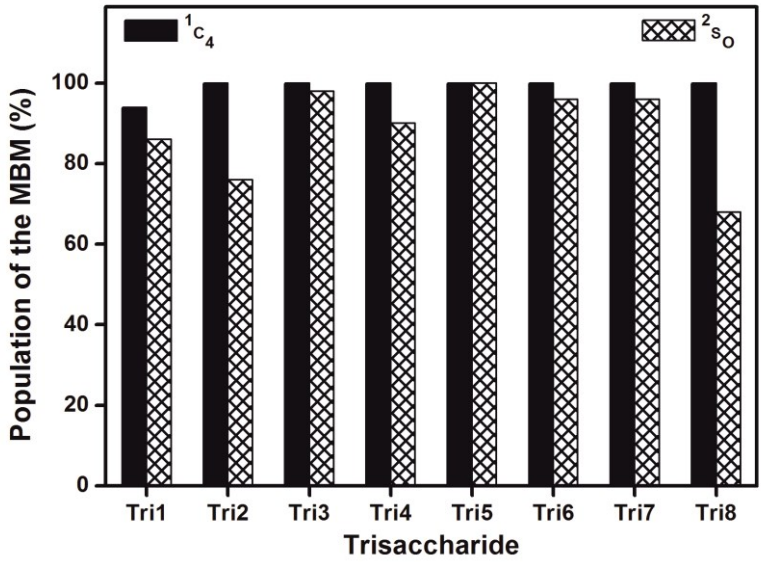

Figure 4. Population of poses adopting the main binding mode for the trisaccharide models containing the iduronate ring in $^1C_4$ chair (filled bars) and ^2S_O skew-boat (lined bars) conformations. The percentages have been calculated for the best 50 *emodel*-ranked poses for each ligand.

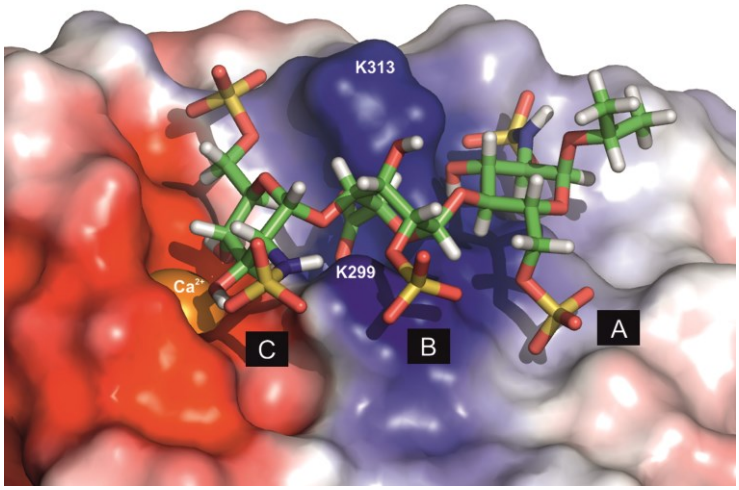

Figure 5. 3D molecular model of **Tri1** bound to Lg ECD (MBM identified from docking calculations). The electrostatic potential is mapped onto the protein surface (blue: positive; red: negative). The positively charged key residues Lys299 and Lys313 are labeled.

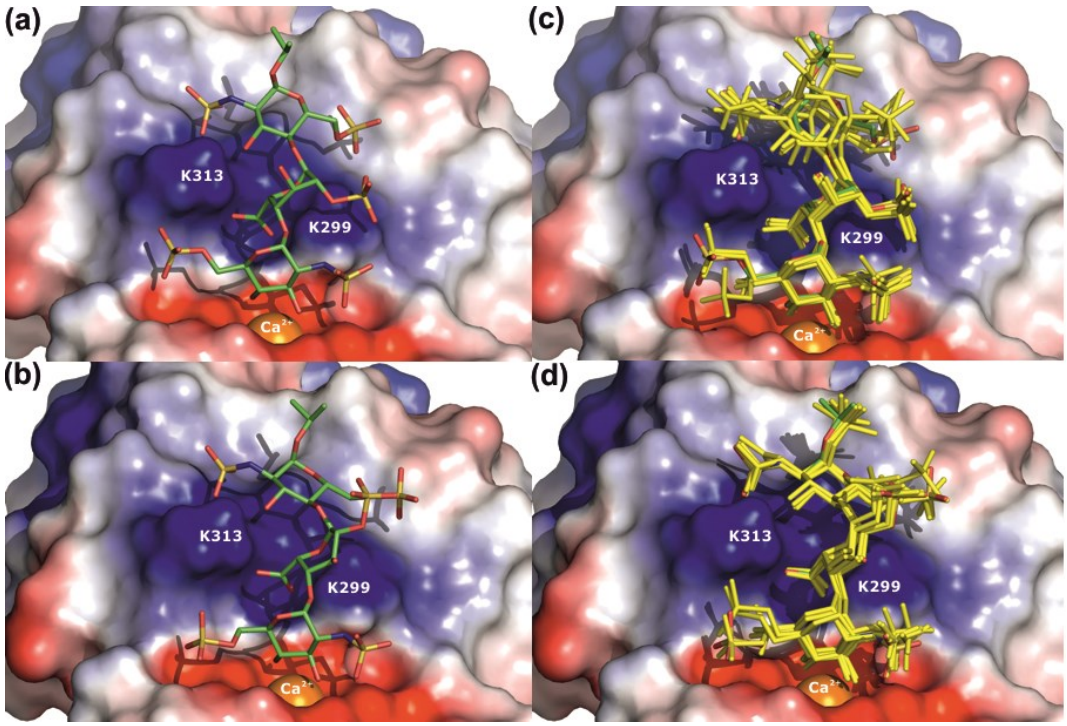

Figure 6. Main binding mode of **Tri1** (colored by atom type; green carbon, red oxygen, yellow sulfur, blue nitrogen) with Lg ECD, for the two possible conformers of the central iduronate ring, $^{1}C_{4}$ (a) and $^{2}S_{O}$, (b). Comparison by superimposition of **Tri2**-**Tri8** main binding modes (in yellow) with that of **Tri1** (coloured by atom type), for the two different conformers of the central iduronate ring, $^{1}C_{4}$ (c) and $^{2}S_{O}$ (d).

On the other hand, we can assume that the SBM (*Appendix*) predicted by docking is dispensable since, in general, it is very lowly populated, it contains high energy poses, and more importantly, it is not able to explain the saturation observed for the protons of the reducing part of the ligands (reducing end GlcN and isopropyl residues stay solvent exposed in the SBM; see *Appendix*).

Going into details, the structural elements that determine the molecular recognition of these heparin-like GAG trisaccharides have been identified. First, the Ca^{2+}-coordination by oxygens O3 and O4 of the non-reducing glucosamine ring explains the Ca^{2+} dependency of the binding; furthermore, and as we would expect, the docking calculations show up that the interactions with the calcium ion act as major stabilizing contributors to ligand binding affinity in all cases. Second, all the trisaccharides showed a double salt bridge involving the coordination of the carboxylate group of the central iduronate ring by the ammonium groups of the side chains of Lys299 and Lys313 (**figure 7**, up). And, third, in all the cases the ammonium group of Lys299 occupies the region below the plane of the hexopyranose ring of the central iduronate ring, in a location that had been previously identified as a specific site for the interaction with calcium ions. In this case it is the ammonium group (instead of a calcium ion) that coordinates the oxygen atoms O6 (COO^- group), OR (oxygen ring) and O4 of the iduronate ring (**figure 7**, down); oxygen O2 of the same residue is also coordinated but only when the iduronate ring adopts the $^1C_4$ chair conformation. Lastly, stabilizing hydrophobic interactions between the reducing end GlcN residue and the aminoacidic residues F315 and/or I250 have been identified as another common structural feature contributing to binding (**figure 8**), although in a minor extent.

Remarkably, none of the above described common key structural elements for the recognition of the GAG trisaccharides by Lg ECD involve the direct participation of any of the variable sulfates of the ligands (N- and 6-O-sulfate groups of residue A, and 6-O-sulfate of residue C), which explains the experimentally observed null effect of the sulphation pattern on the binding epitope. Differently is the possible detrimental or beneficial role that the variable sulfate groups play on binding affinity. In this regard, the comparison of the best *GlideScore*-ranked pose obtained for each trisaccharide model (MBM) suggests that the N-sulphate of the reducing end GlcN residue would slightly increase binding affinity (see *Appendix*), which is reasonable since it establishes favorable interactions with LYS313 (see **figures 5** and **6**). However, the differences are too small for a scoring function (0.5 kcal/mol) as to arrive to solid conclusions regarding the possible different impact of each sulfate position on binding affinity. Furthermore, the same comparison suggest that binding affinity seem to be enhanced (better *GlideScore* score) as long as the iduronate ring adopts the $^1C_4$ chair conformation instead of the ^2S_O puckering (see *Appendix*).

Related to the above commented, to experimentally determine the influence of sulphation pattern on binding affinity, STD NMR competition experiments were carried out with **Tri1** (GlcNS,6S-IdoA2S-GlcNS,6S) and **Tri4** (GlcNS-IdoA2S-GlcNAc,6S). The results demonstrated the higher affinity of the latter (see *Appendix*), so that either the 6-O-sulphate (Glc(C)) or N-sulphate (Glc(A)), or both, were acting as destabilizing forces upon binding to Lg ECD. However, according to the binding mode obtained (see **figures 5** and **6**) the former would establish repulsive interactions

with the binding site (fall on a negative potential energy site), whereas the N-sulphate of the reducing terminal would contribute with favorable interactions with the positively charge LYS313 side chain. This supports that it is probably the 6-O-sulphation at the non-reducing terminal the destabilizing substituent for the binding of heparin-like oligosaccharides to Lg CRD.

Also, aiming to experimentally demonstrate that binding affinity for the interaction of Langerin (CRD) and heparin-like trisaccharides is not only dependent on the favorable interactions of the sugar ring that coordinates the calcium ion, but also on other stabilizing contributions coming from the rest of sugar residues (L-IdoA2S and reducing end GlcN), another STD NMR competition experiment was carried out between a trisaccharide (**Tri2**) and the monosaccharide mannose. In this case, the addition of an equimolar fraction of mannose on a solution containing Lg ECD and **Tri2** did not change the STD intensities of the latter (see *Appendix*). This confirmed that **Tri2** presents a considerably higher affinity for Langerin than mannose, and therefore, that the three trisaccharide residues contribute to binding affinity (with the exception of specific destabilizing substitutions, as commented above).

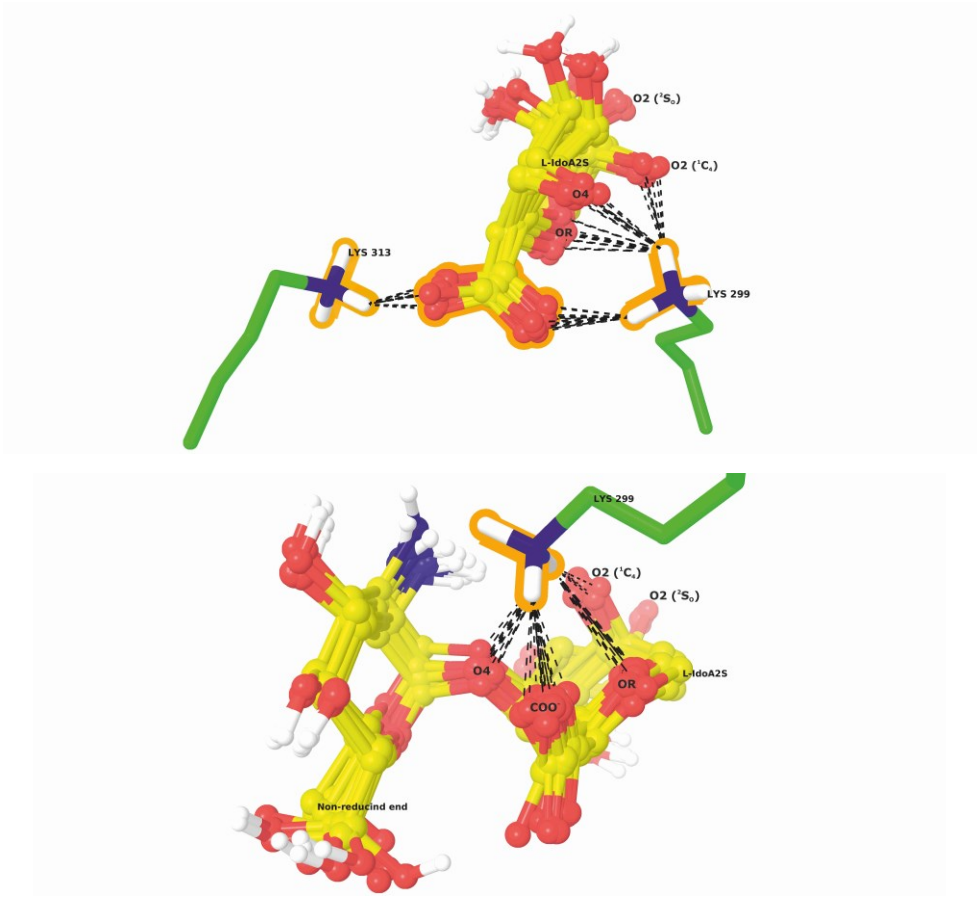

Figure 7. Graphical representation of the two main common structural motifs of the best docking pose obtained for the 16 trisaccharide models (8 with the IdoA2S ring in $^{1}C_{4}$ chair and 8 in $^{2}S_{O}$ skew-boat conformations). (Up) K299-COO^{-}(L-IdoA2S)-K313 salt bridge. (Down) K299 occupying heparin binding site type A for calcium ion[81b]. Note that the oxygen atom O2 of the IdoA2S residue participates in both structural motifs (hydrogen bond with LYS299) only when the iduronate ring is in $^{1}C_{4}$ chair conformation.

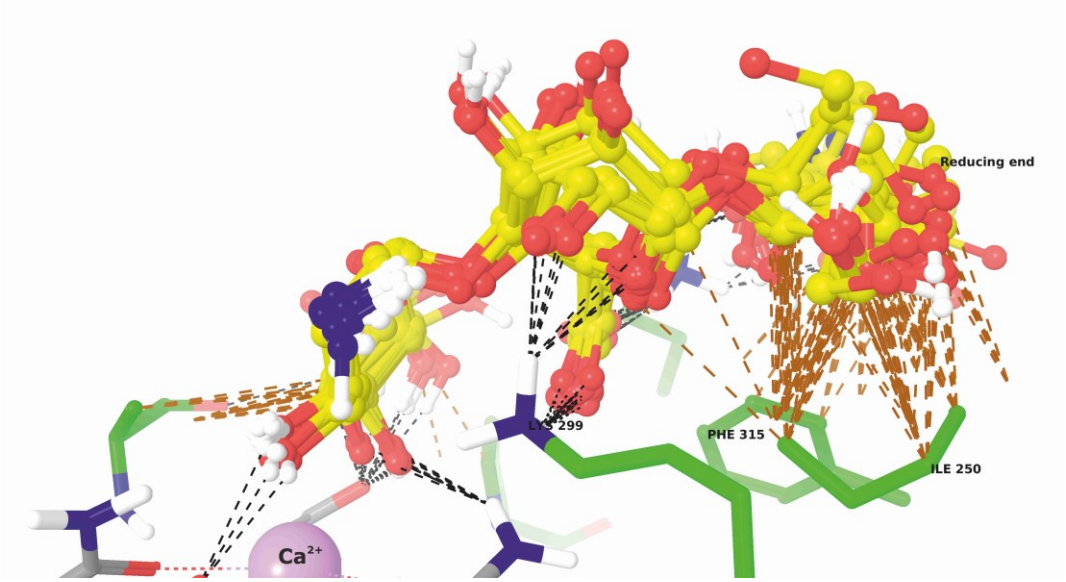

Figure 8. Picture representing the stabilizing hydrophobic interactions (brown dashed lines) between the reducing end GlcN residues and the aminoacidic residues F315 and/or I250. The best *emodel*-docked poses for each of the 16 trisaccharide models appear superimposed. Hydrogen bonds are shown in black dashed lines.

Molecular Dynamics

Despite the very good performance of docking calculations in generating high quality poses able to qualitatively explain the STD-derived binding epitope, there are still some dark zones we need to shed light on related to the agreement between docking and NMR results. Specifically, we refer to the reducing end GlcN residue. Thus, looking at the best docked pose of **Tri1**, shown in **figure 5**, we could reasonably expect, for example, that protons H2 and H4 of the reducing end GlcN ring, closely pointing to the protein surface, received a level of saturation similar to protons H3 and/or H4 of the non-reducing end GlcN residue. However, the **Tri1** binding epitope (**figure 2b**) shows STD intensities of 47/50% for the former pair of protons while 80/100% for the later, thus, indicating that protons H2 and H4 of the reducing end GlcN residue establish further contacts with the protein surface. This behavior is also observed for the other seven ligands (*Appendix*). So, in order to understand the origin of these differences, and based on the known higher dynamic flexibility of the IdoA-GlcN glycosidic linkages, we have carried out a long MD simulation (100 ns) in explicit water with Lg ECD complexed to **Tri1**. The starting structure consisted of the whole trimeric receptor including one copy of the best docking solution of **Tri1** in each of the three Ca^{2+} binding sites. This increases 3-fold the conformational sampling space for a given simulation time. The analysis was focused on the level of flexibility of **Tri1** in the bound state by following the RMSd per sugar residue backbone with respect to the protein binding site, and the conformational space sampled around the ϕ and ψ torsions.

The production dynamics was considered to be fully equilibrated after the first 20 ns, according to the protein backbone RMSd calculated (*Appendix*). The RMSd per sugar residue clearly highlights, on average, the significantly higher mobility of the reducing end GlcN ring with respect to the binding

site, in the three monomer units (*Appendix*). This translates into a larger separation of this residue from the protein surface by variation of both the orientation of the principal axis of the ligand with respect to an imaginary plane defined by LYS299 and LYS313 and the conformation around the flexible IdoA2S-GlcN linkage (**figure 9**). In contrast, the non-reducing GlcN ring stays well fixed on the Ca^{2+} site along the whole simulation. Therefore, this is in very good agreement with the STD NMR epitope, and verifies the stability of the 3D molecular model of the complex obtained by docking calculations.

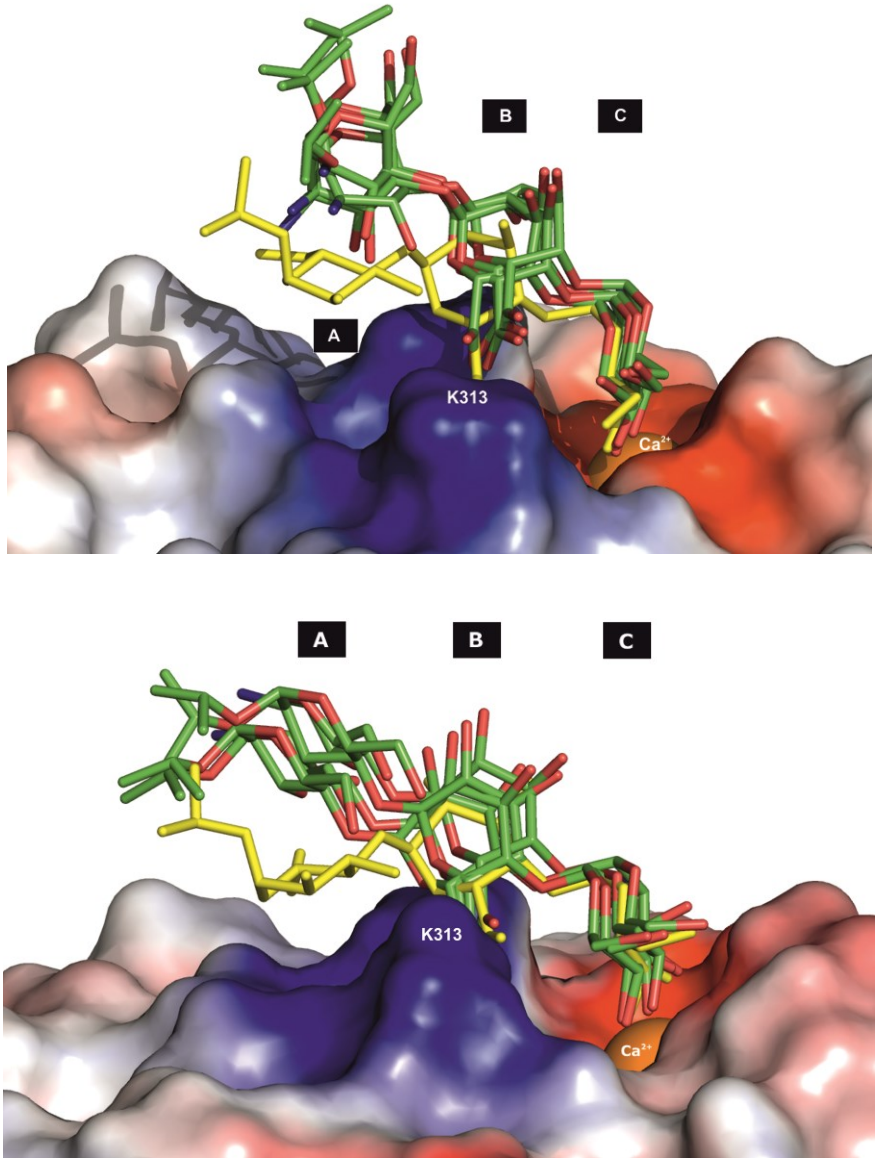

Figure 9. Evidence of the greater mobility of the reducing end GlcN ring of **Tri1** in the bound state coming from some frames of a Langevin Molecular Dynamics (red and green sticks), with the iduronate ring in both $^{1}C_{4}$ chair (up) and $^{2}S_{O}$ skew-boat (down) conformations. The best solution from docking calculations is shown in yellow sticks. The exocyclic groups have been removed for clearance.

4.1.2 Non-sulphated GAGs: hyaluronic acid disaccharides

Results and Discussion

Two hyaluronan-like (HA) disaccharides, of $D-GlcA-\beta(1\rightarrow 3)-D-GlcNAc-O(CH_2)_3N_3$ and $D-GlcNAc-\beta(1\rightarrow 4)-D-GlcA-O(CH_2)_3N_3$ sequences (**figure 10**), abbreviated as **disaccharide 1** and **disaccharide 2**, respectively, have been used to study the structural features of Langerin interactions with hyaluronic acid derivatives.

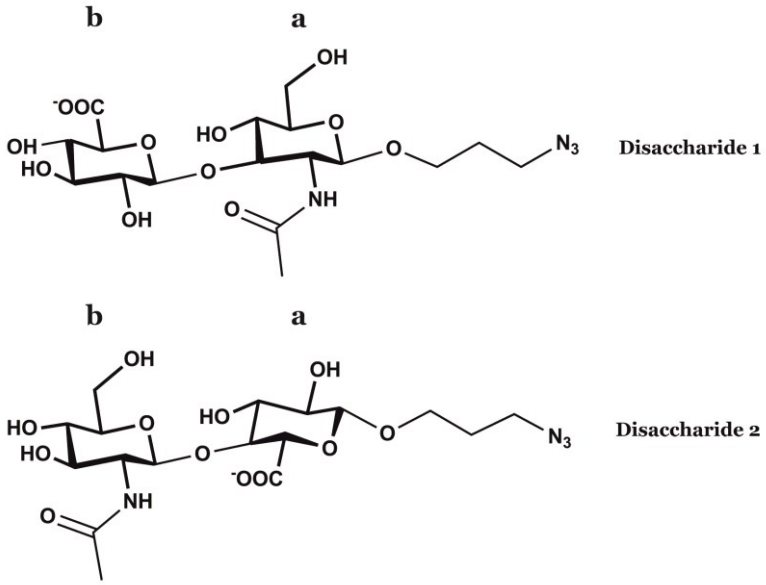

Figure 10. Scheme representation of the hyaluronan-like disaccharides D-GlcA-β(1-3)-D-GlcNAc-$O(CH_2)_3N_3$ (up) and D-GlcNAc-β(1-4)-D-GlcA-$O(CH_2)_3N_3$ (down). The reducing and non-reducing residues are labeled as *a* and *b*, respectively. The 3D structure of hyaluronan (PDB code 2BVK[95]), in terms of the relative disposition of the exocyclic groups, has been considered in this 2D representation.

Prior to the discussion of results, by just considering the chemical structure of both disaccharides, we might expect them to interact with mannose-type specificity with C-type lectins, since both present sites with two consecutive equatorial oxygen atoms ready to chelate the calcium ion. Furthermore, considering that in the case of heparin-like oligosaccharides sequences longer than trisaccharides are needed for the non-calcium dependent interaction in the interface of trimerization[114] (see *subchapter 4.2*, below), HA disaccharides are not expected to interact in a calcium independent manner.

Disaccharide 1 features two possible Ca^{2+}-chelating sites, formed by the pairs of oxygens O2-O3 and O3-O4 of the non-reducing GlcA residue. However, it has to be noted that no experimental evidence of a glucuronate residue interacting in the Ca^{2+}-binding site 2 of a C-type lectin (also known as CRD) has been reported so far. Regarding **disaccharide 2**, it also has two Ca^{2+}-chelating sites, one formed by oxygens O2 and O3 of the reducing end GlcA residue and the other by oxygens O3 and O4 of the

non-reducing GlcNAc residue. Alike for **disaccharide 1**, we would not expect **disaccharide 2** to interact through the GlcA residue, but through the non-reducing terminal.

Again, the combination of protein-ligand interactions based NMR experiments (STD, tr-NOE) and molecular modeling techniques has allowed us to identify the structural characteristics defining the binding of hyaluronan-like oligosaccharides to Langerin.

Nuclear Magnetic Resonance

To deepen into the binding mode of **disaccharides 1** and **2** to Lg ECD, series of STD NMR experiments were carried out at 500 MHz varying the saturation time to record the growth of STD signals. A protein to ligand ratio of 1:60 was used to be at good conditions for the STD experiment, since with an appropriate ligand excess over protein rebinding effects are highly minimized, allowing a real quantification of the STD intensities.

Figure 11 shows the STD NMR experiment for **disaccharide 1**, in which the absence of ligand signals manifests the non-binding properties to Lg ECD for the D-GlcA-D-GlcNAc sequence. On the other hand, binding is observed for **disaccharide 2** (**figure 12b**). Besides, the calcium dependency of **disaccharide 2** interaction was corroborated upon addition of an excess of EDTA-d_{12} (**figure 12c**).

The build-up curves for **disaccharide 2** were obtained (**figure 13**), showing that the non-reducing GlcNAc residue receives the largest amount of saturation. Furthermore, a lower and heterogeneous saturation is observed for the reducing GlcA residue whereas the 3-azidopropyl glycoside did not give rise to STD signals, suggesting solvent exposure. Therefore, from the STD build-up curves we could guess a binding mode in which the sugar interacts through its non-reducing terminal with solvent exposure increasing just as we move away from it.

From these STD build-up curves, we mapped out the main contacts of the ligand with Langerin in the bound state (binding epitope), by determining the initial growth rates of the curves and normalizing all the values within a given ligand by the highest one, to which arbitrarily a value of 100% was assigned. Again, remember that the use of initial slopes approximation increases the accuracy of the method by avoiding the detrimental effects of different relaxation properties of the ligand protons on the determination of the binding epitope.

According to the NMR-derived binding epitope of **disaccharide 2** (**figure 14**), the protons that establish the closest contacts with the protein surface are H3, H4, H5, H6 and the acetamide protons of the non-reducing GlcNAc ring, with H3 showing the largest saturation transfer (100 % relative STD) among them, thus highlighting the importance of this part of the molecule for the molecular recognition by Lg ECD. This result, along with the observation of the abolishment of binding to Lg ECD upon removal of the divalent cation, strongly supports that **disaccharide 2** bind Langerin at the Ca^{2+}-dependent binding pocket, by a classical coordination of the calcium ion through the

di-equatorial oxygen atoms O3 and O4 of the non-reducing GlcNAc ring. Furthermore, STD signals are also observed for the reducing end GlcA residue, with proton H3 showing the highest intensity (comparable to proton H1 of the non-reducing terminal; **figure 14**), which highlights that the GlcA residue participate, although in a minor extent, upon binding. It should be noted that GlcA protons H4 and H5 also gave rise to STD signals but their quantification was not possible due to overlapping. Regarding the methylene protons of the 3-azidopropyl glycoside tail, they do not participate in binding (absence of STD signals), so this substituent will stay solvent exposed in the bound state.

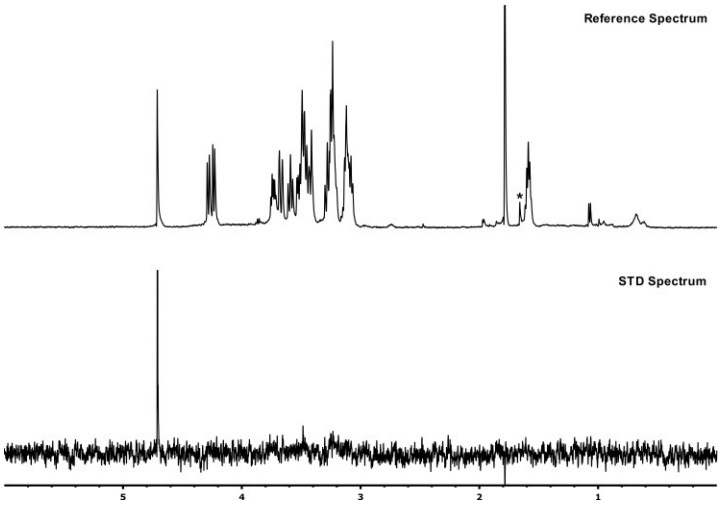

Figure 11. Reference and STD spectra of **disaccharide 1** in the presence of Lg ECD (1:60 protein-to-ligand ratio), at 10 ºC, 6s saturation time and 500 MHz.

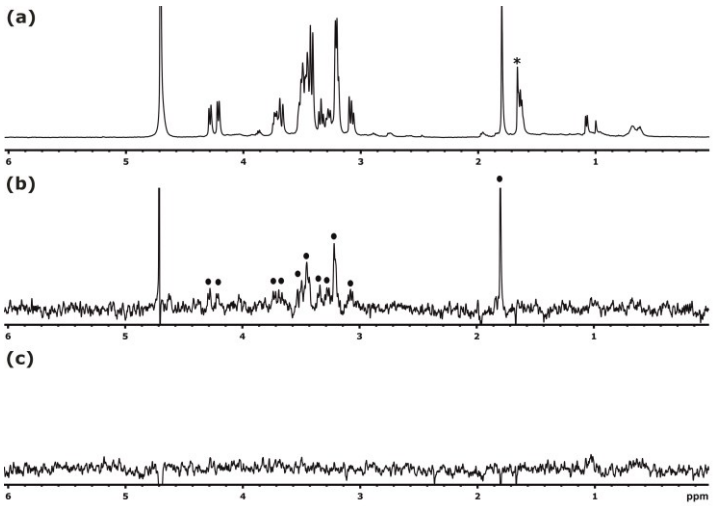

Figure 12. The interaction of the HA-like **disaccharide 2** with Lg ECD is Ca^{2+}-dependent. (a) 1D ^{1}H-NMR (500 MHz) reference spectrum (*off-resonance*) of **disaccharide 2** in the presence of Lg ECD in 4 mM Ca^{2+} Tris-d_{11} buffer (the buffer signal is indicated with a star symbol). (b) 1D STD NMR spectrum of **disaccharide 2** in the presence of Lg ECD in 4 mM Ca^{2+} Tris-d_{11} buffer (black circles mark ligand STD signals). (c) 1D STD NMR spectrum of **disaccharide 2** in the presence of Lg ECD in 10 mM EDTA Tris-d11 buffer. Note: a protein-to-ligand ratio of 1:27, 4s saturation time and a temperature of 10 ºC have been used.

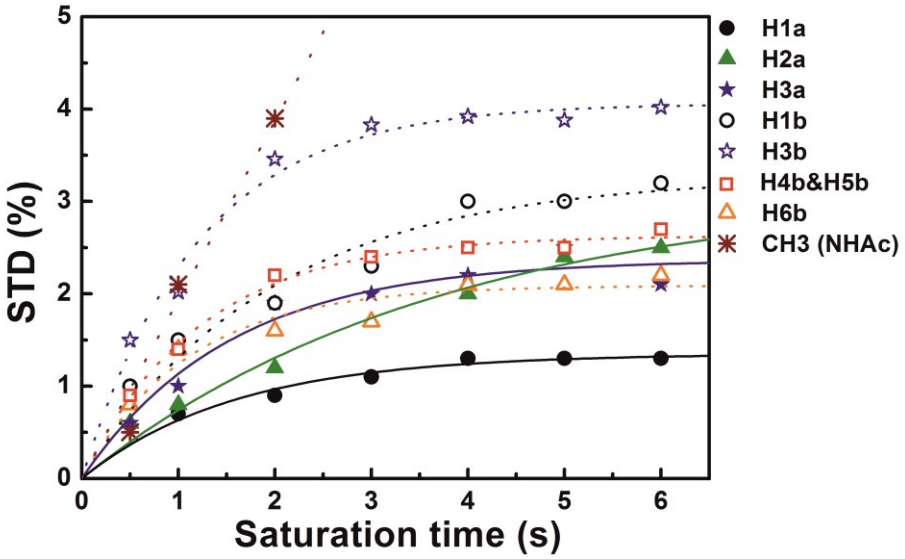

Figure 13. STD growth curves obtained for the interaction of disaccharide 2 with LgECD, at 10 ºC and using a 1:60 protein-to-ligand ratio. Note: Due to overlapping, STD intensities of protons H4b and H5b have been approached to be equivalent. For the methyl (NHAc) group, only the 3 first points with less saturation time have been considered for the regression (linear in this case) due to the deviation from the mono-exponential behavior of its STD curve.

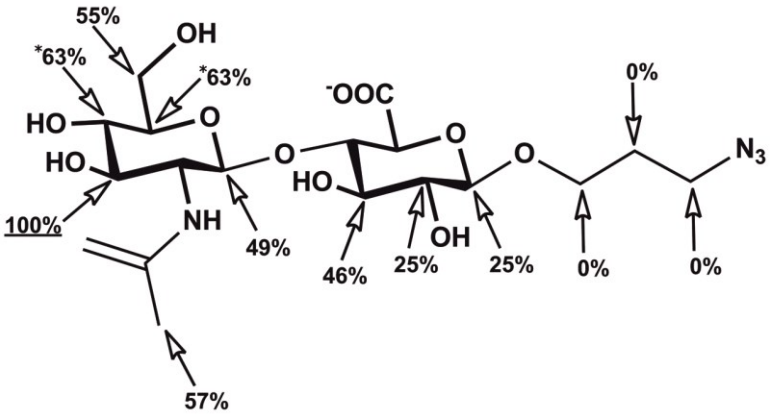

Figure 14. STD-based binding epitope of disaccharide 2 interacting with LgECD. Values calculated from the initial slopes of STD amplification factors (STD-AF_0) and normalized against the most intense STD signal (proton H3 of the GlcNAc residue). The not-assigned protons were not quantified due to overlapping. *STD intensities for protons H4 and H5 have been roughly approximated to be equivalent due to overlapping.

Aiming to investigate the impact on the conformation of **disaccharide 2** upon binding to Lg ECD, we also carried out transferred NOESY experiments on the same samples used for the STD NMR experiments, containing **disaccharide 2**, Lg ECD, and buffer with Ca^{2+}. The resulting spectrum (**figure 15**) confirmed the low-energy *syn-ψ* geometry (H1b/H4a NOE peak) for the β(1-4) glycosidic

linkage in the bound state. Furthermore, no contribution of *anti-ψ* conformations (absence of H1b/H3a NOE peak) was observed. This is in agreement with the reported structure of hyaluronan in solution[95]. The stability of the $^4C_1$ chair upon binding was also confirmed (**figure 15**; NOEs H1b/H3b and H1b/H5b for the GlcNAc residue; NOEs H1a/H3a, H2a/H4a for the GlcA ring). Thus, binding to Langerin does not induce any conformational change on **disaccharide 2**.

As a general summary so far, the STD-NMR experiments described above have demonstrated that hyaluronan-like compounds do not bind to the CRD of Langerin through a glucuronate residue. Further, a hyaluronic acid oligosaccharide with a GlcA residue as the non-reducing terminal (independently of the sugar residue present in the reducing end) will not interact with Langerin CRD due to the lack of consecutive di-equatorial oxygen atoms available in the non-terminal GlcNAc residues (oxygen O3 participates in the glycosidic bond). To sum up, HA oligosaccharides are only able to bind Langerin CRD through a non-reducing GlcNAc residue, and no conformational change occurs upon binding.

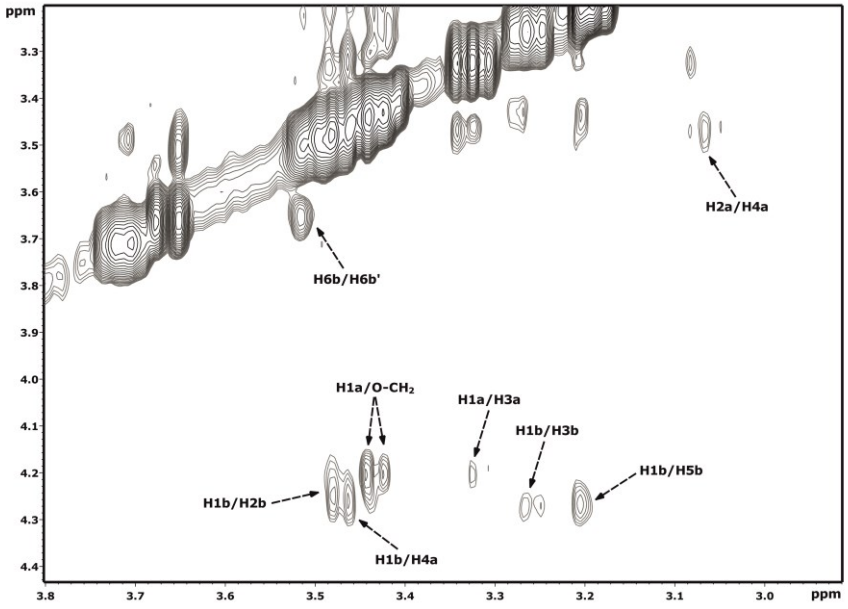

Figure 15. Expansion of the Transferred-NOESY spectrum (mixing time 500 ms) of disaccharide 2 in the presence of Lg ECD (bound state).

Modelling

Generating poses: flexible docking

The computational technique docking has been used to obtain binding modes for **disaccharide 2** which are able to explain its NMR-derived binding epitope (**figure 16**). Thus, the *glide*[246] module of Schrodinger 2012 Suite was employed to carry out flexible docking with the Langerin binder **disaccharide 2**.

Apart from the general factors, already indicated in chapter 1, which complicate the computational docking and scoring of protein-carbohydrate interactions, two more sources of complexity are present in the particular system formed by Langerin CRD and HA disaccharides. First, as a reminder of what was outlined for the docking of heparin-like trisaccharides, the Ca^{2+} binding site of Langerin is thoroughly solvent exposed. Thus, the number of site-points, orientations and conformations to sample and score will be considerably higher than, for instance, in the case of a groove-type binding site. Second, unlike the case of heparin-like trisaccharides interactions with Langerin CRD, in which the high charge density of the ligand played a key role in the quality of the results reproduced by glide SP algorithm (the high number of stabilizing charge-charge interactions with LYS residues 299 and 313 acted as important contributors upon binding), in **disaccharide 2** only one negative charge is present. Thus, it is reasonable to think that those stabilizing forces will not intensively drive **disaccharide 2** docking process, which, in principle, makes this system even more challenging from the docking point of view.

Based on both the known influence of the input ligand geometry on docking accuracy[246a] and the poor docked poses we obtained in a first test using flexible SP docking (default settings) with **disaccharide 2**, we focused on strategies to enhance the sampling and rescue false negative poses. Thus, we carried out 1) a Monte Carlo Multiple Minimum (MCMM) torsional sampling to obtain a set of different starting conformations to be sourced to the docking calculation, and 2) a smooth "induced fit" during the docking process by reducing the van der Waals radii of the non-polar ligand atoms. Also, the distance-dependent dielectric constant was increase up to 4.0 (2.0 by default) since according to our experience it is another variable which improves the docking process for carbohydrates. These three factors resulted in a great enrichment of the sampling and, consequently, the recovery of false negatives. However, we have to note that, at the same time, the number of false positive obtained is also enhanced.

The pair of hundreds of disaccharide 2 conformers (obtained from MCMM sampling) submitted to a flexible SP docking job (using the settings described above) gave rise to many thousands of poses. However, a high percentage of them were immediately rejected since 1) they did not match the GlcNAc calcium coordination experimentally determined and/or 2) the ligand conformation in the bound state contradicted the exo-anomeric effect (*anti-*ϕ conformers were obtained) and/or the *syn-*ψ geometry of the glycosidic linkage experimentally determined (tr-NOESY experiments; see **figure 15**). Related to the latter, it has to be noted that a recent structural study (50 ns of molecular dynamics simulations) of different hyaluronan oligosaccharides[95] also reported rigid *syn-*ψ conformation (no *anti-* ϕ) for the GlcNAc-β(1→4)-GlcA linkage.

Two binding modes have been easily identified among the 1600 poses after the first clustering (the orientation of the non-reducing GlcNAc residue was clustered), which we have named as the crystallographic and the non-crystallographic binding mode. The crystallographic binding mode (**figure 16**, a and b) resembles that of the non-reducing glucose residue of laminaritriose bound to Langerin CRD (chain C of the crystal structure 3P5H). This can be arbitrarily defined as the O3→O4

orientation as long as we look at the sugar leaving LYS299 and LYS313 residues behind and considering the direction from left to right. On the other hand, the non-crystallographic binding mode identified resembles the 180º-rotated of the crystallographic orientation. So, it can be assigned with the O4→O3 orientation (**figure 16**, c and d).

In order to distinguish the different conformers contained in each binding mode, a subsequent torsional clustering was carried out. For the crystallographic and non-crystallographic set of poses, 24 and 18 clusters were generated, respectively. The analysis of the results has been done based on the best scored (*glide emodel*) pose of the 10 most populated clusters obtained for both the O3-O4 (**figure 16**, b) and O4-O3 orientations (**figure 16**, d). Thus, comparing the most populated cluster of each binding mode, the best ranked pose adopts the crystallographic orientation (-22.7 *glide emodel* score, versus -20.2 for the best ranked pose in the non-crystallographic orientation), suggesting that this binding mode might be slightly favored.

The crystallographic binding mode is characterized by hydrogen bonding between the side chains of the aminoacidic residues that coordinate the calcium ion (EPN and WND motifs) and the OH groups in positions 3 and 4 of the non-reducing GlcNAc residue. The proton of the 6-OH group of this residue is also involved in hydrogen bonding with the E285 side chain. Furthermore, the acetamide group points to K313, and the carboxylate group of the GlcA ring partially establishes stabilizing charge-charge interactions with K313 in some poses, although it is quite solvent exposed.

On the other hand, the non-crystallographic binding mode features the same kind of hydrogen bond interactions with the EPN and WND motifs, with the 6-OH acid proton of the glucosamine residue interacting in this case by hydrogen bond with the E293 side chain. In addition, both the carbonyl oxygen of the acetamide group and the carboxylate function interact favorably with lysine 299 (by hydrogen bond and electrostatic interactions, respectively).

In both binding modes predicted by docking calculations, the 3-azidopropyl tail stays solvent exposed in the most populated clusters, in agreement with the STD NMR results. In any case, it has to be noted that some low-populated clusters of poses in which this tail interacts with the protein surface has been obtained. On the other hand, regarding the most representative clusters, the geometry of the glycosidic linkage is equal or very similar to that of the resolved structure of hyaluronan in solution[95]. In addition, alike it happened for Lg ECD interactions with **Tri1-Tri8**, docking calculations predict that **disaccharide 2** interactions with the calcium ion act as major contributors upon binding (obvious for a calcium-dependent interaction), confirming the goodness of docking (see, for instance , **table 5**).

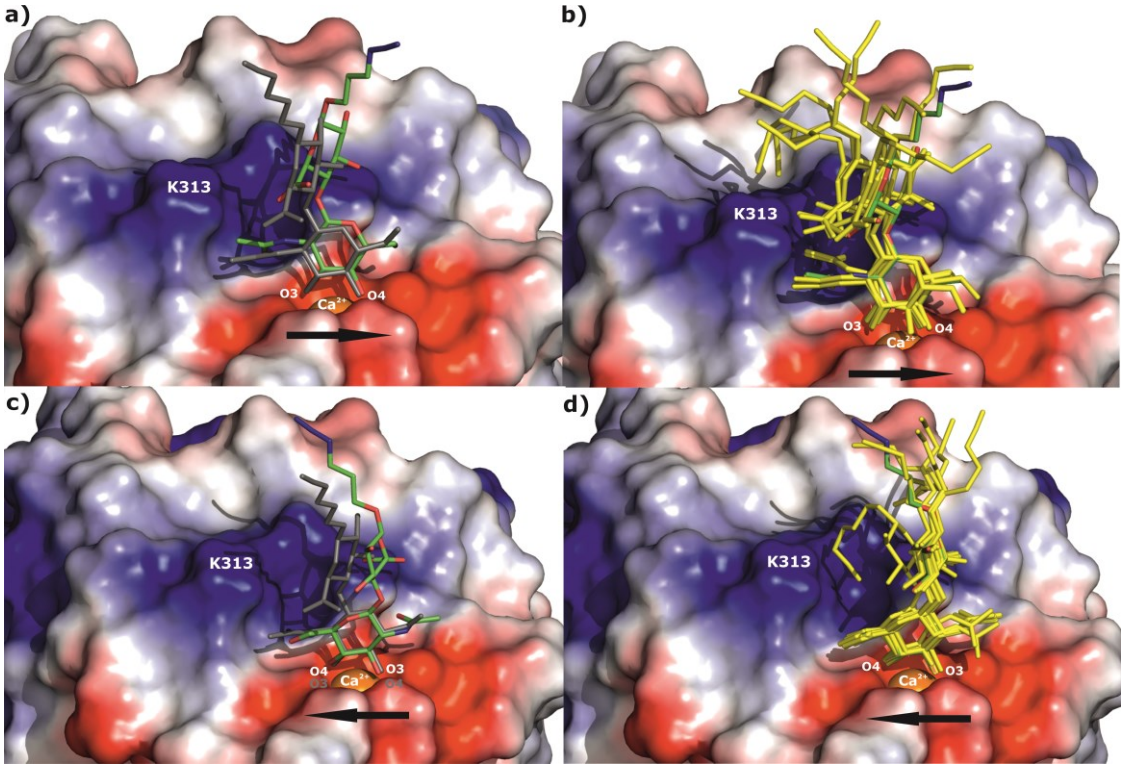

Figure 16. (a) Best ranked pose of the most populated cluster for the crystallographic binding mode (green and red sticks) compared to the initial pose (grey sticks). (b) Same as (a) plus the best rank pose of the following 9 most populated clusters (yellow sticks). (c) Best ranked pose of the most populated cluster for the non-crystallographic binding mode (green and red sticks) compared to the initial pose (grey sticks, O3-O4 orientation). (d) Same as (c) plus the best rank pose of the following 9 most populated clusters (yellow sticks). An arrow indicates the O3-O4 direction in each case.

The detailed comparison of the STD-derived binding epitope with both binding modes obtained from docking calculations lead to interesting conclusions (see **figure 17**). Specifically, regarding proton H3 of the non-reducing terminal, it features slightly closest contacts with the protein surface in the non-crystallographic orientation compared to the crystallographic binding mode. The opposite applies for proton H4 of the same residue. So, considering one binding mode or the other, they should not give rise to very different saturation levels for protons H3 and H4 of the GlcNAc residue, which is logic since they both are symmetrically involved in the calcium-coordinating core of the ligand. At this point, it has to be noted that, although we have observed a significantly higher STD intensity for proton H3 compared to proton H4 (100% and 63% respectively), we have to keep in mind that the STD signal of the later was impossible to integrate accurately due to overlapping. Otherwise, regarding proton H1 of the GlcNAc residue, whereas it points to the solvent in the O3-O4 orientation, contacts with the protein surface in the distance range of the proton receiving the highest amount of saturation, H3b, occur in the O4-O3 orientation. Thus, assuming the non-crystallographic model, both protons should give rise to similar and very high STD intensities, while in the crystallographic binding mode these protons would not receive saturation (or very little). Nevertheless, neither of both binding modes are able to explain on their own the approximately half of STD intensity observed for proton H1

compared to H3 of the GlcNAc residue. An analogous reasoning applies for proton H3 of the GlcA residue, which points to the protein surface and to the solvent in the crystallographic and non-crystallographic binding modes, respectively, and presents a STD factor of about 50%. Interestingly, the methyl protons of the acetamide group and the methylene protons of position 6 of the non-reducing terminal present the same STD factor, in agreement with the equal proximity of both groups to the protein surface in both binding modes (**figure 17**).

All the results described and discussed above lead to conclude that both docked binding modes contributes upon binding. Further, the fact that a STD factor of about 50% is observed for proton H1 of the GlcNAc residue and proton H3 of the GlcA ring, together with the former presenting close interactions and being solvent exposed in the O3-O4 and O4-O3 orientations, respectively, (with the opposite occurring for proton H3a) suggest that the interaction of **disaccharide 2** is favored in a similar extent in both binding modes (approximately equal contribution upon binding or 1:1 ratio).

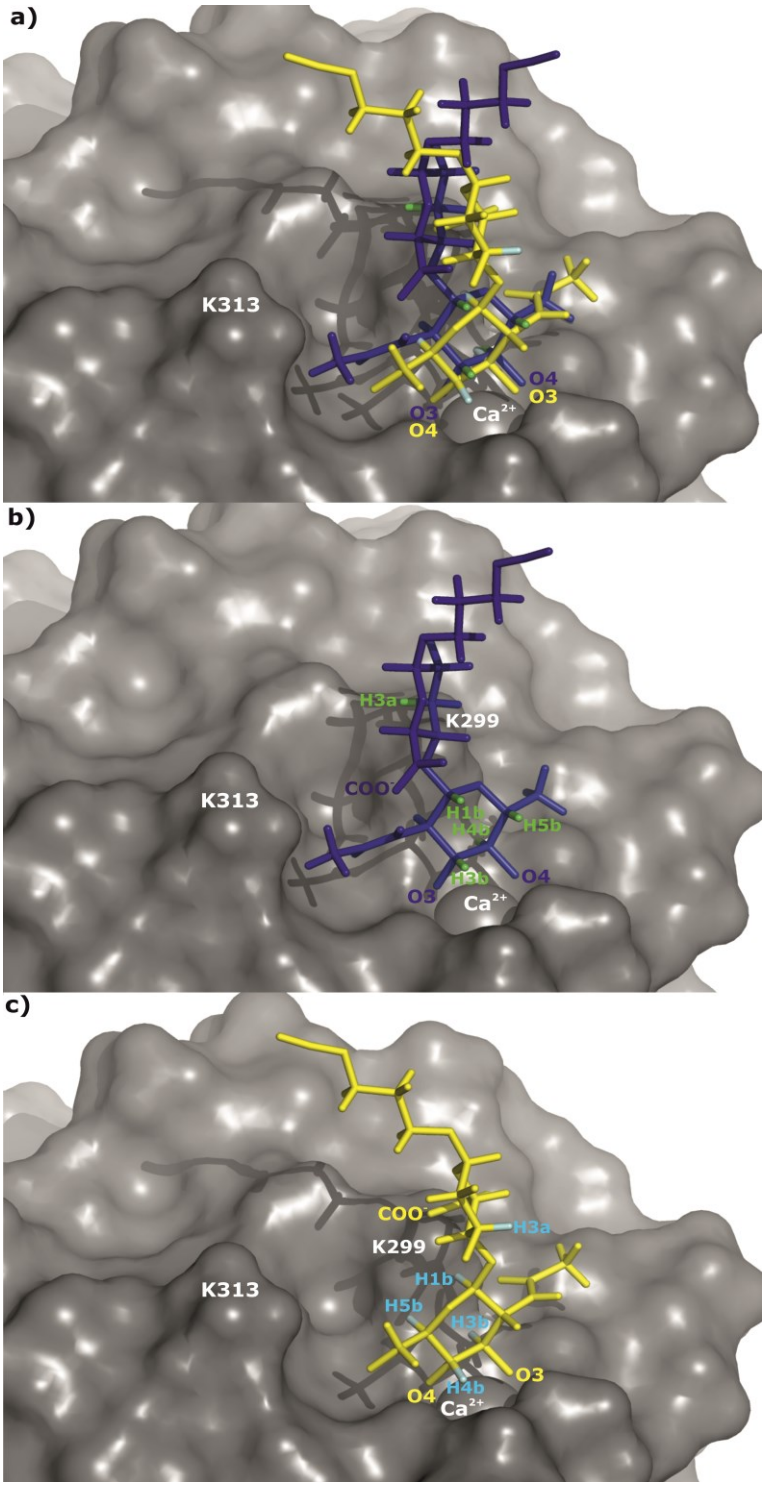

Figure 17. Comparison (a) of the best docked pose obtained in the crystallographic (b, purple) and its 180º-rotated (c, yellow) orientation for **disaccharide 2**. The protons used for comparison with the STD-derived binding epitope are colored green and cyan in the crystallographic and non-crystallographic orientations, respectively.

Structural basis for non-binding of disaccharide 1: refine docking

Aiming to understand the structural reasons that cause HA sequence GlcA-GlcNAc (**disaccharide 1**) to not bind to Langerin CRD as opposed to the binding sequence GlcNAc-GlcA (**disaccharide 2**), both disaccharides were submitted to a *glide refine* calculation (do not dock) adopting both the laminaritriose-type orientation (O3-O4) and its 180º-rotated (O4-O3), to be consistent with the binding modes obtained by flexible docking for **disaccharide 2** and, at the same time, with the x-ray resolved structures of Lg CRD bound to neutral carbohydrates reported to date. The *refine* docking calculation in *glide* consists of an optimization of the ligand structure in the field of the receptor followed by the scoring of the ligand. It has to be noted that another pair of di-equatorial oxygens, O2-O3, are available in the GlcA residue to coordinate the calcium ion. However, we dismissed this type of coordination for *glide refine* docking since steric clashes between the reducing GlcNAc residue and the protein surface are observed when we build the models (with laminaritriose as reference for the superimposition) in both the O2-O3 and O3-O2 orientations (see *Appendix*).

Disaccharide 2 was better scored than 1 in the r*efine* docking calculations, considering the three most representative scoring functions, i.e., *GlideScore*, *Emodel* and G*lideEnergy* (**table 4**). It has to be noted that *GlideScore* is the most convenient scoring function to compare poses of different chemical species in terms of the binding affinity[246a]; however, *GlideScore* is just an estimate of it (remember that computing accurate absolute, or even relative, binding energies is still an extremely challenging task). In any case, the fact that these three scoring functions agree in better punctuating the GlcNAc-GlcA disaccharide is indicative of the right performance of *glide* scoring algorithms in distinguish the active from the non-active compound.

Table 4. *GlideScore*, *Emodel* and *GlideEnergy* scoring functions, calculated from *glide refine* docking, for **disaccharides 1** and **2** in the laminaritriose-type orientation (03-04) and its 180º-rotated (04-03).

	O3-O4 orientation		**O4-O3 orientation**	
	GlcA-GlcNAc (**1**)	GlcNAc-GlcA (**2**)	GlcA-GlcNAc (**1**)	GlcNAc-GlcA (**2**)
GlideScore	-1.0	-1.3	-0.7	-1.5
Emodel	-20.4	-25.9	-18.3	-25.0
GlideEnergy	-16.1	-20.0	-15.6	-18.5

Moreover, in order to deepen into the particular destabilizing interactions which preclude Langerin CRD complexed to **disaccharide 1** to exist (more exactly, to exist during a long enough time as to be observed by STD NMR spectroscopy), per-residue interactions were calculated for the two refined poses of each disaccharide. It is known from previous studies that the networks of interactions between the calcium ion, the carbonyl residues that coordinate it and the sugar hydroxyls are major contributors to the basic binding affinity and specificity of mannose-type monosaccharides to the

Ca^{2+}-binding site 2 of Langerin (conventional sugar binding site)[115]. Thus, we have focused on the non-bonding contributions (Coulomb and Van der Waals) of the calcium ion and the carbonyl residues that coordinate it (EPN and WND motifs) to explain the non-binding and binding properties of **disaccharides 1** and **2**, respectively (see **table 5**).

Table 5. Per-residue non-bonding contributions (sum of the Coulomb and Van der Waals terms) of the calcium ion and Langerin aminoacidic residues of the EPN and WND motifs, obtained from a *glide refine* docking calculation for **disaccharides 1** and **2** in both the laminaritriose-type O3-O4 orientation and its 180°-rotated (O4-O3 orientation).

	O3-O4 orientation		**O4-O3 orientation**	
	GlcA-GlcNAc (**1**)	GlcNAc-GlcA (**2**)	GlcA-GlcNAc (**1**)	GlcNAc-GlcA (**2**)
GLU285	14.9*	4.4	9.3	8.7
ASN287	-2.3	-1.2	-1.3	-2.9
GLU293	11.3	10.2	14.0*	5.6
ASN307	-2.8	-4.5	-3.7	-2.8
ASN308	11.4	9.1	11.3	8.9
Ca^{2+}	-40.5	-32.2	-38.2	-32.2
Total	-8.0	-14.2	-8.7	-14.7

* Most significant destabilizing contributions upon binding.

According to the values reported in **table 5**, the overall interaction energy of the ligand with the set of residues constituted by the calcium ion and the amino acids of the EPN and WND motifs is about 40% less favorable for the GlcA-GlcNAc sequence, with the most significant destabilizing contributions coming from the interactions with GLU285 (in the O3-O4 orientation) and GLU293 (in the O4-O3 orientation). These represent repulsive interactions between the carboxylate groups of the GlcA residue and a glutamate side chain, due to the parallel disposition of the former with respect to GLU285 and GLU293 in the O3-O4 and O4-O3 orientations, respectively (**figure 18**).

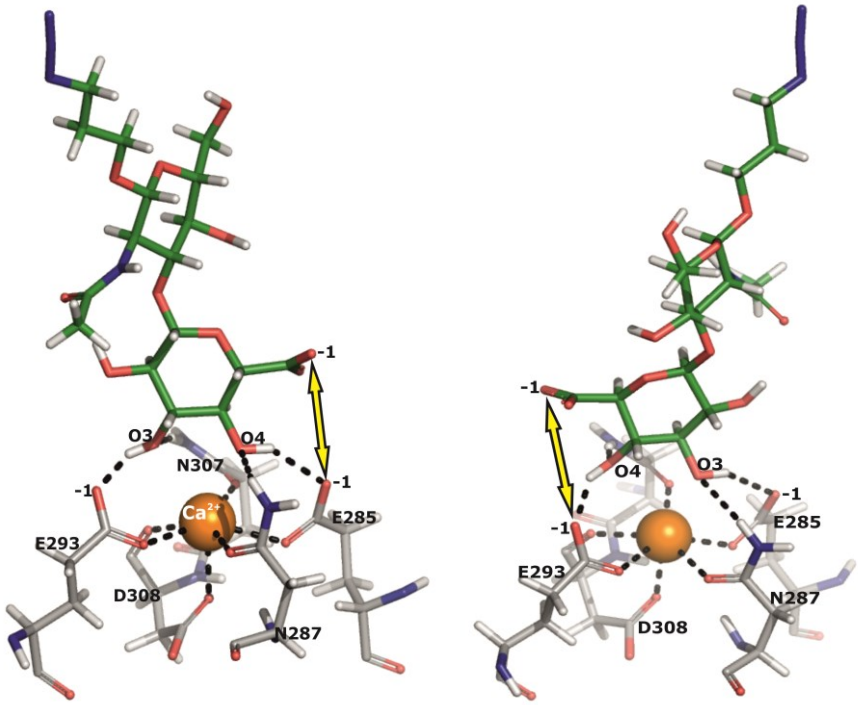

Figure 18. *Glide refine* docked poses for the non-binder **disaccharide 1** in both the crystallographic laminaritriose-type binding mode (left, O3-O4 orientation) and its 180º-rotated (right, O4-O3 orientation).

4.2 Non calcium-dependent interactions

4.2.1 Results and Discussion

The interaction of long heparin-like oligosaccharides with Langerin ECD is Ca^{2+} independent.

In *section 4.1*, we have demonstrated that heparin-like GAG trisaccharides **Tri1-Tri8** interact with Lg ECD in a different mode than larger oligosaccharides of natural heparin (6 and 15 kDa)[114], with the former interacting through Ca^{2+} chelation in the conventional calcium binding site of Langerin CRD and the latter in the interface of trimerization of Lg ECD in a calcium-independent manner. Thus, in order to depict the size requirements of heparin-like oligosaccharide on their ability to bind to the calcium independent binding site of Lg ECD, we decided to investigate the binding of a synthetic hexasaccharide, **Hexa4S** (**figure 19**), which harbors the regular sulphation pattern typical of heparin plus an additional O-sulfate group at position 4 of the non-reducing glucosamine ring. This latter chemical modification should preclude **Hexa4S** from binding at the conventional calcium binding site of Langerin CRD, since one of the adjacent OH groups (OH4) in di-equatorial configuration necessary for the calcium chelation is substituted by a bulky charged sulphate group.

The STD NMR spectra of a sample containing an excess of the hexasaccharide (1:21 protein to ligand ratio) in the presence of Lg ECD are shown in **figure 20**. In the presence of 4 mM Ca^{2+} ions the experiment showed strong STD signals indicative of binding to the protein (**figure 20b**). Interestingly, in contrast to the trisaccharides, the removal of Ca^{2+} ions from the sample by sequestration with a large excess of EDTA did not have an effect on the binding to Lg ECD, as strong signals were observable in the STD spectrum (**figure 20d**). This strongly supports that the binding of the **Hexa4S** to Lg ECD takes place in a different binding site that is not Ca^{2+} specific, as it was guessed from the above mentioned presence of the additional 4-O-sulfate group.

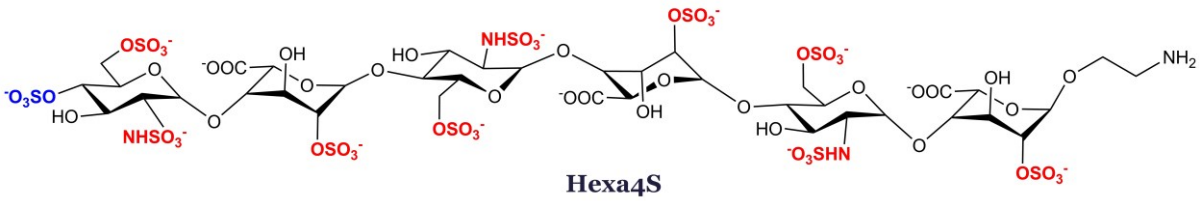

Figure 19. Scheme representation of the heparin-like hexasaccharide **Hexa4S**. Note the 4-O-sulphate group at the non-reducing terminal (bold blue) precluding the interaction in the calcium chelating site of Langerin. The 3D structure of heparin (PDB code 1HPN[65]), in terms of the relative disposition of the exocyclic groups, has been considered in this 2D representation. Sulphate and sulphamate groups are shown in red.

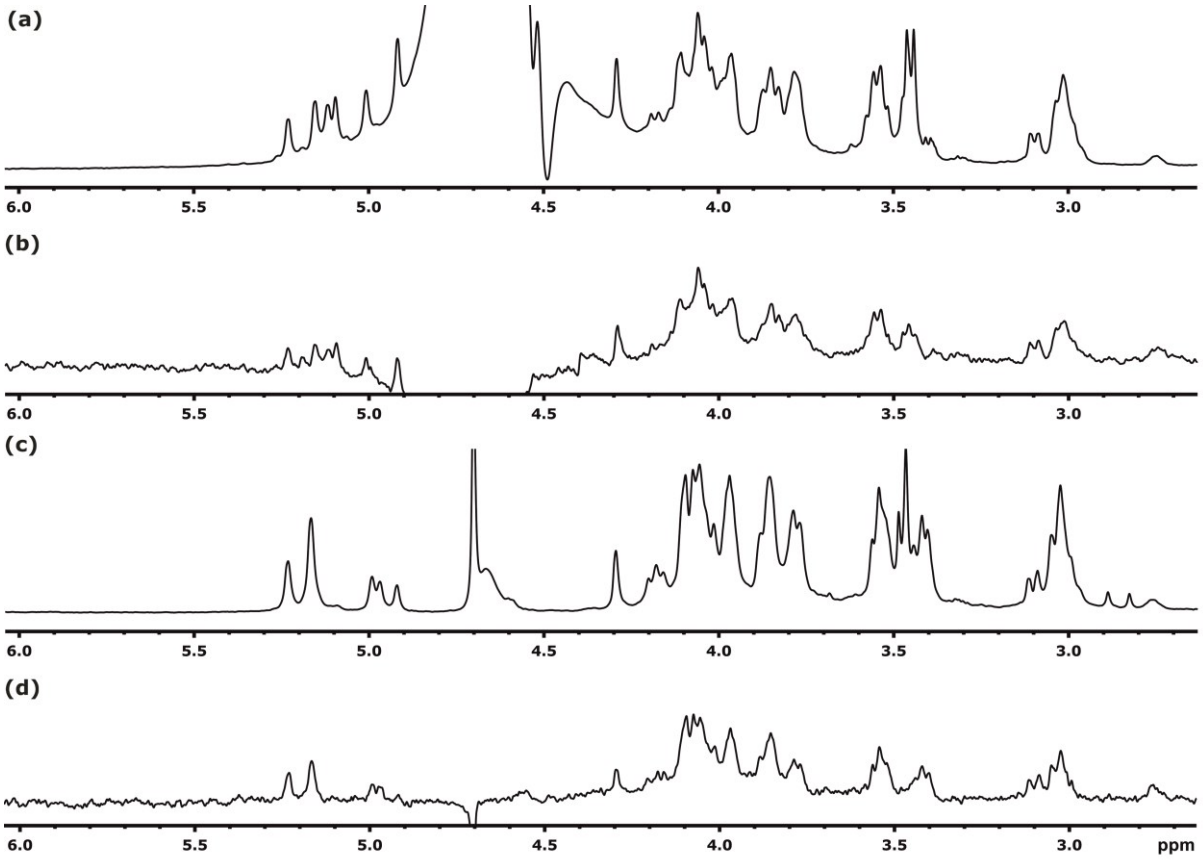

Figure 20. The hexasaccharide binds Lg ECD in a different binding mode in comparison to trisaccharides **Tri1**-**Tri8**, and in a Ca^{2+} independent way. (a) Reference ^{1}H NMR spectrum of **Hexa4S** in the presence of Lg ECD in a buffer containing 4 mM Ca^{2+} ions and (b) STD NMR spectrum of that sample. (c) Reference ^{1}H NMR spectrum of **Hexa4S** in the presence of Lg ECD in a buffer containing 4 mM Ca^{2+} ions and 10 mM EDTA to sequester the divalent cations. (d) STD NMR of the latter sample.

The strong signal overlap of the spectra of **Hexa4S** (**figure 20**) precluded a detailed determination of its binding epitope for its interaction with Lg ECD. Nevertheless, a couple of features emerged clearly from the data: first, there is an homogeneous distribution of the saturation along the whole ligand, indicative of binding in an orientation parallel to the protein surface, without existing a main contact with a specific sugar ring of the ligand; and second, there are changes in the chemical shifts of the hexasaccharide upon removal of Ca^{2+} ions (particularly of the anomeric protons; see **figure 20**, c and d). This is due to the specific interaction of **Hexa4S** with the divalent cations in the free state, a behavior that has been previously described for synthetic equivalent hexasaccharides[81b].

The NMR data on the molecular recognition of **Hexa4S** by Lg ECD are supporting an interaction similar to that described previously for 6 kDa and 15 kDa heparin. In this proposed binding mode, the hexasaccharide would occupy part of the groove formed between two adjacent carbohydrate recognition domains (CRD), repeated three times over the protein surface due to the symmetric nature of Langerin. Also, it would interact with the α2 helix of one CRD and the α1 of the adjacent CRD, and part of the α-helix coil-coil parts of both CRDs. This groove is lined by positively charged side chains of

lysine residues, which accommodate very well the negatively charged groups of the hexasaccharide, making unnecessary the coordination of any hydroxyl group of the ligand with Ca^{2+} on the surface of the protein CRD for the binding to Lg ECD to occur.

As a final remark, we have demonstrated that an hexasaccharide sized heparin is long enough to bind the previously identified calcium-independent binding site in the interface of trimerization of Lg ECD (in that case for natural heparin of 6 and 15 kDa). However, it has to be noted that by using the synthetic hexasaccharide **Hexa4S** we have precluded the interaction with the Ca^{2+} binding site of Langerin CRD (4-O-sulphation in the non-reducing terminal), and thus, driven the interaction towards the calcium-independent binding site. Therefore, it would be interesting as future work to study the selectivity towards each of both Lg ECD binding sites with an hexasaccharide analogue to **Hexa4S** but with the hydroxyl group OH4 (non-reducing end) available to participate in calcium chelation.

4.3 Methodology

4.3.1 Nuclear Magnetic Resonance

All the samples were prepared in 200 μL of 99.9 % D_2O buffer containing 25 mM tris d11 pH 8, 150 mM NaCl and 4 mM $CaCl_2$. Most of the NMR experiments were performed on a Bruker Avance DRX 500 MHz spectrometer equipped with a 5 mm inverse triple-resonance probe, at the *Instituto de Investigaciones Químicas* of the *Consejo Superior de Investigaciones Científicas* (CSIC-US). In the case of heparin-like trisaccharides, the tr-NOESY and STD NMR experiments (only those used for the epitopes mapping) were carried out on a Bruker Digital Avance 800 MHz spectrometer equipped with a 8 mm inverse triple-resonance probe, at the *Laboratorio de RMN de Barcelona* (LRB).

Protein concentrations of 105 (**Tri1-Tri8**), 54 (**Hexa4S**) and 25 μM (**disaccharides**) were employed. On the other hand, ligand concentrations of 1.50 (**Tri1-Tri8** and **disaccharides**) and 1.16 mM (**Hexa4S**) were used.

The STD NMR experiments were carried out with spin-lock and solvent suppression. For the selective saturation, cascades of 49 ms Gaussian pulses and a field strength of 50 Hz were employed, with delay of 1 ms between successive pulses[303]. The on-resonance frequency was set to 6.8 ppm, whereas the off-resonance frequency was set to 40 ppm (blank experiments were performed to assure the absence of direct saturation to the ligand protons). The relaxation delay was properly adjusted for the experiment time length to be constant (generally 6.1 s). Depending on the saturation time, STD NMR experiments were performed with 64/256 (at 800 MHz) or 1k/2k/4k scans (at 500 MHz). More details of each experiment are given in **table 6**.

The STD amplification factors (STD-AF_0) were calculated from the STD initial slopes. To do so, the evolution of the STD-AF with the saturation time (t_{sat}) was fitted to the equation

$$\text{STD-AF}(t) = a(1-\exp(-bt)) \qquad \textit{Eq. 1}$$

Thus, the STD-AF_0 values were obtained as the product of the *ab* coefficients.

Transferred-NOESY experiments were performed using a phase sensitive pulse program with gradient pulses in the mixing time[304] and a relaxation delay of 1.5 seconds. See **table 6** for more details.

Table 6. Summary of the experimental conditions employed in the STD NMR and tr-NOESY experiments performed.

Experiment				*STD-NMR*	*tr-NOESY*
Compound	**Field (MHz)**	**Temp (ºC)**	**prot:lig ratio**	**t_{sat} (s)**	**t_{mix} (ms)**
Tri1-Tri8	800	15	1:14	0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 6	200
HA disaccharides	500	10	1:60 & 1:27	0.5, 1, 2, 3, 4, 5, 6	200, 500
Hexa4S	500	10	1:21	0.5, 1, 1.5, 2, 3, 4, 5	100, 200, 300

Determination of Tri1-Tri8 interglycosidic distances in the bound state *(subchapter 4.1.1)*

Firstly, the ratio between the H2b-H5b and H4b-H5b cross-relaxation rates ($\sigma_{H2b-H5b}/\sigma_{H4b-H5b}$) were calculated in both the free and bound states (see **table 1**). To do so, the H2b-H5b and H4b-H5b normalized volumes (σ_{norm}; cross-peak normalized against its diagonal peak) were divided by the mixing time for comparison purposes between the free and bound states (NOESY and tr-NOESY experiments registered at different mixing times).

On the other hand, the docking-derived distances were first Boltzmann r^{-6} averaged over the best 20 ranked poses, according to the *Emodel* score (see **table 2**). Then, considering that the similarity among the $\sigma_{H2b-H5b}/\sigma_{H4b-H5b}$ ratios obtained indicated alike proportions of the IdoA2S $^1C_4$ and ^2S_O puckers in both the free and bound states, a r^{-6} weighting on the populations of puckers of the iduronate ring at 278K[76] (see *Chapter 3*) was done. Note that the populations at 278K were taken for the weighting because at this temperature were registered the NOESY experiments of the free ligands from which the cross-relaxation rates have been compared to those of the bound state. The final values were obtained as $\langle r^{-6} \rangle^{-1/6}$. Also note that the H1'-H6 docking-derived distances represent the r^{-6} averages over the *proR* and *proS* values, in agreement with the well-known methodology for proton to methylene distance averaging provided that the methylene internal motion is significantly slower than the overall tumbling motion[305].

The experimental distances of the ligand in the bound state were derived from tr-NOESY experiments at 200 ms mixing time (see **table 2**). First, each cross-peak was divided by its corresponding diagonal peak, thus obtaining the normalized NOE volume or σ_{NOE}. Then, considering the *Isolated Spin Pair*

Approximation (ISPA)[261] and using the constant Glc(C) H1-H2 distance as reference (2.5 Å), the key inter-glycosidic proton-proton distances were calculated according to the equation :

$$d_x = d_{ref} \cdot \left({}^{\sigma_{ref}}/_{\sigma_x}\right)^{1/6} \qquad Eq.\ 2$$

where d_x is the unknown distance to be determined, d_{ref} is the distance used as reference, and σ_{ref} and σ_x are the cross-relaxation rates of the reference and unknown distances, respectively.

4.3.2 Modelling

Heparin-like trisaccharides interactions with Lg ECD (subchapter 4.1.1)

Refining the receptor conformation for docking: Induced Fit Docking (IFD)[306]

The starting structure of Langerin CRD was taken from the crystal structure with PDB code 3P5H, which contains four monomers of the CRD of Langerin complexed to laminaritriose. Monomer C was chosen for presenting the better resolved tertiary structure. Furthermore, the backbone RMSd for monomers A, B and D (with monomer C being the reference structure) was calculated, obtaining values below 0.4.

For the preparation of the docking starting structures, the crystal water molecules were first removed. Then, we placed the trisaccharide model Tri1 (with the iduronte ring in $^1C_4$ conformation) in the calcium dependent binding site of Langerin by superimposing the backbone of its non-reducing end GlcN residue to the Glc residue of laminaritriose (3P5H) and prepared the whole structure with PrepWiz tool of Maestro. Thus, the essential zero-order bonds to the calcium ion were created and the hydrogen bond network optimized. Next, Induced Fit Docking[306] was run. To do so, the grid was first generated by defining the centroid of the ligand as the center of the box, and a box size of 30Å. Then, a constrained minimization of the receptor was carried out followed by flexible docking standard precision (SP) without sampling the conformation of the rings, and penalizing the non-planar conformation of the amide bonds. The best 20 poses were then selected for the refinement of the side chains within 6Å of the ligand, followed by flexible docking SP of the receptor structures within 30 kcal/mol of the lowest-energy structure. (Note: glide docking standard precision is more convenient when docking poor binders than the extra precision algorithm, XP, because poor binders are more likely to encounter penalty terms in glide XP). The structure of the complex matching the O3-O4 calcium coordination requirement and presenting the lowest *glide energy* value (sum of Van der Waals and Coulomb contributions), -69,6 kcal/mol, was chosen for the subsequent docking jobs with all the ligands (**figure 21**). It should be highlighted that this new IFD-refined receptor structure does not present any significant conformational change when it is compared to the crystal structure as it is shown by their backbone RMSD value of 0.37 (**figure 22**), but it does refine the orientation of the binding site side chains.

Conformational search

Monte Carlo Multiple Minimum (MCMM) torsional sampling was carried out for each of the 16 ligand models, keeping the initial conformation of the iduronate ring ($^1C_4$ and ^2S_O) by imposing torsional restraints (force constant of 200 kcal/mol) on its ring torsions C1-C2-C3-C4, C4-C5-O5-C1, O5-C1-C2-C3 and C3-C4-C5-O5. The force field OPLS2005, Macromodel implicit water model and an infinite cutoff for the non-bonded interactions were used. A thousand steps of conformational searching per rotatable bond were accomplished and redundant conformers were eliminated using an RMSD cutoff of 0.5Å. During the minimization step, the Truncated Newton Conjugate Gradient method was applied, using a maximum of a thousand iterations. Then, the conformers generated were RMSD clustered using the average-linkage algorithm[307] and those nearest the centroid of each cluster where taken. Among them, the ten lowest energy conformers for each ligand were selected for docking.

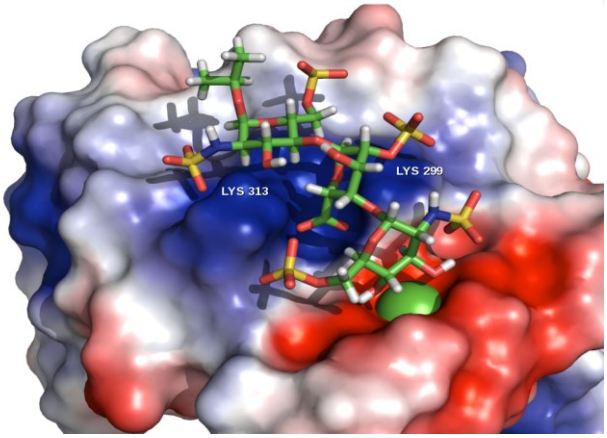

Figure 21. Lowest *glide energy* pose for Tri1, with the IdoA2S ring in the $^1C_4$ puckered conformation, obtained by Induced Fit Docking. Note: the pose of the ligand in the selected Induced Fit Docking structure matches what its later referred as the Main Binding Mode.

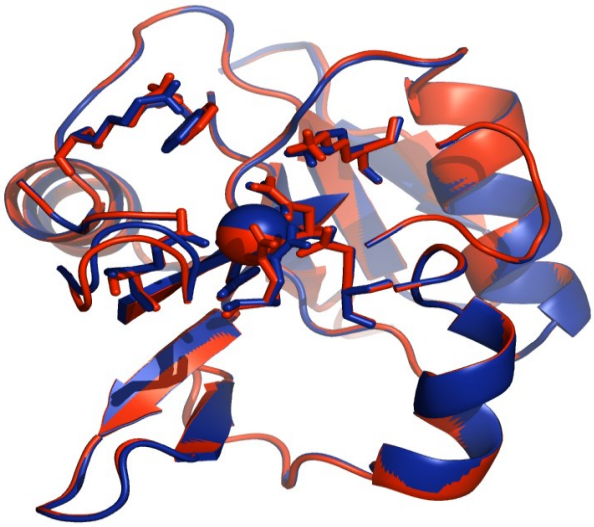

Figure 22. Crystal (red) and IFD (blue) structures of Langerin CRD, superimposed. Backbone RMSd = 0.37.

Grid generation

With Glc(C) residue of **Tri1** adopting the binding mode of laminaritriose bound to Lg CRD (PDB code 3P5H[148b]), a 30Å-sided cube grid from the centroid of the ligand was generated. Inside, a subgrid of 10Å-side from the geometric center of ligand was defined so that the centroid of the output docking solutions had to fall into it.

Flexible docking

The ten lowest energy conformers of each ligand model obtained from MCMM torsional sampling were submitted to *Glide Flexible Docking SP*[246]. Ring conformations sampling was turned off during docking. On the contrary, nitrogen inversions sampling and penalization of non-planar conformations for amide-type torsions were turned on. A distance-dependent dielectric constant of 4 was used and post-docking minimization was performed.

Eventually, the several thousand of poses obtained were filtered according to the calcium-coordination requirement. Then, the best 50 poses for each ligand were selected and backbone RMSD clustered using the average-linkage algorithm[307]. The poses nearest the centroid of each cluster were considered to be the most representative, and thus, they were used for the analysis and discussion.

Molecular Dynamics

Input preparation

The initial coordinates for molecular dynamics were obtained after superimposing the best docking solution for Langerin CRD with Tri1 to each of the three CRDs of the trimeric structure of Lg ECD (PDB code: 3KQG), using VMD 1.9[308]. The topology and coordinates files were built with the *tLEAP* module of AMBER 11[294] package. The system was neutralized with calcium ions and then immersed in a TIP3P[264] water box, giving rise to a molecular system of 81967 atoms. The GLYCAM_06h[298] parameters were used to model the sugar moiety, including the sulfate and sulphamate groups, and the AMBER ff12SB parameters[300] for the protein part, water molecules and calcium ions. Furthermore, the partial charges of GLYCAM06 were employed for the sugar moiety, adjusting the partial charge on the O- and N- atoms bound to the SO3 groups according to GLYCAM philosophy for charge development. For the O-isopropyl group, partial charges were derived from the molecular electrostatics potential (MEP) using the RESP method[295] with a constraint of 0.01, for consistency with the procedure employed in GLYCAM06 force field[298] development. The HF/6-31G* level of theory was used for both the structure optimization and the MEP calculation. Detailing the procedure employed, a methyl-O-isopropyl and a D-Glc-OMe were built and charge constraints imposed as follows: the total charge of both molecular models is set to 0, the methyl group in the D-Glc-OMe must have a charge of +0.194 whereas the charge of the oxygen involved in the glycosidic linkage is set to -0.194, and both methyl groups are set to be equivalent and to be removed during the last step of charge derivation, and, thus, being both compounds merged to form a D-Glc-O-Isopropil. Additionally, the

partial charges on aliphatic hydrogens and on the O-isopropyl group were constrained to 0 and -0.194, respectively, in agreement with GLYCAM philosophy. The standard error and relative root mean square error were, respectively, 0.005 and 0.188. It is noticeable that a similar protocol has been successfully applied for the development of parameters for different sugar derived compounds[296, 299]. The quantum mechanical calculations and the RESP procedure were carried out with ante-R.E.D 2.0 and R.E.D IV of the R.E.D web server[297].

MD of the LgECD-Tri1 trimeric complex

The molecular dynamics simulations have been run on Finis Terrae HPC, which belongs to the *Centro de Supercomputación de Galicia* (CESGA), Spain, taking advantage of the prioritized computing time awarded (ICTS CESGA 2012 second call). They were carried out with AMBER 11[294]. The equilibration protocol consisted of an initial minimization of the water box (100000 steps), followed by a minimization of the whole system (50000 steps). Then, the system was slowly heated (100000 steps) at constant volume until 300 K using a time constant for the heat bath coupling of 1 ps. The equilibration finished with 200 picoseconds of molecular dynamics simulation without restraints, at constant pressure (1bar), applying Periodic Boundary Conditions (PBC) and using the Particle Mesh Ewald Method[282] (PME) to account for the long range electrostatic effect (the grid spacing was approximately 1 Å). The SHAKE algorithm[302] was also employed, thus, allowing a 2 fs time step, and non-bonded interactions were cutoff at 8.0 Å and updated every 25 steps. The production dynamics simulations have been accomplished at a constant temperature of 300 K, by applying the Langevin thermostat[301] with a collision frequency of 2 ps^{-1}, and at constant pressure (1 bar). PME and PBC conditions, the SHAKE algorithm and a 8 Å cutoff have also been employed. During both the last step of the equilibration protocol and the production dynamics, two instantaneous distance restraints, between the calcium ion and the atoms O3 and O4 of the non-reducing end GlcN residue, of 20 kcal/mol, have been imposed to match a maximum distance of 3.2 Å. The equilibration protocol and production dynamics have been performed with the *sander.MPI* and *pmemd.MPI* modules of AMBER 11[294], respectively. One MD simulation of 100 ns have been performed, saving the trajectory coordinates each picosecond. The processing of the MD trajectory has been done with the *ptraj* module of AMBER 11[294]. Specifically, the RMSd per ligand residue with respect to the protein binding site was calculated. To do so, a first protein backbone RMSd fit was done to know when the system started to be in equilibrium. Then, for each of the three monomer units of Lg ECD structure, a protein backbone RMSd fit within 5Å distance from the ligand was applied to the equilibrated part of the trajectory, followed by a RMSd (no fit) per sugar backbone residue (see *Appendix*).

HA disaccharides interactions with Lg ECD (subchapter 4.1.2)

Input structures

The input coordinates of both hyaluronan-like disaccharides were taken from the NMR and molecular modeling resolved average structure of hyaluronan in solution (PDB code 2BVK[95]). The 3-azidopropyl glycoside substituent of the reducing terminal was built with the graphical interface Maestro of

Schrödinger Suite 2012, in a *gauche* orientation with respect to proton H1 of the reducing residue in order to minimize repulsive interactions.

On the other hand, the starting structure of Langerin CRD was taken from the crystal structure with PDB code 3P5H[148b], which contains four monomers of the CRD of Langerin complexed to laminaritriose (no published structure of Langerin bound to hyaluronan-like oligosaccharide exits). Monomer C was chosen for presenting 1) the better resolved tertiary structure and 2) the ligand laminaritriose coordinated to the calcium ion through oxygens O3 and O4 of the non-reducing glucose residue, with the latter facilitating the setup of an appropriate starting disposition of the disaccharide ligands in the binding site for the grid generation.

For the preparation of the docking starting structure, the crystal water molecules were first removed. Next, each disaccharide was placed in the calcium dependent binding site of Langerin by superimposing the backbone of its non-reducing end residue on the non-reducing glucose residue of laminaritriose (3P5H). Then, with *PrepWiz* tool of Maestro, zero-order bonds to the calcium ion (involving the amino acid residues of the EPN and WND motifs) were created, the protein hydrogen bond network optimized and the ligand-receptor complex structure refined with a restrained minimization (*impref* utility).

Conformational search

Monte Carlo Multiple Minimum (MCMM) torsional sampling was carried out for the interacting ligand disaccharide 2, keeping fixed the $^4C_1$ conformation of the sugar units by imposing intra-ring torsional restraints (force constant of 200 kcal/mol) on the torsions C1-C2-C3-C4, C4-C5-O5-C1, O5-C1-C2-C3 and C3-C4-C5-O5. The force field OPLS_2005, Macromodel implicit water model and an infinite *cutoff* for the non-bonded interactions were used. A thousand steps of conformational searching per rotatable bond were accomplished and redundant conformers were eliminated using an RMSD *cutoff* of 0.5Å. During the minimization step, the Truncated Newton Conjugate Gradient method was applied, using a maximum of a thousand iterations. 262 conformers were obtained for disaccharide 2.

Grid generation

With disaccharide 2 adopting a laminaritriose-type orientation, a 30Å-sided cube grid from the centroid of the ligand was generated. Inside, a subgrid of 10Å-side from the geometric center of ligand was defined so that the centroid of the output docking solutions had to fall into it.

Flexible docking

The set of conformers obtained in the conformational search for disaccharide 2 were sourced to a *glide* flexible docking standard precision (SP, version 5.0)[246] calculation, using expanded sampling during the selection of the initial poses. For the energy minimization step, a value of 4.0 was employed as

distance-dependent dielectric constant and the maximum number of minimization steps was set to a thousand. Furthermore, whereas the sampling of the ring conformation was turned off during docking, nitrogen inversions sampling and penalization of non-planar conformations for amide-type torsions were turned on. In addition, the van der Waals radii of the non-polar ligand atoms were scaled to a 0.5 scaling factor. Finally, a minimization step after docking was performed. Per-residue interaction scoring was turned on for residues within 30Å of grid center.

The docked poses obtained (60000) were first filtered according to the distances between the calcium ion and oxygen atoms O3 and O4 of the non-reducing terminal, dismissing those poses with any of these distances beyond 3 Å. Next, those poses violating the previously reported allowed geometry (ϕ and ψ torsions) for the glycosidic linkages of hyaluronan[95], were also rejected. With the 1600 poses left, an atomic RMSd clustering (in place, no superposition) with the set of atoms O3. C3, C4 and O4 of the non-reducing terminal (using the average-linkage algorithm) was carried out to separate the different relative orientations of the sugar with respect to the calcium ion. As a result, two main binding modes were observed. Finally, a torsional RMSd clustering of conformers (with superimposition) was carried out with the set of torsions involving the orientation of the 3-azidopropyl glycoside substituent relative to the reducing sugar residue and the ϕ and ψ dihedral angles. The average-linkage algorithm was also used. The set of conformers nearest the centroid of each cluster (42) were considered for the analysis and discussion. Glide version 5.8 (Schrödinger Suite 2012) was used.

Refine docking

Both disaccharides, adopting the initial laminaritriose-type orientation, were submitted to a *glide refine* calculation (do not dock) consisting of an optimization of the ligand structure in the field of the receptor, followed by the scoring of the ligand. For the energy minimization, a value of 4.0 was employed as distance-dependent dielectric constant and the maximum number of minimization steps was set to a thousand. Per-residue interaction scoring was turned on for residues within 30Å of grid center.

Capítulo 5

Conclusiones

Las características estructurales de oligosacáridos tipo heparina de diferentes tamaños y patrones de sustitución han sido estudiadas en profundidad, en disolución, mediante dinámica molecular (DM) y Resonancia Magnética Nuclear (RMN). A partir de los resultados obtenidos, hemos llegado a las siguientes conclusiones:

- El equilibrio conformacional del anillo de L-IdoA2S presente en los compuestos **Tri1-Tri8** viene exclusivamente descrito por las conformaciones del polo norte ($^1C_4$) y ecuador (^2S_O, $^{2,5}B$, $^5S_1$, etc.) de la esfera de Cremer-Pople, siendo el conformero ^2S_O el que predomina entre sus análogos ecuatoriales.

- El equilibrio conformacional de anillos de iduronato contenidos en derivados de heparina (L-IdoA2S) se desplaza, independientemente de la temperatura, hacia los confórmeros ecuatoriales cuando el residuo de glucosamina del extremo reductor se encuentra sulfatado en su posición 6 (**Tri1-Tri4**). Por otro lado, la substitución de esta voluminosa y cargada función sulfato por un grupo neutro como el hidroxilo (**Tri5-Tri8**) da lugar a que las poblaciones de confórmeros ecuatoriales experimenten una marcada sensibilidad con la temperatura. Además, en éste caso el equilibrio conformaciónal se desplaza hacia la silla $^1C_4$, aunque sólo a baja temperatura, disminuyendo las diferencias de poblaciones entre ambas series conforme la temperatura aumenta (hasta desaparecer a 313 K). Estos resultados pueden ser extrapolados a residuos de L-IdoA2S internos presentes en oligosacáridos de mayor tamaño.

- La sulfatación en posición 6 del residuo de GlcN del extremo reductor (**Tri1-Tri4**) parece promover una considerable flexibilidad adicional sobre la torsión H2-C2-C3-H3 del anillo de L-IdoA2S, lo cual se correlaciona con la mayor tendencia de los compuestos **Tri1-Tri4** a poblar los estados conformacionales ecuatoriales.

- Llevar a cabo simulaciones de dinámica molecular con restricciones ponderadas en el tiempo (*tar*-MD) usando la distancia intraanular H2-H5 del residuo de L-IdoA2S como única restricción experimental ha resultado ser una aproximación de elevada precisión y sencillez a la hora de describir tanto el equilibrio conformacional del anillo de iduronato en derivados de heparina, como la influencia que los patrones de sulfatación de los residuos de GlcN adyacentes ejercen sobre él. Para los compuestos **Tri1-Tri8**, esta metodología ha indicado que los confórmeros ecuatoriales dominan el equilibrio conformacional del anillo de ácido idurónico cuando la glucosamina del extremo terminal se encuentra sulfatada en posición 6.

- La estructura global de los trisacáridos **Tri1-Tri8** viene caracterizada por conformaciones *syn* rígidas y flexibles alrededor de las uniones glicosídicas GlcN-IdoA2S e IdoA2S-GlcN, respectivamente, la última de las cuales adopta conformaciones tipo *anti*-Ψ (generalmente minoritarias) únicamente en algunos compuestos. Notablemente, las intensidades de los NOEs han indicado contribuciones muy similares de las conformaciones *syn*-Ψ y *anti*-Ψ alrededor de la unión IdoA2S-GlcN para la secuencia GlcNS-IdoA2S-GlcNAc (**Tri8**).

- La eliminación del grupo sulfato en posición 6 del extremo reductor (**Tri5-Tri8**) da lugar a la disminución del tiempo de correlación global, lo cual se traduce en una reorientación más rápida en disolución. Comparativamente, dicha sustitución conlleva un mayor impacto sobre el tiempo de correlación global que el resto de sitios de sulfatación presentes en los compuestos **Tri1-Tri8**.

- La conformación del hexasacárido tipo heparina **Hexa3** presenta las mismas características que los hexa- y oligosacáridos de mayor tamaño estudiaos previamente, es decir, una conformación global tipo hélice lineal, un esqueleto rígido a nivel global, plasticidad conformacional de los residuos de iduronato internos, así como un comportamiento hidrodinámico anisotrópico.

- Las afinidad previamente observada para la interacción del **Hexa3** con FGF-1, significativamente menor en relación a los compuestos **Hexa1** y **Hexa2**, se debe a la combinación de su conformación y secuencia, de forma que **Hexa3** (a diferencia de los otros dos hexasacáridos) puede únicamente interaccionar con el sub-sitio de reconocimiento principal de la proteína, dejando el secundario sin ocupar. Esto conlleva una disminución en la intensidad de la interacción.

- La casi inexistente capacidad de la secuencia **Hexa3** para inducir actividad mitogénica mediada por FGF-1 parece estar basada en su considerablemente más baja afinidad por esta proteína, lo cual se traduciría en una menor probabilidad de que el factor FGF-1 genere conformaciones accesibles para interaccionar favorablemente con su receptor y así formar el complejo ternario que desencadena la respuesta biológica.

Además, las características que gobiernan las propiedades de las interacciones de diferentes oligosacáridos tipo heparina y ácido hialurónico (HA) con una lectina tipo C implicada en el reconocimiento del virus VIH, Langerina (Lg), han sido investigadas desde el punto de vista estructural mediante técnicas de RMN basadas en observación de ligando, como STD NMR, simulaciones de dinámica molecular y docking computacional. A partir del análisis de los resultados obtenidos, podemos concluir que:

- La interacción del dominio extracelular de Langerina (Lg ECD) con trisacáridos tipo heparina (**Tri1-Tri8**) y disacáridos tipo HA (**disacáridos 1** y **2**) es dependiente de calcio, y tiene lugar en el sitio de reconocimiento de calcio tipo 2 del dominio de reconocimiento de carbohidratos (CRD) de la proteína.

- La interacción de Lg ECD con trisacáridos tipo heparina (**Tri1-Tri8**) es independiente del patrón de sulfatación variable, es decir, tanto el modo como el epítopo de la interacción son equivalentes para los 8 trisacáridos. Sin embargo, el patrón de sulfatación constante

(N-sulfato del terminal no reductor y 2-sulfato del anillo de iduronato) presente en toda la librería **Tri1-Tri8**, juega un papel clave en la definición del modo de interacción.

- La interacción Lg ECD – trisacáridos (**Tri1-Tri8**) tiene lugar sin selección conformacional. Así, las características estructurales observadas para los 8 trisacáridos, tanto a nivel global (geometría de los enlaces glicosídicos) como local (poblaciones de confórmeros del residuo de L-IdoA2S), son aproximadamente iguales en estado libre y enlazado.

- El modo de interacción de los compuestos **Tri1-Tri8** representa una conformación extendida a través de las cadenas laterales de las lisinas 299 y 313, de forma que cada uno de los residuos del ligando establece contactos con la superficie de la proteína. Las características estructurales principales de la interacción son 1) la coordinación al calcio a través de los oxígenos O3 y O4 del residuo de GlcN del extremo no reductor, 2) la formación del puente salino LYS299-COO^-(L-IdoA2S)-LYS313, y 3) la ocupación del sitio de interacción tipo A para calcio (de la heparina) por la cadena lateral del amino ácido LYS299.

- Los disacáridos de HA son únicamente capaces de interaccionar con el domino de reconocimiento de carbohidratos de Langerina a través de la quelación del ion calcio por los átomos de oxígeno O3 y O4 de un residuo no reductor de GlcNAc, sin que tengan lugar cambios conformacionales durante el reconocimiento molecular (**disacárido 2**). De acuerdo con los modelos de docking, dos modos de interacción, que además contribuyen de forma similar al estado enlazado, son posibles: uno análogo al de la β-D-glucosa en el estado cristalino (3P5H), y el otro con igual geometría que el primero pero con su orientación rotada 180º respecto al eje molecular principal. En ambos modos de interacción, el ligando se orienta de forma perpendicular respecto a la superficie del sitio de reconocimiento, aumentando así la exposición del ligando al disolvente en sentido terminal no reductor – terminal reductor.

- La secuencia de ácido hialurónico GlcA-GlcNAc (**disacárido 1**) no es capaz de interaccionar en el dominio de reconocimiento de carbohidratos de Langerina, lo cual se debe, de acuerdo a los estudios de docking, a que en el estado enlazado tendría lugar una interacción repulsiva altamente desestabilizante entre grupo carboxilato del anillo de glucurónico y una de las cadenas laterales de los amino ácidos GLU285 o GLU293.

- El compuesto **Hexa4S** interacciona en el sitio de reconocimiento independiente de calcio situado en la interfaz de trimerización de Lg ECD, del mismo modo que lo hace la heparina de alto peso molecular. Por tanto, la sustitución de un oligosacárido tipo heparina de pequeño tamaño (trisacárido) por uno de tamaño hexasacarídico modifica la selectividad de la interacción con Lg ECD, cambiando el sitio de reconocimiento dependiente de calcio (trisacárido) por el independiente de calcio (hexasacárido).

References

[1] a) T. W. Rademacher, R. B. Parekh and R. A. Dwek, *Annual Review of Biochemistry* **1988**, *57*, 785-838; b) Y. C. Lee and R. T. Lee, *Accounts of Chemical Research* **1995**, *28*, 321-327; c) H. J. Gabius, H. C. Siebert, S. Andre, J. Jimenez Barbero and H. Rudiger, *ChemBioChem* **2004**, *5*, 740-764.
[2] J. E. Turnbull and R. A. Field, *Nat Chem Biol* **2007**, *3*, 74-77.
[3] R. Raman, S. Raguram, G. Venkataraman, J. C. Paulson and R. Sasisekharan, *Nature Methods* **2005**, *2*, 817-824.
[4] A. A. Weiss and S. S. Iyer, *Microbe* **2007**, *2*, 489-497.
[5] T. K. Lindhorst, *Wiley-VCH* **2007**.
[6] a) K. N. Kirschner and R. J. Woods, *Proc Natl Acad Sci U S A* **2001**, *98*, 10541-10545; b) F. Corzana, M. S. Motawia, C. H. Du Penhoat, S. Perez, S. M. Tschampel, R. J. Woods and S. B. Engelsen, *J Comput Chem* **2004**, *25*, 573-586; c) F. Corzana, M. S. Motawia, C. Herve du Penhoat, F. van den Berg, A. Blennow, S. Perez and S. B. Engelsen, *J Am Chem Soc* **2004**, *126*, 13144-13155.
[7] J. P. B. Kamerling, G., *Elsevier: Amsterdam* **2007**.
[8] G. Boons, *Blackie Academic & Professional: London* **1998**.
[9] J. F. Stoddart, *Wiley-Interscience: New-York* **1971**.
[10] D. R. Ferro, A. Provasoli, M. Ragazzi, G. Torri, B. Casu, G. Gatti, J. C. Jacquinet, P. Sinaÿ, M. Petitou and J. Choay, *Journal of the American Chemical Society* **1986**, *108*, 6773-6778.
[11] V. S. Raghavendra Rao, *Harwood Academic Publishers* **1998**.
[12] P. Vidal, B. Vauzeilles, Y. Blériot, M. Sollogoub, P. Sinaÿ, J. Jiménez-Barbero and J. F. Espinosa, *Carbohydrate Research* **2007**, *342*, 1910-1917.
[13] a) G. A. Jeffrey, J. A. Pople and L. Radom, *Carbohydrate Research* **1972**, *25*, 117-131; b) G. A. Jeffrey, J. A. Pople and L. Radom, *Carbohydrate Research* **1974**, *38*, 81-95; c) G. A. Jeffrey, J. A. Pople, J. S. Binkley and S. Vishveshwara, *Journal of the American Chemical Society* **1978**, *100*, 373-379.
[14] G. A. Jeffrey and J. H. Yates, *Carbohydrate Research* **1979**, *74*, 319-322.
[15] R. A. Laine, *Glycobiology* **1994**, *4*, 759-767.
[16] A. K. Dunker, I. Silman, V. N. Uversky and J. L. Sussman, *Curr Opin Struct Biol* **2008**, *18*, 756-764.
[17] B. M. Sattelle, S. U. Hansen, J. Gardiner and A. Almond, *Journal of the American Chemical Society* **2010**, *132*, 13132-13134.
[18] I. Tvaroška and S. Pérez, *Carbohydrate Research* **1986**, *149*, 389-410.
[19] A. Varki, *Cold Spring Harbor Laboratory Press, New York* **1999**.
[20] H. Iha and M. Yamada, *Glycan Profiling of Adult T-Cell Leukemia (ATL) Cells with the High Resolution Lectin Microarrays*, **2013**, p.
[21] B. Lepenies, J. Yin and P. H. Seeberger, *Current Opinion in Chemical Biology* **2010**, *14*, 404-411.
[22] I. Cumpstey, *Carbohydrate Research* **2009**, *344*, 2285-2310.
[23] E. Fernandez-Megia, J. Correa, I. Rodríguez-Meizoso and R. Riguera, *Macromolecules* **2006**, *39*, 2113-2120.
[24] C. C. Lee, J. A. MacKay, J. M. J. Fréchet and F. C. Szoka, *Nature Biotechnology* **2005**, *23*, 1517-1526.
[25] B. Weinhold, R. Seidenfaden, I. Röckle, M. Mühlenhoff, F. Schertzinger, S. Conzelmann, J. D. Marth, R. Gerardy-Schahn and H. Hildebrandt, *Journal of Biological Chemistry* **2005**, *280*, 42971-42977.
[26] L. Jin, J. P. Abrahams, R. Skinner, M. Petitou, R. N. Pike and R. W. Carrell, *Proceedings of the National Academy of Sciences of the United States of America* **1997**, *94*, 14683-14688.
[27] R. S. Haltiwanger and J. B. Lowe in *Role of glycosylation in development*, *Vol. 73* **2004**, pp. 491-537.
[28] R. W. Sanders, M. Venturi, L. Schiffner, R. Kalyanaraman, H. Katinger, K. O. Lloyd, P. D. Kwong and J. P. Moore, *J Virol* **2002**, *76*, 7293-7305.
[29] A. Almond and J. K. Sheehan, *Glycobiology* **2000**, *10*, 329-338.
[30] K. Karaveg, A. Siriwardena, W. Tempel, Z. J. Liu, J. Glushka, B. C. Wang and K. W. Moremen, *Journal of Biological Chemistry* **2005**, *280*, 16197-16207.
[31] a) A. Varki, *Glycobiology* **1993**, *3*, 97-130; b) R. A. Dwek, *Chemical Reviews* **1996**, *96*, 683-720.
[32] I. W. Sutherland, *Trends in Biochemical Sciences* **1979**, *4*, 55-59.

[33] B. H. A. Rehm, *Nature Reviews Microbiology* **2010**, *8*, 578-592.
[34] a) E. S. Jeremiah and S. Geetha, *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **2002**, *54*, 177-186; b) K. Sugahara and H. Kitagawa, *Iubmb Life* **2002**, *54*, 163-175.
[35] L. F. James, *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **2002**, *54*, 187-194.
[36] I. Naoki and K. Koji, *IUBMB Life (International Union of Biochemistry and Molecular Biology: Life)* **2002**, *54*, 195-199.
[37] a) A. Varki and Varki, *Glycobiology* **1993**, *3*, 97-130; b) H.-J. Gabius, S. Andre, J. Jimenez Barbero, A. Romero and D. Solis, *Trends in biochemical sciences* **2011**, *36*, 298-313; c) H. J. Gabius, H. C. Siebert, S. Andre, J. Jimenez-Barbero and H. Rudiger, *Chembiochem* **2004**, *5*, 740-764.
[38] a) B. Casu and U. Lindahl, *Advances in Carbohydrate Chemistry and Biochemistry, Vol 57* **2001**, *57*, 159-206; b) D. J. Hamel, I. Sielaff, A. E. I. Proudfoot, T. M. Handel, M. H. Tracy and J. H. Damon in *Chapter 4 Interactions of Chemokines with Glycosaminoglycans*, *Vol. Volume 461* Academic Press, **2009**, pp. 71-102; c) U. Lindahl and J.-p. Li in *INTERACTIONS BETWEEN HEPARAN SULFATE AND PROTEINS-DESIGN AND FUNCTIONAL IMPLICATIONS*, *Vol. 276* (Ed. K. W. Jeon), **2009**, pp. 105-159; d) J. Jimenez-Barbero, M. D. Diaz and P. M. Nieto, *Anti-Cancer Agents in Medicinal Chemistry* **2008**, *8*, 52-63.
[39] J. R. Bishop, M. Schuksz and J. D. Esko, *Nature* **2007**, *446*, 1030-1037.
[40] M. Mohammadi, S. K. Olsen and O. A. Ibrahimi, *Cytokine & Growth Factor Reviews* **2005**, *16*, 107-137.
[41] T. M. Handel, Z. Johnson, S. E. Crown, E. K. Lau, M. Sweeney and A. E. Proudfoot, *Annual Review of Biochemistry* **2005**, *74*, 385-410.
[42] W. Li, D. J. D. Johnson, C. T. Esmon and J. A. Huntington, *Nature Structural & Molecular Biology* **2004**, *11*, 857-862.
[43] J. M. Whitelock and R. V. Iozzo, *Chemical Reviews* **2005**, *105*, 2745-2764.
[44] R. Sasisekharan, R. Raman and V. Prabhakar, *Annual Review of Biomedical Engineering* **2006**, *8*, 181-231.
[45] a) B. Casu in *Structure and Biological Activity of Heparin*, *Vol. 43* **1985**, pp. 51-134; b) W. D. Comper, *Heparin (and related polysaccharides): structural and functional properties*, Gordon and Breach Science Publishers, **1981**, p.
[46] In this thesis, the common termssulphate and carboxylate are used instead of sulfonato and carboxylato as recommended by the IUPAC publication Nomenclature of Carbohydrates.
[47] a) R. J. T. Linhardt, T., *Marcel Dekker, New Yok* **1997**, 277-341; b) H. E. Conrad, *Heparin Binding Proteins*, Academic Press, San Diego, California, **1988**, p.
[48] D. A. Lane and U. Lindahl, *Heparin: Chemical and Biological Properties, Clinical Applications*, CRC Press, **1989**, p.
[49] J. T. Gallagher, J. E. Turnbull and M. Lyon, *Adv Exp Med Biol* **1992**, *313*, 49-57.
[50] C. C. Griffin, R. J. Linhardt, C. L. Van Gorp, T. Toida, R. E. Hileman, R. L. Schubert Ii and S. E. Brown, *Carbohydrate Research* **1995**, *276*, 183-197.
[51] M. Bernfield, R. Kokenyesi, M. Kato, M. T. Hinkes, J. Spring, R. L. Gallo and E. J. Lose, *Annual Review of Cell Biology* **1992**, *8*, 365-393.
[52] U. Lindahl, K. Lidholt, D. Spillmann and L. Kjellen, *Thrombosis Research* **1994**, *75*, 1-32.
[53] M. Bernfield, M. Götte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum and M. Zako in *Functions of cell surface heparan sulfate proteoglycans*, *Vol. 68* **1999**, pp. 729-777.
[54] R. J. Linhardt, *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]* **2001**, *Chapter 17*.
[55] I. Capila and R. J. Linhardt, *Angewandte Chemie - International Edition* **2002**, *41*, 391-412.
[56] a) A. Imberty and S. Perez, *Chemical Reviews* **2000**, *100*, 4567-4588; b) W. Tempel, S. Tschampel and R. J. Woods, *J Biol Chem* **2002**, *277*, 6615-6621; c) N. K. Vyas, M. N. Vyas, M. C. Chervenak, M. A. Johnson, B. M. Pinto, D. R. Bundle and F. A. Quiocho, *Biochemistry* **2002**, *41*, 13575-13586.
[57] a) H. Kogelberg, D. Solís and J. Jiménez-Barbero, *Curr Opin Struct Biol* **2003**, *13*, 646-653; b) C. Landersjö and G. Widmalm, *Biopolymers* **2002**, *64*, 283-291; c) E. W. Sayers and J. H. Prestegard,

Biophysical Journal **2002**, *82*, 2683-2699; d) M. R. Wormald, A. J. Petrescu, Y. L. Pao, A. Glithero, T. Elliott and R. A. Dwek, *Chemical Reviews* **2002**, *102*, 371-386.
[58] K. G. Rice, P. Wu, L. Brand and Y. C. Lee, *Biochemistry* **1991**, *30*, 6646-6655.
[59] J. F. G. W. Vliegenthart, R.J., *NMR Spectroscopy and Computer Modeling of Carbohydrates: Recent Advances*, **2006**, p. 235-257.
[60] E. Wrenn Wooten, C. J. Edge, R. Bazzo, R. A. Dwek and T. W. Rademacher, *Carbohydrate Research* **1990**, *203*, 13-17.
[61] F. H. Cano, C. Foces-Foces, J. Jiménez-Barbero, A. Alemany, M. Bernabé and M. Martín-Lomas, *Journal of Organic Chemistry* **1987**, *52*, 3367-3372.
[62] J. Gonzalez-Outeirino, R. Kadirvelraj and R. J. Woods, *Carbohydr Res* **2005**, *340*, 1007-1018.
[63] B. L. Foley, M. B. Tessier and R. J. Woods, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 652-697.
[64] R. J. Linhardt, *Journal of Medicinal Chemistry* **2003**, *46*, 2551-2564.
[65] B. Mulloy, M. J. Forster, C. Jones and D. B. Davies, *Biochemical Journal* **1993**, *293*, 849-858.
[66] a) B. Mulloy and R. J. Linhardt, *Current Opinion in Structural Biology* **2001**, *11*, 623-628; b) B. Mulloy and M. J. Forster, *Glycobiology* **2000**, *10*, 1147-1156.
[67] J. Angulo, R. Ojeda, J. L. De Paz, R. Lucas, P. M. Nieto, R. M. Lozano, M. Redondo-Horcajo, G. Giménez-Gallego and M. Martín-Lomas, *ChemBioChem* **2004**, *5*, 55-61.
[68] D. R. Ferro, A. Provasoli, M. Ragazzi, G. Torri, B. Casu, G. Gatti, J. C. Jacquinet, P. Sinay, M. Petitou and J. Choay, *Journal of the American Chemical Society* **1986**, *108*, 6773-6778.
[69] a) D. R. Ferro, A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossennec, B. Perly, P. Sinay, M. Petitou and J. Choay, *Carbohydrate Research* **1990**, *195*, 157-167; b) M. Ragazzi, D. R. Ferro, B. Perly, G. Torri, B. Casu, P. Sinay, M. Petitou and J. Choay, *Carbohydrate Research* **1987**, *165*, C1-C5.
[70] B. Casu, M. Petitou, M. Provasoli and P. Sinaÿ, *Trends in Biochemical Sciences* **1988**, *13*, 221-225.
[71] I. Capila and R. J. Linhardt, *Angewandte Chemie - International Edition* **2002**, *41*, 390-412.
[72] P. N. Sanderson, T. N. Huckerby and I. A. Nieduszynski, *Biochemical Journal* **1987**, *243*, 175-181.
[73] a) M. Ragazzi, D. R. Ferro and A. Provasoli, *Journal of Computational Chemistry* **1986**, *7*, 105-112; b) M. J. Forster and B. Mulloy, *Biopolymers* **1993**, *33*, 575-588; c) Y. Kurihara and K. Ueda, *Carbohydrate Research* **2006**, *341*, 2565-2574.
[74] a) D. A. Rees, E. R. Morris, J. F. Stoddart and E. S. Stevens, *Nature* **1985**, *317*, 480; b) B. Casu, J. Choay, D. R. Ferro, G. Gatti, J. C. Jacquinet, M. Petitou, A. Provasoli, M. Ragazzi, P. Sinay and G. Torri, *Nature* **1986**, *322*, 215-216; c) S. K. Das, J. M. Mallet, J. Esnault, P. A. Driguez, P. Duchaussoy, P. Sizun, J. P. Hérault, J. M. Herbert, M. Petitou and P. Sinaÿ, *Angewandte Chemie - International Edition* **2001**, *40*, 1670-1673.
[75] D. R. Ferro, A. Provasoli, M. Ragazzi, B. Casu, G. Torri, V. Bossennec, B. Perly, P. Sinaÿ, M. Petitou and J. Choay, *Carbohydrate Research* **1990**, *195*, 157-167.
[76] J. C. Muñoz-García, J. López-Prados, J. Angulo, I. Díaz-Contreras, N. Reichardt, J. L. de Paz, M. Martín-Lomas and P. M. Nieto, *Chemistry – A European Journal* **2012**, *18*, 16319-16331.
[77] L. Jin, M. Hricovíni, J. A. Deakin, M. Lyon and D. Uhrín, *Glycobiology* **2009**, *19*, 1185-1196.
[78] Z. Zhang, S. A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jimenez-Barbero, M. Chen, J. Liu and R. J. Linhardt, *J Am Chem Soc* **2008**, *130*, 12998-13007.
[79] S. Ernst, G. Venkataraman, V. Sasisekharan, R. Langer, C. L. Cooney and R. Sasisekharan, *Journal of the American Chemical Society* **1998**, *120*, 2099-2107.
[80] M. Petitou and C. A. A. van Boeckel, *Angewandte Chemie - International Edition* **2004**, *43*, 3118-3133.
[81] a) C. A. A. van Boeckel, S. F. van Aelst, G. N. Wagenaars, J. R. Mellema, H. Paulsen, T. Peters, A. Pollex and V. Sinnwell, *Recueil des Travaux Chimiques des Pays-Bas* **1987**, *106*, 19-29; b) J. Angulo, J.-L. De Paz, P. M. Nieto and M. Martín-Lomas, *Israel Journal of Chemistry* **2000**, *40*, 289-299.
[82] J. Angulo, *CSIC-US* **2002**.
[83] a) V. P. Eswarakumar, I. Lax and J. Schlessinger, *Cytokine & Growth Factor Reviews* **2005**, *16*, 139-149; b) J. Kreuger, D. Spillmann, J. P. Li and U. Lindahl, *Journal of Cell Biology* **2006**, *174*, 323-327; c)

M. Bernfield, M. Gotte, P. W. Park, O. Reizes, M. L. Fitzgerald, J. Lincecum and M. Zako, *Annual Review of Biochemistry* **1999**, *68*, 729-777.
[84] a) J. Schlessinger, A. N. Plotnikov, O. A. Ibrahimi, A. V. Eliseenkova, B. K. Yeh, A. Yayon, R. J. Linhardt and M. Mohammadi, *Molecular Cell* **2000**, *6*, 743-750; b) L. Pellegrini, D. F. Burke, F. von Delft, B. Mulloy and T. L. Blundell, *Nature* **2000**, *407*, 1029-1034.
[85] a) J. L. De Paz, J. Angulo, J. M. Lassaletta, P. M. Nieto, M. Redondo-Horcajo, R. M. Lozano, G. Giménez-Gallego and M. Martín-Lomas, *ChemBioChem* **2001**, *2*, 673-685; b) R. Ojeda, J. Angulo, P. M. Nieto and M. Martin-Lomas, *Canadian Journal of Chemistry-Revue Canadienne De Chimie* **2002**, *80*, 917-936; c) R. Lucas, J. Angulo, P. M. Nieto and M. Martin-Lomas, *Organic & Biomolecular Chemistry* **2003**, *1*, 2253-2266; d) J. L. de Paz and M. Martin-Lomas, *European Journal of Organic Chemistry* **2005**, 1849-1858.
[86] B. Mulloy, M. J. Forster, C. Jones and D. B. Davies, *Biochemical Journal* **1993**, *293*, 849-858.
[87] A. D. DiGabriele, I. Lax, D. I. Chen, C. M. Svahn, M. Jaye, J. Schlessinger and W. A. Hendrickson, *Nature* **1998**, *393*, 812-817.
[88] A. Canales, R. Lozano, B. Lopez-Mendez, J. Angulo, R. Ojeda, P. M. Nieto, M. Martin-Lomas, G. Gimenez-Gallego and J. Jimenez-Barbero, *Febs Journal* **2006**, *273*, 4716-4727.
[89] M. Mohammadi, S. K. Olsen and R. Goetz, *Current Opinion in Structural Biology* **2005**, *15*, 506-516.
[90] T. C. Laurent, U. B. G. Laurent and J. R. E. Fraser, *Immunology and Cell Biology* **1996**, *74*, A1-A7.
[91] B. P. Toole, *Current Opinion in Cell Biology* **1990**, *2*, 839-844.
[92] A. J. Day and G. D. Prestwich, *Journal of Biological Chemistry* **2002**, *277*, 4585-4588.
[93] a) T. C. Laurent and J. R. E. Fraser, *Faseb Journal* **1992**, *6*, 2397-2404; b) C. B. Knudson and W. Knudson, *Faseb Journal* **1993**, *7*, 1233-1241; c) B. P. Toole, *Journal of Internal Medicine* **1997**, *242*, 35-40; d) G. Abatangelo and P. H. Weigel, *Redefining Hyaluronan* **2000**.
[94] T. E. Hardingham and A. J. Fosang, *Faseb Journal* **1992**, *6*, 861-870.
[95] A. Almond, P. L. DeAngelis and C. D. Blundell, *Journal of Molecular Biology* **2006**, *358*, 1256-1269.
[96] a) E. D. T. Atkins, D. Meader and J. E. Scott, *Int. J. Biol. Macromol.* **1980**, *2*, 318-319; b) F. Heatley and J. E. Scott, *Biochemical Journal* **1988**, *254*, 489-493; c) A. Almond, A. Brass and J. K. Sheehan, *Glycobiology* **1998**, *8*, 973-980; d) S. M. A. Holmbeck, P. A. Petillo and L. E. Lerner, *Biochemistry* **1994**, *33*, 14246-14255.
[97] A. Donati, A. Magnani, C. Bonechi, R. Barbucci and C. Rossi, *Biopolymers* **2001**, *59*, 434-445.
[98] C. D. Blundell, P. L. DeAngelis and A. Almond, *Biochemical Journal* **2006**, *396*, 487-498.
[99] a) F. Heatley, J. E. Scott, R. W. Jeanloz and E. Walker-Nasir, *Carbohydrate Research* **1982**, *99*, 1-11; b) J. E. Scott, F. Heatley, D. Moorcroft and A. H. Olavesen, *Biochemical Journal* **1981**, *199*, 829-832.
[100] T. B. H. Geijtenbeek and S. I. Gringhuis, *Nat Rev Immunol* **2009**, *9*, 465-479.
[101] P. Redelinghuys and G. D. Brown, *Immunology Letters* **2011**, *136*, 1-12.
[102] C. G. Figdor, Y. van Kooyk and G. J. Adema, *Nat Rev Immunol* **2002**, *2*, 77-84.
[103] F. Sallusto, M. Cella, C. Danieli and A. Lanzavecchia, *J. Exp. Med* **1995**, *182*, 389-400.
[104] a) M. Groger, *J. Immunol.* **2000**, *165*, 5428-5434; b) M. Fukuda, N. Hiraoka and J. C. Yeh, *J. Cell Biol.* **1999**, *147*, 467-470.
[105] W. Jiang, *Nature* **1995**, *375*, 151-155.
[106] G. D. Brown and S. Gordon, *Nature* **2001**, *413*, 36-37.
[107] K. Ariizumi, *J. Biol. Chem.* **2000**, *275*, 20157-20167.
[108] K. Ariizumi, *J. Biol. Chem.* **2000**, *275*, 11957-11963.
[109] J. Valladeau, O. Ravel, C. Dezutter-Dambuyant, K. Moore, M. Kleijmeer, Y. Liu, V. Duvert-Frances, C. Vincent, D. Schmitt, J. Davoust, C. Caux, S. Lebecque and S. Saeland, *Immunity* **2000**, *12*, 71-81.
[110] L. De Witte, A. Nabatov, M. Pion, D. Fluitsma, M. A. W. P. De Jong, T. De Gruijl, V. Piguet, Y. Van Kooyk and T. B. H. Geijtenbeek, *Nature Medicine* **2007**, *13*, 367-371.

[111] T. B. H. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. F. van Duijnhoven, G. J. Adema, Y. van Kooyk and C. G. Figdor, *Cell* **2000**, *100*, 575-585.
[112] T. B. H. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. Van Vliet, G. C. F. Van Duijnhoven, J. Middel, I. L. M. H. A. Cornelissen, H. S. L. M. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor and Y. Van Kooyk, *Cell* **2000**, *100*, 587-597.
[113] T. B. H. Geijtenbeek, D. J. E. B. Krooshoop, D. A. Bleijs, S. J. Van Vliet, G. C. F. Van Duijnhoven, V. Grabovsky, R. Alon, C. G. Figdor and Y. Van Kooyk, *Nature Immunology* **2000**, *1*, 353-357.
[114] E. Chabrol, A. Nurisso, A. Daina, E. Vassal-Stermann, M. Thepaut, E. Girard, R. R. Vivès and F. Fieschi, *PLoS One* **2012**, *7*.
[115] A. N. Zelensky and J. E. Gready, *FEBS Journal* **2005**, *272*, 6179-6217.
[116] I. Sutkeviciute, *Université de Grenoble* **2012**.
[117] A. N. Zelensky and J. E. Gready, *Proteins: Structure, Function, and Bioinformatics* **2003**, *52*, 466-477.
[118] a) K. Drickamer and R. B. Dodd, *Glycobiology* **1999**, *9*, 1357-1369; b) R. B. Dodd and K. Drickamer, *Glycobiology* **2001**, *11*, 71R-79R.
[119] A. C. S. Saphire, M. D. Bobardt, Z. Zhang, G. David and P. A. Gallay, *Journal of Virology* **2001**, *75*, 9187-9200.
[120] a) J. Arthos, C. Cicala, E. Martinelli, K. Macleod, D. Van Ryk, D. Wei, Z. Xiao, T. D. Veenstra, T. P. Conrad, R. A. Lempicki, S. McLaughlin, M. Pascuccio, R. Gopaul, J. McNally, C. C. Cruz, N. Censoplano, E. Chung, K. N. Reitano, S. Kottilil, D. J. Goode and A. S. Fauci, *Nat Immunol* **2008**, *9*, 301-309; b) C. Cicala, E. Martinelli, J. P. McNally, D. J. Goode, R. Gopaul, J. Hiatt, K. Jelicic, S. Kottilil, K. Macleod, A. O'Shea, N. Patel, D. Van Ryk, D. Wei, M. Pascuccio, L. Yi, L. McKinnon, P. Izulla, J. Kimani, R. Kaul, A. S. Fauci and J. Arthos, *Proceedings of the National Academy of Sciences* **2009**, *106*, 20877-20882.
[121] S. L. Orloff Gm Fau - Orloff, M. S. Orloff Sl Fau - Kennedy, P. J. Kennedy Ms Fau - Maddon, J. S. Maddon Pj Fau - McDougal and J. S. McDougal, *Journal of Immunology* **1991**, *146*, 2578-2587.
[122] J. C. Wilen Cb Fau - Tilton, R. W. Tilton Jc Fau - Doms and R. W. Doms, **2012**.
[123] N. M. Sherer, J. Jin and W. Mothes, *Journal of Virology* **2010**, *84*, 3248-3258.
[124] K. Miyauchi, Y. Kim, O. Latinovic, V. Morozov and G. B. Melikyan, *Cell* **2009**, *137*, 433-444.
[125] G. Melikyan, *Retrovirology* **2008**, *5*, 1-13.
[126] A. K. Abbas, A. H. H. Lichtman and S. Pillai, *Cellular and Molecular Immunology: with STUDENT CONSULT Online Access*, Elsevier Health Sciences, **2011**, p.
[127] a) P. Cameron, P. Freudenthal, J. Barker, S. Gezelter, K. Inaba and R. Steinman, *Science* **1992**, *257*, 383-387; b) M. Pope, M. G. H. Betjes, N. Romani, H. Hirmand, P. U. Cameron, L. Hoffman, S. Gezelter, G. Schuler and R. M. Steinman, *Cell* **1994**, *78*, 389-398.
[128] D. S. Kwon, G. Gregorio, N. Bitton, W. A. Hendrickson and D. R. Littman, *Immunity* **2002**, *16*, 135-144.
[129] N. Sol-Foulon, A. Moris, C. Nobile, C. Boccaccio, A. Engering, J. P. Abastado, J. M. Heard, Y. Van Kooyk and O. Schwartz, *Immunity* **2002**, *16*, 145-155.
[130] a) K. B. Gurney, J. Elliott, H. Nassanian, C. Song, E. Soilleux, I. McGowan, P. A. Anton and B. Lee, *Journal of Virology* **2005**, *79*, 5762-5773; b) S. Sattin, A. Daghetti, M. Thepaut, A. Berzi, M. Sanchez-Navarro, G. Tabarani, J. Rojo, F. Fieschi, M. Clerici and A. Bernardi, *ACS Chemical Biology* **2010**, *5*, 301-312; c) A. Berzi, J. J. Reina, R. Ottria, I. Sutkeviciute, P. Antonazzo, M. Sanchez-Navarro, E. Chabrol, M. Biasin, D. Trabattoni, I. Cetin, J. Rojo, F. Fieschi, A. Bernardi and M. Clerici, *Aids* **2012**, *26*, 127-137.
[131] a) L. de Witte, A. Nabatov and T. B. H. Geijtenbeek, *Trends in Molecular Medicine* **2008**, *14*, 12-19; b) L. Burleigh, P. Y. Lozach, C. Schiffer, I. Staropoli, V. Pezo, F. Porrot, B. Canque, J. L. Virelizier, F. Arenzana-Seisdedos and A. Amara, *Journal of Virology* **2006**, *80*, 2949-2957.
[132] F. Hladik and M. J. McElrath, *Nat Rev Immunol* **2008**, *8*, 447-457.
[133] a) E. Fithian, P. Kung, G. Goldstein, M. Rubenfeld, C. Fenoglio and R. Edelson, *Proc Natl Acad Sci U S A* **1981**, *78*, 2541-2544; b) B. K. Patterson, A. Landay, J. N. Siegel, Z. Flener, D. Pessis, A. Chaviano and R. C. Bailey, *American Journal of Pathology* **2002**, *161*, 867-873.

[134] R. E. Hunger, P. A. Sieling, M. T. Ochoa, M. Sugaya, A. E. Burdick, T. H. Rea, P. J. Brennan, J. T. Belisle, A. Blauvelt, S. A. Porcelli and R. L. Modlin, *Journal of Clinical Investigation* **2004**, *113*, 701-708.
[135] P. Stoitzner and N. Romani, *European Journal of Immunology* **2011**, *41*, 2526-2529.
[136] a) L. Klareskog, U. M. Tjernlund, U. Forsum and P. A. Peterson, *Nature* **1977**, *268*, 248-250; b) G. Rowden, M. G. Lewis and A. K. Sullivan, *Nature* **1977**, *268*, 247-248.
[137] J. Banchereau and R. M. Steinman, *Nature* **1998**, *392*, 245-252.
[138] M. S. Birbeck, A. S. Breathnach and J. D. Everall, *The Journal of Investigative Dermatology* **1961**, *37*, 51-64.
[139] O. Schwartz, *Nat Med* **2007**, *13*, 245-246.
[140] a) N. Romani, B. E. Clausen and P. Stoitzner, *Immunological Reviews* **2010**, *234*, 120-141; b) J. Idoyaga, N. Suda, K. Suda, C. G. Park and R. M. Steinman, *Proc Natl Acad Sci U S A* **2009**, *106*, 1524-1529.
[141] J. Valladeau, C. Dezutter-Dambuyant and S. Saeland, *Immunologic Research* **2003**, *28*, 93-107.
[142] S. G. Turville, J. Arthos, K. MacDonald, G. Lynch, H. Naif, G. Clark, D. Hart and A. L. Cunningham, *Blood* **2001**, *98*, 2482-2488.
[143] R. E. Hunger, P. A. Sieling, M. T. Ochoa, M. Sugaya, A. E. Burdick, T. H. Rea, P. J. Brennan, J. T. Belisle, A. Blauvelt, S. A. Porcelli and R. L. Modlin, *The Journal of Clinical Investigation* **2004**, *113*, 701-708.
[144] a) K. Takahara, Y. Yashima, Y. Omatsu, H. Yoshida, Y. Kimura, Y. S. Kang, R. M. Steinman, C. G. Park and K. Inaba, *International Immunology* **2004**, *16*, 819-829; b) K. Takahara, Y. Omatsu, Y. Yashima, Y. Maeda, S. Tanaka, T. Iyoda, B. Clusen, K. Matsubara, J. Letterio, R. M. Steinman, Y. Matsuda and K. Inaba, *International Immunology* **2002**, *14*, 433-444.
[145] a) F. Hladik, P. Sakchalathorn, L. Ballweber, G. Lentz, M. Fialkow, D. Eschenbach and M. J. McElrath, *Immunity* **2007**, *26*, 257-270; b) S. G. Turville, P. U. Cameron, A. Handley, G. Lin, S. Pöhlmann, R. W. Doms and A. L. Cunningham, *Nature Immunology* **2002**, *3*, 975-983; c) L. A. Hussain and T. Lehner, *Immunology* **1995**, *85*, 475-484.
[146] L. Ballweber, B. Robinson, A. Kreger, M. Fialkow, G. Lentz, M. J. McElrath and F. Hladik, *Journal of Virology* **2011**, *85*, 13443-13447.
[147] L. Chatwell, A. Holla, B. B. Kaufer and A. Skerra, *Molecular Immunology* **2008**, *45*, 1981-1994.
[148] a) N. S. Stambach and M. E. Taylor, *Glycobiology* **2003**, *13*, 401-410; b) H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer and W. I. Weis, *Journal of Molecular Biology* **2011**, *405*, 1027-1039.
[149] M. Thépaut, J. Valladeau, A. Nurisso, R. Kahn, B. Arnou, C. Vivès, S. Saeland, C. Ebel, C. Monnier, C. Dezutter-Dambuyant, A. Imberty and F. Fieschi, *Biochemistry* **2009**, *48*, 2684-2698.
[150] W. I. Weis and K. Drickamer in *Structural basis of lectin-carbohydrate recognition*, *Vol. 65* **1996**, pp. 441-473.
[151] W. I. Weis and K. Drickamer, *Structure* **1994**, *2*, 1227-1240.
[152] a) H. Feinberg, D. A. Mitchell, K. Drickamer and W. I. Weis, *Science* **2001**, *294*, 2163-2166; b) K. Drickamer, *Curr Opin Struct Biol* **1999**, *9*, 585-590.
[153] H. Feinberg, A. S. Powlesland, M. E. Taylor and W. I. Weis, *Journal of Biological Chemistry* **2010**, *285*, 13285-13293.
[154] H. Feinberg, R. Castelli, K. Drickamer, P. H. Seeberger and W. I. Weis, *Journal of Biological Chemistry* **2007**, *282*, 4202-4209.
[155] a) H. Tateno, K. Ohnishi, R. Yabe, N. Hayatsu, T. Sato, M. Takeya, H. Narimatsu and J. Hirabayashi, *Journal of Biological Chemistry* **2010**, *285*, 6390-6400; b) C. Galustian, C. G. Park, W. Chai, M. Kiso, S. A. Bruening, Y. S. Kang, R. M. Steinman and T. Feizi, *Int Immunol* **2004**, *16*, 853-866.
[156] a) L. Bousarghin, P. Hubert, E. Franzen, N. Jacobs, J. Boniver and P. Delvenne, *Journal of General Virology* **2005**, *86*, 1297-1305; b) L. de Witte, Y. Zoughlami, B. Aengeneyndt, G. David, Y. van Kooyk, L. Gissmann and T. B. H. Geijtenbeek, *Immunobiology* **2008**, *212*, 679-691.
[157] G. M. O'Sullivan, C. M. Boswell and G. M. Halliday, *Experimental Dermatology* **2000**, *9*, 25-33.
[158] a) M. J. Keiser, J. J. Irwin and B. K. Shoichet, *Biochemistry* **2010**, *49*, 10267-10276; b) J. A. Wells and C. L. McClendon, *Nature* **2007**, *450*, 1001-1009; c) J. T. Metz and P. J. Hajduk, *Current Opinion in*

Chemical Biology **2010**, *14*, 498-504; d) M. R. Arkin and J. A. Wells, *Nat Rev Drug Discov* **2004**, *3*, 301-317; e) X. Morelli, R. Bourgeas and P. Roche, *Current Opinion in Chemical Biology* **2011**, *15*, 475-481.
[159] R. J. Woods and M. B. Tessier, *Curr Opin Struct Biol* **2010**, *20*, 575-583.
[160] a) J.-P. Renaud and M.-A. Delsuc, *Current Opinion in Pharmacology* **2009**, *9*, 622-628; b) F. H. Niesen, H. Berglund and M. Vedadi, *Nat. Protocols* **2007**, *2*, 2212-2221; c) K. C. Luk, E. G. Hyde, J. Q. Trojanowski and V. M. Y. Lee, *Biochemistry* **2007**, *46*, 12522-12529; d) N. Jonker, J. Kool, H. Irth and W. M. A. Niessen, *Analytical and Bioanalytical Chemistry* **2011**, *399*, 2669-2681.
[161] a) O. Trott and A. J. Olson, *Journal of Computational Chemistry* **2010**, *31*, 455-461; b) S.-Y. Huang, S. Z. Grinter and X. Zou, *Physical Chemistry Chemical Physics* **2010**, *12*, 12899-12908.
[162] a) L. E. Kay, *Journal of Magnetic Resonance* **2011**, *213*, 492-494; b) A. Joachimiak, *Curr Opin Struct Biol* **2009**, *19*, 573-584.
[163] S. Pérez, K. Mazeau and C. Hervé du Penhoat, *Plant Physiology and Biochemistry* **2000**, *38*, 37-55.
[164] R. Chandrasekaran and A. Giacometti, **1997**, p. 54.
[165] S. Pérez and H. Chanzy, *Journal of electron microscopy technique* **1989**, *11*, 280-285.
[166] a) M. Pellecchia, I. Bertini, D. Cowburn, C. Dalvit, E. Giralt, W. Jahnke, T. L. James, S. W. Homans, H. Kessler, C. Luchinat, B. Meyer, H. Oschkinat, J. Peng, H. Schwalbe and G. Siegal, *Nature Reviews Drug Discovery* **2008**, *7*, 738-745; b) J. W. Peng, J. Moore and N. Abdul-Manan, *Progress in Nuclear Magnetic Resonance Spectroscopy* **2004**, *44*, 225-256; c) M. Goldflam, T. Tarragó, M. Gairí and E. Giralt in *NMR Studies of Protein–Ligand Interactions*, *Vol. 831* Eds.: A. Shekhtman and D. S. Burz), Humana Press, **2012**, pp. 233-259.
[167] a) M. Billeter, G. Wagner and K. Wüthrich, *Journal of Biomolecular NMR* **2008**, *42*, 155-158; b) J. Cavanagh, W. J. Fairbrother, A. G. Palmer and N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Elsevier Science, **1995**, p.
[168] B. Meyer and T. Peters, *Angewandte Chemie International Edition* **2003**, *42*, 864-890.
[169] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier Science Limited, **2009**, p.
[170] B. T. Farmer Ii, S. Macura and L. R. Brown, *Journal of Magnetic Resonance (1969)* **1988**, *80*, 1-22.
[171] M. Ravikumar and A. A. Bothner-By, *Journal of the American Chemical Society* **1993**, *115*, 7537-7538.
[172] a) H. Desvaux, P. Berthault, N. Birlirakis and M. Goldman, *Journal of Magnetic Resonance Series A* **1994**, *108*, 219-229; b) P. Berthault, N. Birlirakis, G. Rubinstenn, P. Sinay and H. Desvaux, *Journal of Biomolecular NMR* **1996**, *8*, 23-35.
[173] a) P. Balaram, A. A. Bothner-By and E. Breslow, *Journal of the American Chemical Society* **1972**, *94*, 4017-4018; b) P. Balaram, A. A. Bothner-By and J. Dadok, *Journal of the American Chemical Society* **1972**, *94*, 4015-4017; c) G. M. Clore and A. M. Gronenborn, *Journal of Magnetic Resonance (1969)* **1983**, *53*, 423-442; d) H. N. B. Moseley, E. V. Curto and N. R. Krishna, *Journal of Magnetic Resonance Series B* **1995**, *108*, 243-261.
[174] a) D. Neuhaus and M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, Wiley-VCH, New York, **2000**, p; b) F. Ni, *Progress in Nuclear Magnetic Resonance Spectroscopy* **1994**, *26*, 517-606; c) C. B. Post, *Current Opinion in Structural Biology* **2003**, *13*, 581-588.
[175] a) J. L. Asensio, F. J. Canada and J. Jimenez-Barbero, *European Journal of Biochemistry* **1995**, *233*, 618-630; b) M. Hricovini, M. Guerrini and A. Bisio, *European Journal of Biochemistry* **1999**, *261*, 789-801; c) P. M. Enríquez-Navas, M. Marradi, D. Padro, J. Angulo and S. Penadés, *Chemistry- A European Journal* **2011**, *17*, 1547-1560; d) M. Thépaut, C. Guzzi, I. Sutkeviciute, S. Sattin, R. Ribeiro-Viana, N. Varga, E. Chabrol, J. Rojo, A. Bernardi, J. Angulo, P. M. Nieto and F. Fieschi, *Journal of the American Chemical Society* **2013**, *135*, 2518-2529.
[176] A. A. Bothnerby, R. L. Stephens, J. M. Lee, C. D. Warren and R. W. Jeanloz, *Journal of the American Chemical Society* **1984**, *106*, 811-813.
[177] C. Guzzi, *CSIC-University of Sevilla* **2013**.

[178] B. Meyer, T. Weimar and T. Peters, *European Journal of Biochemistry* **1997**, *246*, 705-709.
[179] M. Mayer and B. Meyer, *Angewandte Chemie International Edition* **1999**, *38*, 1784-1788.
[180] P. M. Enríquez-Navas, *CSIC-University of Sevilla* **2011**.
[181] J. L. Yan, A. D. Kline, H. P. Mo, M. J. Shapiro and E. R. Zartler, *Journal of Magnetic Resonance* **2003**, *163*, 270-276.
[182] M. Mayer and T. L. James, *Journal of the American Chemical Society* **2004**, *126*, 4453-4460.
[183] M. Mayer and B. Meyer, *Journal of the American Chemical Society* **2001**, *123*, 6108-6117.
[184] J. Angulo, P. M. Enriquez-Navas and P. M. Nieto, *Chemistry- A European Journal* **2010**, *16*, 7803-7812.
[185] A. A. Roberts, *Israel Journal of Chemistry* **1986**, *27*, 198-210.
[186] a) B. M. Sattelle and A. Almond, *Journal of Computational Chemistry* **2010**, *31*, 2932-2947; b) J. W. Brady, *Advances in biophysical chemistry* **1990**, *1*, 155-202; c) R. J. Woods, *Reviews in computational chemistry* **1996**, *9*, 129-165; d) T. Weimar, *Combining NMR and simulation methods in oligosaccharide conformational analysis*
NMR of Glycoconjugates, **2002**, p. 111-144.
[187] a) S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *Journal of Computational Chemistry* **1986**, *7*, 230-252; b) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.* **1983**, *4*, 187-217; c) M. Karplus, *The prediction and analysis of mutant structures*
Protein Engineering, **1987**, p. 35-44; d) U. a. H. Dinur, A.T. , *Review of computational chemistry. (eds. K.B. Lipkowitz and D.B. Boyd) Wiley-VCH, New York, NY.* **1991**, 99-164.
[188] a) S. Lifson and P. S. Stern, *The Journal of Chemical Physics* **1982**, *77*, 4542-4550; b) G. Nemethy, M. S. Pottle and H. A. Scheraga, *The Journal of Physical Chemistry* **1983**, *87*, 1883-1887; c) Hermans, Jr., *Biopolymers* **1984**, *23*, 1513-1518; d) S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta Jr and P. Weiner, *Journal of the American Chemical Society* **1984**, *106*, 765-784; e) L. Nilsson and M. Karplus, *Journal of Computational Chemistry* **1986**, *7*, 591-616; f) R. A. Engh and R. Huber, *Acta Crystallographica Section A* **1991**, *47*, 392-400.
[189] T. A. Halgren, *Curr Opin Struct Biol* **1995**, *5*, 205-210.
[190] D. Bordo and P. Argos, *Journal of Molecular Biology* **1994**, *243*, 504-519.
[191] a) K. A. Dill, *Biochemistry* **1990**, *29*, 7133-7155; b) A. R. Fersht and L. Serrano, *Curr Opin Struct Biol* **1993**, *3*, 75-83.
[192] A. Fersht, *Enzyme structure and function 2nd ed*, **1985**, p.
[193] a) W. W. Cleland and M. M. Kreevoy, *Science* **1994**, *264*, 1887-1890; b) P. A. Frey, S. A. Whitt and J. B. Tobin, *Science* **1994**, *264*, 1927-1930; c) M. C. Hutter and V. Helms, *Protein Science* **2000**, *9*, 2225-2231.
[194] a) J. A. Ippolito, R. S. Alexander and D. W. Christianson, *Journal of Molecular Biology* **1990**, *215*, 457-471; b) Y. Mandel-Gutfreund, O. Schueler and H. Margalit, *Journal of Molecular Biology* **1995**, *253*, 370-382.
[195] E. N. Baker and R. E. Hubbard, *Progress in Biophysics and Molecular Biology* **1984**, *44*, 97-179.
[196] a) A. T. Hagler, S. Lifson and P. Dauber, *Journal of the American Chemical Society* **1979**, *101*, 5122-5130; b) W. L. R. Jorgensen, J.T., *Journal of the American Chemical Society* **1988**, *110*, 657-666; c) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *Journal of the American Chemical Society* **1995**, *117*, 5179-5197; d) A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B* **1998**, *102*, 3586-3616.
[197] F. A. Momany, R. F. McGuire, A. W. Burgess and H. A. Scheraga, *Journal of Physical Chemistry* **1975**, *79*, 2361-2381.
[198] C. S. Ewig, T. S. Thacher and A. T. Hagler, *The Journal of Physical Chemistry B* **1999**, *103*, 6998-7014.

[199] A. T. Hagler, E. Huler and S. Lifson, *Journal of the American Chemical Society* **1974**, *96*, 5319-5327.
[200] J.-H. Lii and N. L. Allinger, *Journal of Computational Chemistry* **1998**, *19*, 1001-1016.
[201] a) C. Gorbitz, *Acta Crystallographica Section B* **1989**, *45*, 390-395; b) D. F. Sticke, L. G. Presta, K. A. Dill and G. D. Rose, *Journal of Molecular Biology* **1992**, *226*, 1143-1159; c) E. D. Isaacs, A. Shukla, P. M. Platzman, D. R. Hamann, B. Barbiellini and C. A. Tulk, *Physical Review Letters* **1999**, *82*, 600-603.
[202] J. A. McCammon, B. R. Gelin and M. Karplus, *Nature* **1977**, *267*, 585-590.
[203] D. L. Beveridge and G. Ravishanker, *Curr Opin Struct Biol* **1994**, *4*, 246-255.
[204] J. W. Brady, *Journal of the American Chemical Society* **1986**, *108*, 8153-8160.
[205] D. French Alfred and J. W. Brady in *Vol. 430* American Chemical Society, **1990**.
[206] M. Karplus and J. A. McCammon, *Nat Struct Biol* **2002**, *9*, 646-652.
[207] J. Gonzalez-Outeiriño, K. N. Kirschner, S. Thobhani and R. J. Woods, *Canadian Journal of Chemistry* **2006**, *84*, 569-579.
[208] a) H. Alonso, A. A. Bliznyuk and J. E. Gready, *Medicinal Research Reviews* **2006**, *26*, 531-568; b) C. A. Taft, V. B. da Silva and C. H. T. d. P. da Silva, *Journal of Pharmaceutical Sciences* **2008**, *97*, 1089-1098; c) W. L. Jorgensen, *Science* **2004**, *303*, 1813-1818.
[209] R. Woods and A. Yongye in *Computational Techniques Applied to Defining Carbohydrate Antigenicity*, *Vol.* Eds.: P. Kosma and S. Müller-Loennies), Springer Vienna, **2012**, pp. 361-383.
[210] a) S. P. Kawatkar, D. A. Kuntz, R. J. Woods, D. R. Rose and G. J. Boons, *J Am Chem Soc* **2006**, *128*, 8310-8319; b) S. Pedatella, M. De Nisco, B. Ernst, A. Guaragna, B. Wagner, R. J. Woods and G. Palumbo, *Carbohydr Res* **2008**, *343*, 31-38; c) X. Wen, Y. Yuan, D. A. Kuntz, D. R. Rose and B. M. Pinto, *Biochemistry* **2005**, *44*, 6729-6737; d) A. Germer, M. G. Peter and E. Kleinpeter, *J Org Chem* **2002**, *67*, 6328-6338.
[211] a) Z. Shreif, P. Adhangale, S. Cheluvaraja, R. Perera, R. Kuhn and P. Ortoleva, *Scientific Modeling and Simulation SMNS* **2008**, *15*, 363-380; b) G. B. McGaughey, G. Barbato, E. Bianchi, R. M. Freidinger, V. M. Garsky, W. M. Hurni, J. G. Joyce, X. Liang, M. D. Miller, A. Pessi, J. W. Shiver and M. J. Bogusky, *Current HIV research* **2004**, *2*, 193-204; c) R. Kadirvelraj, J. Gonzalez-Outeirino, B. L. Foley, M. L. Beckham, H. J. Jennings, S. Foote, M. G. Ford and R. J. Woods, *Proc Natl Acad Sci U S A* **2006**, *103*, 8149-8154.
[212] a) A. D. Hill and P. J. Reilly, *Journal of Computational Chemistry* **2008**, *29*, 1131-1141; b) L. Alain and J. R. Peter, *Journal of Computational Chemistry* **2003**, *24*, 1748-1757; c) A. Kerzmann, J. Fuhrmann, O. Kohlbacher and D. Neumann, *Journal of Chemical Information and Modeling* **2008**, *48*, 1616-1625.
[213] a) S. Pérez, A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jimenez-Barbero, A. Poveda, J. F. Espinosa, B. P. Van Eyck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten and K. Rasmussen, *Carbohydrate Research* **1998**, *314*, 141-155; b) T. Yoda, Y. Sugita and Y. Okamoto, *Chemical Physics Letters* **2004**, *386*, 460-467.
[214] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *Journal of the American Chemical Society* **1996**, *118*, 11225-11236.
[215] M. T. Sarah, N. K. Karl and J. W. Robert in *Incorporation of Carbohydrates into Macromolecular Force Fields: Development and Validation*, *Vol. 930* American Chemical Society, **2006**, pp. 235-257.
[216] S. Anita and P. Serge in *Protein?Carbohydrate Interactions*, *Vol.* CRC Press, **2012**, pp. 71-110.
[217] a) S. W. Homans, R. A. Dwek, D. L. Fernandes and T. W. Rademacher, *FEBS Letters* **1983**, *164*, 231-235; b) J. R. Brisson and J. P. Carver, *Biochemistry* **1983**, *22*, 1362-1368.
[218] M. Kuttel, J. W. Brady and K. J. Naidoo, *Journal of Computational Chemistry* **2002**, *23*, 1236-1243.
[219] E. Fadda and R. J. Woods, *Drug Discov Today* **2010**, *15*, 596-609.
[220] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley and R. J. Woods, *J Comput Chem* **2008**, *29*, 622-655.
[221] M. B. Tessier, M. L. Demarco, A. B. Yongye and R. J. Woods, *Mol Simul* **2008**, *34*, 349-363.

[222] X. Biarnés, J. Nieto, A. Planas and C. Rovira, *Journal of Biological Chemistry* **2006**, *281*, 1432-1441.
[223] a) M. Basma, S. Sundara, D. Calgan, T. Vernali and R. J. Woods, *J Comput Chem* **2001**, *22*, 1125-1137; b) R. J. Woods, M. Khalil, W. Pell, S. H. Moffat and V. H. Smith, *Journal of Computational Chemistry* **1990**, *11*, 297-310.
[224] R. J. Woods and R. Chappelle, *Journal of Molecular Structure: THEOCHEM* **2000**, *527*, 149-156.
[225] R. D. Lins and P. H. Hünenberger, *Journal of Computational Chemistry* **2005**, *26*, 1400-1412.
[226] a) O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell Jr, *Journal of Chemical Theory and Computation* **2009**, *5*, 2353-2370; b) O. Guvench, S. N. Greenr, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell Jr, *Journal of Computational Chemistry* **2008**, *29*, 2543-2564.
[227] D. Kony, W. Damm, S. Stoll and W. F. Van Gunsteren, *Journal of Computational Chemistry* **2002**, *23*, 1416-1429.
[228] a) L. D. Schuler and W. F. Van Gunsteren, *Mol Simul* **2000**, *25*, 301-319; b) L. D. Schuler, X. Daura and W. F. van Gunsteren, *Journal of Computational Chemistry* **2001**, *22*, 1205-1218.
[229] H. J. C. Berendsen, J. P. M. Postma, W. F. V. Gunsteren and J. Hermans, *Interaction Models for Water in Relation to Protein Hydration*, Intermolecular Forces, Pullman, B., Dordrecht: Reidel Publishing Company. , **1981**, p.
[230] N. S. Gandhi and R. L. Mancera, *Carbohydrate Research* **2010**, *345*, 689-695.
[231] C. J. M. Huige and C. Altona, *Journal of Computational Chemistry* **1995**, *16*, 56-79.
[232] a) A. D. MacKerell, N. Banavali and N. Foloppe, *Biopolymers* **2000**, *56*, 257-265; b) A. D. Mackerell Jr, M. Feig and C. L. Brooks Iii, *Journal of Computational Chemistry* **2004**, *25*, 1400-1415.
[233] W. Damm, A. Frontera, J. Tirado-Rives and W. L. Jorgensen, *Journal of Computational Chemistry* **1997**, *18*, 1955-1970.
[234] a) N. L. Allinger, F. Li, L. Yan and J. C. Tai, *J. Comput. Chem.* **1990**, *11*, 868-895; b) J.-H. Lii and N. L. Allinger, *Journal of Computational Chemistry* **1991**, *12*, 186-199.
[235] M. Clark, R. D. Cramer and N. Vanopdenbosch, *Journal of Computational Chemistry* **1989**, *10*, 982-1012.
[236] a) A. Imberty, K. D. Hardman, J. P. Carver and S. Perez, *Glycobiology* **1991**, *1*, 631-642; b) S. Pérez, C. Meyer and A. Imberty, *Molecular Engineering* **1995**, *5*, 271-300.
[237] R. K. a. E. R. P. Z. W.F. van Gunsteren, *"Proceedings NATO/CECAM workshop on nucleic acid conformation and dynamics" W.K. Olson ed* **1984**, 79-92.
[238] D. A. Cumming and J. P. Carver, *Biochemistry* **1987**, *26*, 6664-6676.
[239] a) A. E. Torda, R. M. Scheek and W. F. van Gunsteren, *Chemical Physics Letters* **1989**, *157*, 289-294; b) D. A. Pearlman and P. A. Kollman, *Journal of Molecular Biology* **1991**, *220*, 457-479; c) A. E. Torda, R. M. Scheck and W. F. Van Gunsteren, *Journal of Molecular Biology* **1990**, *214*, 223-235; d) A. E. Torda, R. M. Brunne, T. Huber, H. Kessler and W. F. van Gunsteren, *Journal of Biomolecular NMR* **1993**, *3*, 55-66; e) D. A. Pearlman, *Journal of Biomolecular NMR* **1994**, *4*, 279-299; f) A. P. Nanzer, W. F. van Gunsteren and A. E. Torda, *Journal of Biomolecular NMR* **1995**, *6*, 313-320.
[240] a) H. Kessler, C. Griesinger, J. Lautz, A. Müller, W. F. Van Gunsteren and H. J. C. Berendsen, *Journal of the American Chemical Society* **1988**, *110*, 3393-3396; b) J. Tropp, *The Journal of Chemical Physics* **1979**, *72*, 6035-6043.
[241] P. M. S. Hendrickx, F. Corzana, S. Depraetere, D. A. Tourwé, K. Augustyns and J. C. Martins, *Journal of Computational Chemistry* **2010**, *31*, 561-572.
[242] a) P. Bultinck, *Computational Medicinal Chemistry for Drug Discovery*, Taylor & Francis, **2004**, p; b) R. Eklund, *Computational Analysis of Carbohydrates: Dynamical Properties and Interactions*, Department of Organic Chemistry, Stockholm University, **2005**, p; c) H.-D. Höltje and G. Folkers in *Molecular Modeling: Basic Principles and Applications*, *Vol.* Wiley-VCH Verlag GmbH, **2008**; d) A. Kranjc, *International School for Advanced Studies (SISSA), Italy* **2009**; e) R. C. Walker, *Imperial College London, UK* **2003**.
[243] T. J. A. Ewing, S. Makino, A. G. Skillman and I. D. Kuntz, *Journal of Computer-Aided Molecular Design* **2001**, *15*, 411-428.

[244] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *Journal of Computational Chemistry* **1998**, *19*, 1639-1662.
[245] C. Wellock, Ross, B.J., *Proceedings of Late Breaking Papers, Genetic and Evolutionary Computation Conference (GECCO-2001)* **2001**, 474-481.
[246] a) R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of Medicinal Chemistry* **2004**, *47*, 1739-1749; b) T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *Journal of Medicinal Chemistry* **2004**, *47*, 1750-1759.
[247] a) N. Huang, C. Kalyanaraman, K. Bernacki and M. P. Jacobson, *Physical Chemistry Chemical Physics* **2006**, *8*, 5166-5177; b) E. Yuriev and P. A. Ramsland, *Structural Glycobiology*, CRC Press,Taylor & Francis Group, **2012**, p.
[248] G. M. Morris, H. Ruth, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *Journal of Computational Chemistry* **2009**, *30*, 2785-2791.
[249] a) D. N. A. Boobbyer, P. J. Goodford, P. M. McWhinnie and R. C. Wade, *Journal of Medicinal Chemistry* **1989**, *32*, 1083-1094; b) R. C. Wade, K. J. Clark and P. J. Goodford, *Journal of Medicinal Chemistry* **1993**, *36*, 140-147; c) R. C. Wade and P. J. Goodford, *Journal of Medicinal Chemistry* **1993**, *36*, 148-156.
[250] I. Muegge and Y. C. Martin, *Journal of Medicinal Chemistry* **1999**, *42*, 791-804.
[251] a) B. Casu and U. Lindahl in *Structure and biological interactions of heparin and heparan sulfate*, *Vol. 57* **2001**, pp. 159-206; b) B. Mulloy and R. J. Linhardt, *Curr Opin Struct Biol* **2001**, *11*, 623-628; c) U. Lindahl and J. Li in *Chapter 3 Interactions Between Heparan Sulfate and Proteins-Design and Functional Implications*, *Vol. 276* **2009**, pp. 105-159.
[252] J. D. Esko and S. B. Selleck, *Annual Review of Biochemistry* **2002**, *71*, 435-471.
[253] a) R. Raman, V. Sasisekharan and R. Sasisekharan, *Chemistry & Biology* **2005**, *12*, 267-277; b) S. Khan, J. Gor, B. Mulloy and S. J. Perkins, *Journal of Molecular Biology* **2010**, *395*, 504-521; c) S. Khan, E. Rodriguez, R. Patel, J. Gor, B. Mulloy and S. J. Perkins, *Journal of Biological Chemistry* **2011**, *286*, 24842-24854.
[254] A. Canales, J. Angulo, R. Ojeda, M. Bruix, R. Fayos, R. Lozano, G. Giménez-Gallego, M. Martín-Lomas, P. M. Nieto and J. Jiménez-Barbero, *Journal of the American Chemical Society* **2005**, *127*, 5778-5779.
[255] L. Pellegrini, *Curr Opin Struct Biol* **2001**, *11*, 629-634.
[256] M. Guerrini, T. Agulles, A. Bisio, M. Hricovini, L. Lay, A. Naggi, L. Poletti, L. Sturiale, G. Torri and B. Casu, *Biochemical and Biophysical Research Communications* **2002**, *292*, 222-230.
[257] J. L. de Paz, *CSIC-University of Sevilla* **2000**.
[258] E. A. Yates, F. Santini, B. De Cristofano, N. Payre, C. Cosentino, M. Guerrini, A. Naggi, G. Torri and M. Hricovini, *Carbohydrate Research* **2000**, *329*, 239-247.
[259] S. Macura, B. T. Farmer and L. R. Brown, *Journal of Magnetic Resonance* **1986**, *70*, 493-499.
[260] H. Hu, *Journal of Magnetic Resonance* **2006**, *182*, 173-177.
[261] G. M. Clore and A. M. Gronenborn, *Journal of Magnetic Resonance (1969)* **1985**, *61*, 158-164.
[262] C. A. A. Van Boeckel, S. F. Vanaelst, G. N. Wagenaars, J. R. Mellema, H. Paulsen, T. Peters, A. Pollex and V. Sinnwell, *Recueil Des Travaux Chimiques Des Pays-Bas-Journal of the Royal Netherlands Chemical Society* **1987**.
[263] K. J. Murphy, N. McLay and D. A. Pye, *Journal of the American Chemical Society* **2008**, *130*, 12435-12444.
[264] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics* **1983**, *79*, 926-935.
[265] C. J. M. Huige and C. Altona, *Journal of Computational Chemistry* **1995**, *16*, 56-79.
[266] a) B. R. Brooks, Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. , *J. Comput. Chem.* **1983**, *4*, 187–217; b) S. J. Weiner, P. A. Kollman, D. T. Nguyen and D. A. Case, *Journal of Computational Chemistry* **1986**, *7*, 230-252.
[267] D. M. York, T. A. Darden and L. G. Pedersen, *The Journal of Chemical Physics* **1993**, *99*, 8345-8348.

[268] R. J. Woods, R. A. Dwek, C. J. Edge and B. Fraser-Reid, *Journal of Physical Chemistry* **1995**, *99*, 3832-3846.
[269] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Structure, Function and Genetics* **2006**, *65*, 712-725.
[270] C. A. G. Haasnoot, F. A. A. M. de Leeuw and C. Altona, *Tetrahedron* **1980**, *36*, 2783-2792.
[271] J. Angulo, P. M. Nieto and M. Martín-Lomas, *Chemical Communications* **2003**, *9*, 1512-1513.
[272] N. M. Laurendeau, *Cambridge University Press* **2010**.
[273] Z. Zhang, S. A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jimenez-Barbero, M. Chen, J. Liu and R. J. Linhardt, *J. Am. Chem. Soc.* **2008**, *130*, 12998-13007.
[274] a) J. DeChancie and K. N. Houk, *J. Am. Chem. Soc.* **2007**, *129*, 5419-5429; b) S. Forli and A. J. Olson, *Journal of Medicinal Chemistry* **2012**, *55*, 623-638.
[275] a) G. Lipari and A. Szabo, *Journal of the American Chemical Society* **1982**, *104*, 4546-4559; b) G. Lipari and A. Szabo, *Journal of the American Chemical Society* **1982**, *104*, 4559-4570.
[276] R. Ojeda, J. Angulo, P. M. Nieto and M. Martín-Lomas, *Canadian Journal of Chemistry* **2002**, *80*, 917-936.
[277] L. Pellegrini, *Current Opinion in Structural Biology* **2001**, *11*, 629-634.
[278] a) M. Hricovini and G. Torri, *Carbohydrate Research* **1995**, *268*, 159-175; b) D. Mikhailov, K. H. Mayo, I. R. Vlahov, T. Toida, A. Pervin and R. J. Linhardt, *Biochemical Journal* **1996**, *318*, 93-102.
[279] J. Angulo, M. Hricovíni, M. Gairi, M. Guerrini, J. L. de Paz, R. Ojeda, M. Martín-Lomas and P. M. Nieto, *Glycobiology* **2005**, *15*, 1008-1015.
[280] a) R. Boelens, T. M. G. Koning and R. Kaptein, *Journal of Molecular Structure* **1988**, *173*, 299-311; b) B. A. Borgias and T. L. James, *Journal of Magnetic Resonance (1969)* **1990**, *87*, 475-487; c) B. Borgias, Thomas, P.D., James, T.L. , *Complete Relaxation Matrix Analysis (CORMA)*, University of California, San Francisco, **1989**, p.
[281] M. Hricovíni, P. M. Nieto and G. Torri in *NMR of Sulfated Oligo- and Polysaccharides*, *Vol.* Wiley-VCH Verlag GmbH & Co. KGaA, **2003**, pp. 189-229.
[282] a) T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics* **1993**, *98*, 10089-10092; b) H. G. Petersen, *The Journal of Chemical Physics* **1995**, *103*, 3668-3679.
[283] M. Hricovíni, *Carbohydrate Research* **2006**, *341*, 2575-2580.
[284] M. Hricovíni, *The Journal of Physical Chemistry B* **2011**, *115*, 1503-1511.
[285] a) A. Canales-Mayordomo, R. Fayos, J. Angulo, R. Ojeda, M. Martin-Pastor, P. M. Nieto, M. Martin-Lomas, R. Lozano, G. Gimenez-Gallego and J. Jimenez-Barbero, *Journal of Biomolecular Nmr* **2006**, *35*, 225-239; b) A. Canales, J. Angulo, R. Ojeda, M. Bruix, R. Fayos, R. Lozano, G. Gimenez-Gallego, M. Martin-Lomas, P. M. Nieto and J. Jimenez-Barbero, *J Am Chem Soc* **2005**, *127*, 5778-5779.
[286] K. Stott, J. Keeler, Q. N. Van and A. J. Shaka, *Journal of Magnetic Resonance* **1997**, *125*, 302-324.
[287] A. A. Shaw, C. Salaun, J. F. Dauphin and B. Ancian, *Journal of Magnetic Resonance Series A* **1996**, *120*, 110-115.
[288] A. Bax and D. G. Davis, *Journal of Magnetic Resonance* **1985**, *65*, 355-360.
[289] J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *Journal of Chemical Physics* **1979**, *71*, 4546-4553.
[290] A. G. Palmer, J. Cavanagh, P. E. Wright and M. Rance, *Journal of Magnetic Resonance* **1991**, *93*, 151-170.
[291] H. Desvaux, P. Berthault and N. Birlirakis, *Chemical Physics Letters* **1995**, *233*, 545-549.
[292] M. Clark, R. D. Cramer and N. Van Opdenbosch, *Journal of Computational Chemistry* **1989**, *10*, 982-1012.
[293] H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *The Journal of Chemical Physics* **1984**, *81*, 3684-3690.
[294] T. A. D. D.A. Case, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R., R. C. W. Luo, W. Zhang, K.M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg,, I. K. G. Seabra, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell,, H. G. T. Steinbrecher, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui,

D.R. Roe, D.H., M. G. S. Mathews, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and and P. A. Kollman, *University of California, San Francisco* **2010**.
[295] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *Journal of Physical Chemistry* **1993**, *97*, 10269-10280.
[296] S. G. Gouin, E. Vanquelef, J. M. G. Fernández, C. O. Mellet, F. Y. Dupradeau and J. Kovensky, *Journal of Organic Chemistry* **2007**, *72*, 9032-9045.
[297] F. Y. Dupradeau, C. Cézard, R. Lelong, É. Stanislawiak, Pê, J. cher, J. C. Delepine and P. Cieplak, *Nucleic Acids Research* **2008**, *36*, D360-D367.
[298] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley and R. J. Woods, *Journal of Computational Chemistry* **2008**, *29*, 622-655.
[299] N. Sapay, E. Cabannes, M. Petitou and A. Imberty, *Glycobiology* **2011**, *21*, 1181-1193.
[300] T. A. D. D.A. Case, T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman, *University of California, San Francisco* **2012**.
[301] J. A. Lzaguirre, D. P. Catarello, J. M. Wozniak and R. D. Skeel, *Journal of Chemical Physics* **2001**, *114*, 2090-2098.
[302] a) J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics* **1977**, *23*, 327-341; b) S. Miyamoto and P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952-962.
[303] M. Mayer and B. Meyer, *Angewandte Chemie - International Edition* **1999**, *38*, 1784-1788.
[304] a) J. Jeener, B. H. Meier, P. Bachmann and R. R. Ernst, *The Journal of Chemical Physics* **1979**, *71*, 4546-4553; b) R. Wagner and S. Berger, *Journal of Magnetic Resonance - Series A* **1996**, *123*, 119-121.
[305] H. Liu, P. D. Thomas and T. L. James, *Journal of Magnetic Resonance (1969)* **1992**, *98*, 163-175.
[306] a) W. Sherman, H. S. Beard and R. Farid, *Chemical Biology & Drug Design* **2006**, *67*, 83-84; b) W. Sherman, T. Day, M. P. Jacobson, R. A. Friesner and R. Farid, *Journal of Medicinal Chemistry* **2005**, *49*, 534-553.
[307] J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham Iii, *Journal of Chemical Theory and Computation* **2007**, *3*, 2312-2334.
[308] W. Humphrey, A. Dalke and K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33-38.

Publications derived from the present thesis

1. *Effect of the Substituents of the Neighboring Ring in the Conformational Equilibrium of Iduronate in Heparin-like Trisaccharides*. Muñoz-García, J.C.; López-Prados, J.; Angulo, J.; Díaz-Contreras, I.; Reichardt, N.; de Paz, J.L.; Martín-Lomas, M.; Nieto, P.M. *Chemistry A-European Journal* **2012**, 18, 16319-16331.

2. *Conformations of the Iduronate Ring in Short Heparin Fragments Described by Time-averaged Distance Restrained Molecular Dynamics (tar-MD)*. Muñoz-García, J.C.; Corzana, F.; de Paz, J.L.; Angulo, J.; Nieto P.M. *Glycobiology* **2013**, accepted with revisions.

3. *Structures of Glycans Bound to Receptors from Saturation Transfer Difference (STD) NMR Spectroscopy: Quantitative Analysis by using CORCEMA-ST*. Enríquez-Navas, P.M; Guzzi, C.; Muñoz-Garcia, J.C.; Nieto, P.M.; Angulo J.* *Methods in Molecular Biology* **2013**, in press.

4. *NMR Studies on Carbohydrate Interactions with DC-SIGN Towards a Quantitative STD Analysis*. Guzzi, C.; Muñoz-García, J.C.; Enriquez-Navas, P.M.; Rojo, J.; Angulo, J.; Nieto, P.M. *Pure and Applied Chemistry* **2013**, invited contribution sent.

5. *3D Structure of a Heparin Mimetic Analogue to a FGF-1 Activator. A NMR and Molecular Modelling Study*. Muñoz-García, J.C.; Solera, C.; Carrero, P.; de Paz, J.L.; Angulo, J.; Nieto, P.M. In preparation.

6. *Molecular Recognition of Small Heparin-like Oligosaccharides by Langerin. Structural Bases from NMR Spectroscopy*. Muñoz-García, J.C.; Chabrol, E.; Thomas, A.; de Paz, J.L.; Rojo, J.; Fieschi, F.; Imberty, A.; Nieto, P.M.; Angulo, J. In preparation.

7. *NMR and Molecular Modelling Studies of the Interaction between Synthetic Hyaluronic Acid Disaccharides and Langerin*. Muñoz-García, J.C.; Macchione, G.; López-Prados, J.; Fieschi, F.; de Paz, J.L.; Angulo, J.; Nieto, P.M. In preparation