UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Química Orgánica y Farmacéutica

TESIS DOCTORAL

Computational studies of carbohydrate-protein interactions

Estudios computacionales de interacciones carbohidrato-proteína

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Alessandra Lacetera

Directora

Sonsoles Martín Santamaría

Madrid, 2018

FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA ORGÁNICA Y FARMACÉUTICA

COMPUTATIONAL STUDIES OF CARBOHYDRATE-PROTEIN INTERACTIONS

PhD Dissertation presented by

ALESSANDRA LACETERA

Madrid, 2017

This Thesis have been realized in Centro de Investigaciones Biológicas (CSIC)

Supervisor: Sonsoles Martín Santamaría

FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA ORGÁNICA Y FARMACÉUTICA

ESTUDIOS COMPUTACIONALES DE INTERACCIONES CARBOHIDRATO-PROTEÍNA

Memoria para optar el grado de Doctor presentada por

ALESSANDRA LACETERA

Madrid, 2017

Esta Tesis ha sido realizada en el Centro de Investigaciones Biológicas (CSIC)

Directora: Sonsoles Martín Santamaría

To my parents

Imagination is the discovering faculty, pre-eminently.
It is that which penetrates into the unseen worlds around us, the worlds of Science.

(Ada Lovelace)

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Dr. Sonsoles Martín-Santamaría for supervising my PhD Thesis work and for her support, for trusting in me and for giving me the opportunity to work in her group.

I would like also to thank Profs. Jesús Jimenéz-Barbero and Francisco Javier Cañada for teaching me NMR techniques during my secondment in their laboratories and for their collaboration over these four years.

I extend special thanks to Prof. Stefan Oscarson for giving me the opportunity to work in Dublin for a secondment and to learn carbohydrate synthesis. Also, I would like to thank him and Prof. José Carlos Menéndez for the synthesis of some compounds reported in this Thesis.

I also thank Dr. Margarita Menéndez for her help in the isothermal titration calorimetry assays.

I would like to thank the European Commission and the Spanish Ministry for Economy and competiveness for the funds over these years (Ref. FP7-MCA-ITN-GLYCOPHARM and CTQ2014-57141-R).

I would like to extend my acknowledgments to my local family that are more than colleagues Javi, Lucía, Joan, Laura and Jean-Marc that make the lab my home. I would like also to thank all the people come through the lab and my stay in Madrid that contribute to get this important step of my life and to support me and trusting in me.

Thanks also to all the friends spare in the world who have always made me feel their presence despite the distance.

And most important I would like to dedicate this Thesis to my parents, to make me feel the warming and their presence everyday although the distance that separates us.

CONTENTS

CONTENTS **i**

ABBREVIATIONS AND SUGAR **v**

INDEX ABSTRACT **ix**

RESUMEN **xiii**

CHAPTER 1. INTRODUCTION **1**

1.1 Carbohydrates 2

1.1.1 Monosaccharides 3

1.1.2 Disaccharides 5

1.1.3 Oligosaccharides, polysaccharides and others 7

1.1.4 Glycoproteins and glycolipids 7

1.1.5 Glycans in medicinal chemistry 9

1.2 Carbohydrate-protein interaction 20

1.2.1 Pea Lectin 22

1.2.2 Galectins 23

1.2.3 *Maackia amurensis* lectins 26

1.2.4 Glycosyltransferases 27

1.3 Computational techniques in drug design 28

1.3.1 *Ab initio* calculation and semi-empirical approaches 29

1.3.2 Docking 31

1.3.2.1 AutoDock4 35

1.3.2.2 Glide 37

1.3.2.3 VINA 38

1.3.3 Molecular Mechanics and Molecular Dynamics Simulations 39

1.3.3.1 AMBER ff14SB 43

1.3.3.2 GLYCAM06 44

1.3.4 MM-PBSA/MM-GBSA 45

1.3.5 Virtual Screening 47

1.3.5.1 FLAP 47

1.3.5.2 Glide 48

1.3.6 Conformational analysis 49

1.3.7 Homology modelling 50

1.4 Objectives 52

Bibliography 54

CHAPTER 2. MOLECULAR RECOGNITION STUDIES OF ^{19}F GLYCOMIMETICS TARGETING PEA LECTIN. A COMBINED NMR & COMPUTATIONAL APPROXIMATION 71

2.1. Introduction 71

2.2. Results 74

2.2.1 The free state: conformational analysis by MM3* calculations and *ab initio* calculation, Molecular Dynamics simulations, and comparison to the NMR experimental data 76

2.2.2 The bound state: integrating NMR results, docking calculations and MD simulations 81

2.3. Material and Methods 92

Bibliography 95

Annex II 98

CHAPTER 3. STRUCTURAL INSIGHTS ON THE RECOGNITION OF THE ALPHA-TYPE HISTO-BLOOD ALPHA-GAL EPITOPE BY GALECTIN-3 101

3.1. Introduction 101

3.2. Results 104

3.2.1. The recognition of LacdiNAc by Gal-3 104

3.2.2. The recognition of the α-Gal epitope by Gal-3 108

3.2.3. The recognition of blood group A type II tetrasaccharide 115

3.3. Material and Methods 118

Bibliography 121

Annex III 123

CHAPTER 4. COMPUTATIONALLY-DESIGNED GALECTIN BINDERS WITH HIGH-AFFINITY PROFILE **127**

4.1 Introduction 127

4.2 Results 130

4.2.1 Design of selective galectin ligands 130

4.2.2 Virtual screening of a fragment library 131

4.2.3 Docking of the designed methyl-lactose derivatives 132

4.3 Material and methods 134

Bibliography 139

CHAPTER 5. TARGETING SIALIC ACID-MODIFIED RECEPTORS AS A POTENTIAL THERAPY FOR OSTEOARTHRITIS **143**

5.1. Introduction 143

5.2. Results 146

5.2.1. Molecular binding characteristics of the interaction of PDPN with MASL 146

5.2.2. α-2,3-sialylated glycoproteins are induced in arthritic chondrocytes 148

5.2.3. PDPN and MASL: structural background for the study of sialyllactose interaction 149

5.2.4. Computational studies of the PDPN/MASL recognition 152

5.3. Material and Methods 156

Bibliography 160

CHAPTER 6. GLYCOMIMETICS TARGETING GLYCOSYLTRANSFERASES: SYNTHETIC, COMPUTATIONAL AND STRUCTURAL STUDIES OF LESS-POLAR CONJUGATES **165**

6.1. Introduction 165

6.2. Results 169

6.2.1. Crystallographic and binding studies 169

6.2.2. Docking Studies 173

6.2.3. tMD simulation 176

6.3. Material and Methods 180

Bibliography 182

Annex VI 186

CONCLUSIONS **187**

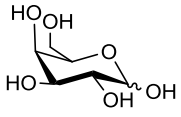
CONCLUSIONES **189**

ABBREVIATIONS

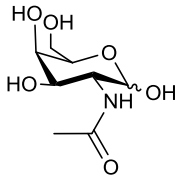
bNAbs	broadly Neutralizing Antibodies
CBAs	Carbohydrate-Binding Agents
CEA	CarcinoEmbryonic Antigen
CPS	Capsular Polysaccharides
CRD	Carbohydrate Recognition Domain
DC-SIGN	Dendritic-Cell Specific ICAM-3 Grabbing Nonintegrin
DFT	Density Functional Theory
DNA	Deoxyribonucleic acid
EA	Evolutionary Algorithms
EPO	Erythropoietin
FA	Fatty Acids
FF	Force Field
FLAP	Fingerprints for Ligands And Proteins
GA	Genetic Algorithm
GAGs	Glycosaminoglycans
GB/SA	Generalized-Born/Surface Area
GBPs	Glycan Binding Protein
gg	*gauche-gauche*
gt	*gauche-trans*
GTO	Gaussian Type Orbitals
GTs	Glycosyltransferases
HIV	Human Immunodeficiency Virus
KB	Knowledge-Based
LOS	Lipooligosaccharides
LPS	Lipopolysaccharides
LS	Local Search
LSDs	Lysosomal Storage Disorders
mAb	monoclonal Antibody
MAH	*Maackia amurensis* Hemagglutinin
MAL	*Maackia amurensis* Leukoagglutinin
MC	Monte Carlo
MD	Molecular Dynamics
MD-2	Myeloid Differentiation factor 2
MetS	Metabolic Syndrome
MIFs	Molecular Interaction Fields
MM	Molecular Mechanics
MM/PBSA	Molecular Mechanics/Poisson-Boltzmann Surface Area
MMGBSA	Molecular Mechanics/Generalized Born Surface Area
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance

Nod	Nodulation
PB/SA	Poisson-Boltzmann/Surface Area
PDB	*Protein Data Bank*
PME	Particle Mesh Ewald
PSA	*Pisum sativum*
QM	Quantum Mechanics
QM/MM	Quantum Mechanics/Molecular Mechanics
QMC	Quantum Monte Carlo
RBVS	Receptor-Based Virtual Screening
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
SASA	Solvent Accessible Surface Area
SLea	Sialyl Lewis a
SLex	Sialyl Lewis x
SO	Swarm Optimization
STD	Saturation-Transfer Difference
STn	Sialyl Tn antigen
STO	Slater Type Orbitals
TACAs	Tumour-Associated Carbohydrate Antigens
tg	*trans-gauche*
TI-2	T-cell Independent type 2
TLR4	Toll-like Receptor 4
UDP	Uridine diphosphate
VINA	Vina Is Not AutoDock
VS	Virtual screening

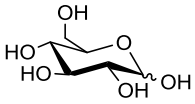
SUGAR ABBREVIATIONS



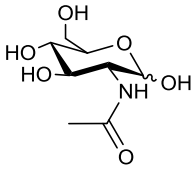
Galactose (Gal)



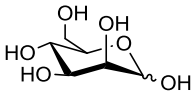
N-Acetil-Galactosamine (GalNAc)



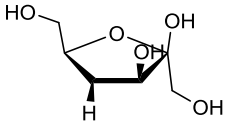
Glucose (Glc)



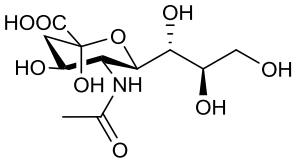
N-Acetil-Glucosamine (GlcNAc)



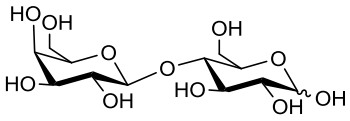
D-Mannose (Man)



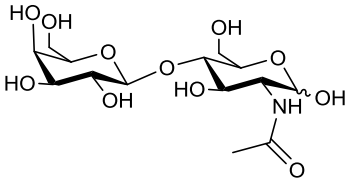
D-Fructose (Fru)



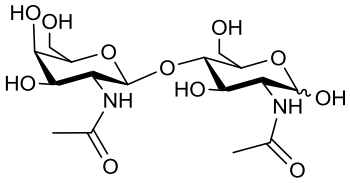
N-Acetyl-Neuraminic acid (Neu5Ac)



Lactose (Lac)



N-Acetyl-lactosamine (LacNAc)



N,N’-diacetyl-lactosamine (LacdiNAc)

ABSTRACT

Molecular modelling techniques have been extensively applied to the elucidation of glycan-protein interactions at the atomic level of several systems with biological and therapeutic interests, and have provided new insights for the understanding of the molecular recognition events underlying the biological functions of these systems. In this Thesis, the general objective is to elucidate the carbohydrate-protein interactions at the atomic level through computational techniques. In particular, the following systems have been studied: human galectins -1, -3, and -7, *Pisum sativum* lectin, *Maackia amurensis* seed lectin, and glycosyltransferase GalNAc-T2. The final aim is to provide new insights for the understanding of the molecular recognition events underlying the biological functions of these proteins.

Chapter 2: A pluridisciplinary approach, in a collaborative work with experimental groups, has been carried out for the study of carbohydrate-protein interaction. In particular, we seek the elucidation of the binding modes of a fluorinated trimannoside α1→3 and α1→6 linked and its fluorinated disaccharide derivatives with a model lectin from *Pisum sativum,* combining computational studies with ^{19}F-NMR experimental data. This combined study has required the use of synthetic fluorinated complex glycan mimics to provide detailed structural and binding information for every sugar ring, allowing the elucidation of the interacting ligand epitope and the bound conformations, not available when using non-fluorinated sugars. Computational techniques have been proved to assist the ^{19}F NMR experiments and to provide the necessary 3D perspective for the elucidation of this example of molecular recognition of high complexity. Specifically, the combination of the modelling and NMR experimental data indicates the co-existence of two different binding modes for the trifluorinated trimannoside when interacting with the plant lectin from *Pisum sativum*. This example accounts for the complementarity between these techniques for the elucidation of complex molecular recognition problems where experimental techniques cannot provide a full response.

Chapter 3: Computational study of the binding mode of several β-galactoside ligands (lactose, LacdiNAc, and three glycans containing the LacNAc motif: blood group A type II tetrasaccharide, and two compounds derived from the important xenoantigen α-Gal epitope) towards human galectin-3 has been carried out. These ligands contain the epitopes LacdiNAc and α-Gal, proposed as molecular patterns for galectin-3 mediated immune response. We seek the characterization of the key interactions required for the distinct recognition by galectin-3, pointing out to the subtle differences in the recognition event, by means of combined molecular modelling and NMR techniques. Using this combined approach, we here report the ligand-receptor interactions accounting for the gal-3 recognition of the important xenoantigen α-Gal epitope, showing the important role of gal-3 in the initial recognition and adhesion of human monocytes to porcine aortic endothelial cells which may contribute to delayed xenograft rejection. Our findings have unraveled important details at atomic level of the gal-3 recognition.

Chapter 4: Computational design of novel selective galectin binders with predicted improved affinity has been performed. In particular, we have focused our efforts in adhesion/growth-regulatory galectins -1, -3, and -7, whose functions have been shown to be correlated among them in several tumoral processes. Our computational studies have been focused on the pocket adjacent to the CRD as additional anchorage point, and have provided structural basis for a fragment-based design of novel galectin binders with predicted selectivity and high affinity. With this approach, we identified the appropriate fragments/moieties able to be anchored to this pocket. In collaboration with other groups, selected compounds have been synthesized and preliminary NMR and ITC results are very promising. These selective ligands may contribute to the understanding of the highly relevant biological functions of galectins, and their role in cancer.

Chapter 5: A computational model for the interaction of the podoplanin (PDPN) epitope (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) and *Maackia amurensis* seed lectin (MASL) isoforms has accounted for the protective action of these lectins against cartilage degradation in osteoarthritis (OA). We explored the binding site of the two MASL isoforms, *Maackia amurensis* leukoagglutinin (MAL) and *Maackia amurensis* hemagglutinin (MAH), and the possible recognition by the epitope of the glycoprotein podoplanin. The ability of MASL to target sialylated glycoproteins such as PDPN, which is also induced during OA and rheumatoid arthritis, and to protect chondrocytes from insults leading to cartilage degradation might offer further possibilities for therapeutic interventions and novel arthritis treatments that may include the regulation of sialylation during acute disease stages. In particular, the docking of the tetrasaccharide (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) shows the preference of this ligand to bind the MAH isoform, due to the presence of disialyl-core and the specific Galβ1-3GalNAc glycosidic linkage. The experimental results indicate that MASL preserves the structure and function of cartilage under diverse arthritic insults by interfering with the function of α-2,3-sialylated transmembrane receptors, such as the mucin-type transmembrane glycoprotein PDPN. Moreover, these findings suggest that MASL inhibits the activation of signal transduction pathways that lead to progressive cartilage destruction during the pathogenesis of arthritis by increasing reactive oxygen species, inflammatory cytokines, and metalloproteinases. We propose the direct interaction of MASL with PDPN as a plausible mechanism for this protective activity of the lectin, and our computational studies have provided a 3D molecular model for such interaction.

Chapter 6: We have studied the binding mode of a series of synthetic nucleoside mimetics as inhibitors of a model glycosyltransferase, the GalNAc-T2. These less-polar nucleotide sugar analogues, derived from uridine only containing the β-phosphate, could be efficient ligands for the enzyme, as deduced from the experimental data. We have proposed a computational 3D model for the binding mode of these ligands to provide 3D clues regarding the mechanism of the catalytic cycle of this family of enzymes. In particular, from the docking studies, it is inferred that the β-phosphate is required for binding to the Mn^{+2}, and thus the replacing of the β-phosphate while keeping the phosphate or alike groups in the α position should maintain the binding of these compounds. The docking study suggests that the less polar uridyl-sugar analogues are putative GalNAc-T2 binders where the uridine moiety plays a major role in the binding in agreement with the crystallographic binding poses, while the phosphonate group and the sugar moiety provide additional interactions, not finding differences among the type of sugar. Additionally, the tMD simulation has given us an approximation to the binding pose changes that the UDP-GalNAc suffers in the closing/opening process when bound to the GT. The specific interactions for the GalNAc in the closing/opening process may be useful to understand the catalytic mechanism cycle of this enzyme, and to help for the designing of selective GT binders. In fact, we provide a plausible explanation for the role of the *N*-acetylamido group in the binding, which supports its higher affinity to the enzyme in comparison to UDP-Gal.

Summarizing, molecular modelling techniques have been extensively applied to the elucidation of glycan-protein interactions at the atomic level of several systems with biological interest, and have provided new insights for the understanding of the molecular recognition events underlying the biological functions of these systems.

RESUMEN

En esta Tesis Doctoral, se han empleado técnicas de modelado molecular para la elucidación a nivel atómico de las interacciones carbohidrato-proteína de varios sistemas con interés biológico y terapéutico. Estos estudios han permitido aportar nuevas perspectivas para la comprensión de los procesos de reconocimiento molecular implicados en las funciones biológicas de estos sistemas. En particular, se han estudiado los siguientes sistemas: galectinas humanas -1, -3 y -7, lectina de *Pisum sativum*, lectina de semillas de *Maackia amurensis*, y la glicosiltransferasa GalNAc-T2. El objetivo final es proporcionar nuevas perspectivas para la comprensión de los eventos de reconocimiento molecular que subyacen en las funciones biológicas de estas proteínas.

Capítulo 2: Se ha llevado a cabo un estudio pluridisciplinar, en colaboración con grupos experimentales, abordándose el estudio de los modos de unión de una trimanosida trifluorada, con enlaces α1 → 3 y α1 → 6, y sus derivados disacáridos fluorados, con una lectina modelo de *Pisum sativum*, combinando los estudios computacionales con datos experimentales de ^{19}F-RMN. Para llevar a cabo este estudio combinado, se han necesitado glicómiméticos fluorados sintéticos para la elucidación de detalles estructurales y detalles de la interacción de cada manosa fluorada con la proteína, identificando el epítopo y las conformaciones de unión. Este tipo de estudio no hubiera sido posible sin la presencia del flúor. Las técnicas computacionales han ayudado de forma importante a los experimentos de ^{19}F-RMN, y han proporcionado la perspectiva 3D necesaria para la elucidación de este ejemplo de reconocimiento molecular de gran complejidad. Específicamente, la combinación de los datos experimentales de modelado y los datos de RMN ha indicado la coexistencia de dos modos de unión diferentes para la trimanosida trifluorada cuando se une a la lectina de *Pisum sativum*. Este ejemplo ilustra la complementariedad entre estas dos técnicas para la elucidación de problemas complejos de reconocimiento molecular donde las técnicas experimentales no pueden proporcionar una respuesta completa.

Capítulo 3: Estudio computacional de la unión de varios ligandos de β-galactósido (lactosa, LacdiNAc y tres glicanos que contienen el motivo LacNAc: tetrasacárido del grupo sanguíneo A tipo II, y dos compuestos derivados del importante epítopo α-Gal de xenoantigeno) hacia la galectina-3 humana. Los epítopos LacdiNAc and α-Gal contienen el motivo LacNAc, que ha sido propuesto como patrón molecular para la respuesta inmune mediada por galectina-3. Se han caracterizado las interacciones clave necesarias para el reconocimiento distinto por la galectina-3, señalando las sutiles diferencias en el evento de reconocimiento, mediante técnicas combinadas de modelado molecular y RMN. Utilizando este enfoque combinado, aquí presentamos las interacciones ligando-receptor que representan el reconocimiento gal-3 del epítopo importante de xenoantógeno α-Gal, mostrando el importante papel de gal-3 en el reconocimiento inicial y adhesión de monocitos humanos a endotelio aórtico porcino, que pueden contribuir al rechazo tardío del xenoinjerto. Nuestros estudios han aportado importantes detalles a nivel atómico del reconocimiento gal-3.

Capítulo 4: Se ha realizado el diseño computacional de nuevos ligandos selectivos de galectinas con afinidad mejorada predicha. En particular, hemos centrado nuestros estudios en las galectinas -1, -3 y -7, involucradas en la adhesión, crecimiento y regulación celular. Además se ha demostrado que las galectinas -1, -3 y -7 tienen diferentes papeles en varios procesos tumorales. Nuestros estudios computacionales se han dirigido al bolsillo de unión adyacente al dominio de reconocimiento de carbohidratos, y nos han proporcionado una base estructural para llevar a cabo un diseño basado en fragmentos de nuevos ligandos de galectina con selectividad y alta afinidad. Con esta aproximación, identificamos los mejores fragmentos capaces de anclarse a este bolsillo en cada una de las galectinas. En colaboración con otros grupos, se han sintetizado algunos compuestos seleccionados y los resultados preliminares de RMN e ITC son muy prometedores. Estos ligandos selectivos pueden contribuir a la comprensión de las funciones biológicas de las galectinas, altamente relevantes, y su papel en el cáncer.

Capítulo 5: Se ha propuesto un modelo computacional para la interacción del epítopo de podoplanina (PDPN) (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) con las isoformas de lectina de semilla de *Maackia amurensis* (MASL), explicando la acción protectora de estas lectinas en la degradación del cartílago en la osteoartritis (OA). Hemos explorado el sitio de unión de las dos especies moleculares de la MASL, la leucoaglutinina de *Maackia amurensis* (MAL) y la hemaglutinina de *Maackia amurensis* (MAH), y el posible reconocimiento por el epítopo de la glicoproteína podoplanina. La capacidad de MASL de unirse a glicoproteínas sialiladas como la podoplanina, que se inducen durante la osteoartritis y la artritis reumatoide, protege a los condrocitos de insultos que conducen a la degradación del cartílago. Este mecanismo puede suponer nuevas formas de intervención terapéutica y nuevos tratamientos para la artritis que incluyan la regulación de sialilación durante estadios agudos de la enfermedad. En particular, la unión del tetrasacárido (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) muestra la preferencia de este ligando para unirse a la isoforma de la MAH, debido a la presencia de dos unidades de acido siálico y a la presencia del enlace glucosídico específico Galβ1-3GalNAc. Los resultados experimentales indican que la MASL preserva la estructura y la función del cartílago bajo diversos insultos artríticos interfiriendo con la función de los receptores transmembrana α-2,3-sialilados, tales como la glicoproteína transmembrana podinaplanina de tipo mucina (PDPN). Por otra parte, estudios sugieren que MASL inhibe la activación de las vías de transducción de señales que conducen a la destrucción progresiva del cartílago durante la patogénesis de la artritis mediante el aumento de las especies de oxígeno reactivo (ROS), citoquinas inflamatorias y metaloproteinasas. Proponemos la interacción directa de MASL con PDPN como un mecanismo plausible para esta actividad protectora de la lectina, y nuestros estudios computacionales han proporcionado un modelo molecular 3D para dicha interacción.

Capítulo 6: Se ha estudiado el modo de unión de una serie de miméticos sintéticos de nucleósidos como inhibidores de una glicosiltransferasa (GT) modelo, la GalNAc-T2. Estos análogos de nucleótidos son compuestos menos polares derivados de la uridina que contiene sólo el β-fosfato y podrían ser ligandos eficaces para la inhibición de la enzima, como se deduce de los datos experimentales. Hemos propuesto un modelo 3D computacional para el modo de unión de estos ligandos para proporcionar más información sobre el mecanismo del ciclo catalítico de esta familia de enzimas. En particular, a partir de los estudios de modelado molecular se deduce que el β-fosfato no es necesario para la unión al Mn^{+2}, y por lo tanto la sustitución del α-fosfato, manteniendo el fosfato o grupos similares en la posición β, debe mantener la unión de estos compuestos. Los estudios de docking sugieren que los análogos menos polares del azúcar uridilo son posibles ligandos de GalNAc-T2, donde la parte de la uridina desempeña un papel principal en el enlace, de acuerdo además con los modos de unión observados en las estructuras cristalográficas. El grupo fosfonato y el azúcar establecen interacciones adicionales, no encontrándose diferencias al estudiar otros tipos de azúcar. Además, la simulación de tMD nos ha dado ha permitido estudiar los cambios en la pose de unión a la GT que el UDP-GalNAc sufre en el proceso de cierre/apertura. Las interacciones específicas de GalNAc en este proceso de cierre/apertura pueden ser útiles entender el ciclo del mecanismo catalítico de esta enzima y para el diseño de ligandos selectivos de GT. De hecho, nos han proporcionado una explicación plausible para el papel del grupo N-acetilamido en el enlace, que apoya su mayor afinidad a la enzima en comparación con la UDP-Gal.

Resumiendo, se han empleado técnicas de modelado molecular a la elucidación a nivel atómico de las interacciones glicano-proteína de varios sistemas con interés biológico. Estos estudios han proporcionado nuevos conocimientos para la comprensión de eventos de reconocimiento molecular subyacentes a las funciones biológicas de estos sistemas.

CHAPTER 1

INTRODUCTION

In biological systems, cells communicate each other and transmit signals to other cells through mediators of different chemical nature. During the years, the most studied molecules were the nucleic acids and proteins. However, this view has changed and also lipids and carbohydrates have been included in this scenario.[1] Carbohydrates (named also sugars or saccharides) are the most abundant biomolecules in nature and they are often conjugated with proteins (glycoproteins) or lipids (glycolipids). These molecules play different roles in the cells, and for this reason recent studies are focused on carbohydrates as selective mediator in the molecular recognition processes. They are involved in physiological and pathological events and in a variety of intra- and extra-cellular events. In most of the cases, the initial contacts of saccharides with cells take place at cell-cell and cell-matrix level.[2] An important and rather complex process is the glycosilation. The glycosilation, the enzymatic process that links glycans with proteins or lipids, is usually rather heterogeneous and it strictly depends on the activity and/or availability of specific enzymes, such as glycosyl transferases and glycosyl hydrolases (glycosidases). Modifications on the regular cell glycosylation pattern are often observed during pathological conditions, for example in inflammation and cancer development. Therefore, glycans have become a topic of major interest to develop diagnostic tools or therapeutic drugs.[3] To date, several sugar-based drugs have already been developed and currently employed as therapeutics, diagnostics, and vaccines and many more are expected to come in the next few years.

1.1 Carbohydrates

As mentioned above, carbohydrates are very abundant in nature and they play numerous biological roles. They are the fundamental source of metabolic energy, in fact they represent an important element in our diet; some insoluble carbohydrate polymers serve to lubricate skeletal joints and provide adhesion between cells and others serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues and cell coats of animals. Sugars can exist in nature as single entities or as a forming part of glycoconjugates, essentially glycolipids and glycoproteins. Chemically, carbohydrates are polyhydroxy aldehydes or ketones, and their chemical formula is usually $C_n(H_2O)_n$ (carbon "hydrates"), but some "natural" carbohydrates can also contain nitrogen, phosphorus, sulphur. Their chemical variability is enormous, due to the possibilities of branching at either hydroxyl group as well as the possible presence of pendant substituents, either neutral or charged.[1, 4] In turn, this variability generates a huge possibility of constitutional isomers and a plethora of conformations. Moreover, their chemical properties in terms of the presence of polar and non-polar patches make them to exhibit peculiar stability and, acting as ligands for receptors, to be fairly adaptable to a variety of environments. They may adopt different shapes and display rather distinct conformational and dynamic properties. Glycan diversity depends on the monosaccharides (basic unit of saccharide) which are made by, on the linkages that connect monosaccharides (glycosidic linkage), and on other factors such as the *anomeric effects*, the orientation of all the torsional angles,[5-6] and the intramolecular hydrogen bonding between adjacent OH groups.[7] Therefore, it is of paramount importance to predict and characterize their conformational and structural properties in a systematic manner. Computational techniques have proven to be essential in this aspect.[8]

1.1.1 Monosaccharides

Carbohydrates are made by small constituents, named monosaccharides (or simply sugars). Monosaccharides are composed of a single polyhydroxy aldehyde or ketone unit and may have different number of carbons normally from five to nine. The most abundant monosaccharide in nature is D-glucose. Usually the carbons of a monosaccharide have hydroxyl groups attached and they are normally stereogenic centres.

Monosaccharides of six carbons (pyranose) usually adopt chair conformations and the substituents can be found either in the axial or equatorial position. In pyranose sugars, if there is a hydroxyl group at the anomeric carbon (the hemiacetal/acetal carbon) the monosaccharide could assume α or β configuration (Figure 1.1-right). α and β Isomers are both optically active and usually present similar physical properties. If the substituent on the anomeric carbon is electronegative (like a methoxy group) there is a preference of the substituent to adopt an axial configuration rather than an equatorial one. This phenomenon is called *anomeric effect* that is contrasted by steric factors, reason why D-sugars present more stability as α isomer (Figure 1.1-left). If the hydroxyl group is attached to a different carbon from the anomeric one the effect of the different position leads to a different monosaccharide (epimer) with completely different properties. For example, glucose and mannose differ on the position of the OH group in position 2 (in equatorial position in glucose and in axial in mannose) and glucose and galactose on the OH of position 4 (in equatorial position in glucose and in axial in galactose) (Figure 1.1-left).

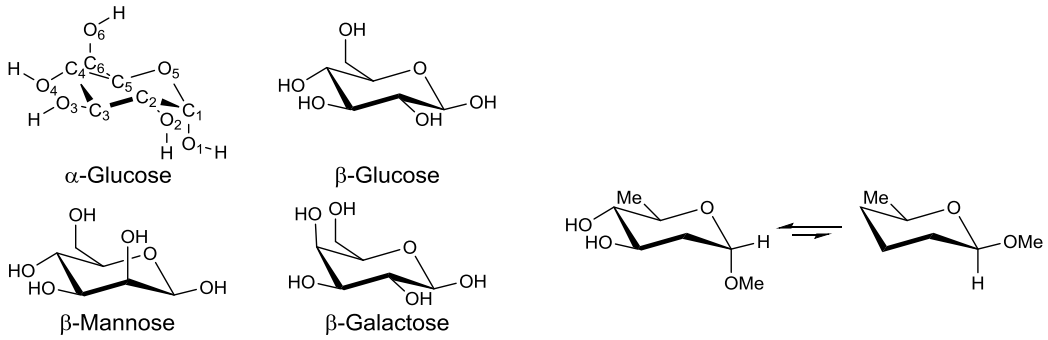


Figure 1.1. Left: α and β configuration of glucose and epimers of the β-glucose on C2 (mannose) and C4 (galactose). Right: anomeric effect is shown. The presence of a methoxy group on the anomeric carbon determines the preference of the axial position (α).

Moreover, simple monosaccharides may adopt distinct three-dimensional shapes. The most common shape for a pyranose is the chair conformation that minimizes steric congestion and includes the *anomeric effect*. However, also other shapes are allowed, depending on the orientation, nature and number of substituents. The most prevalent alternative shapes for a pyranose are the boat, the skew and the envelope (Figure 1.2).

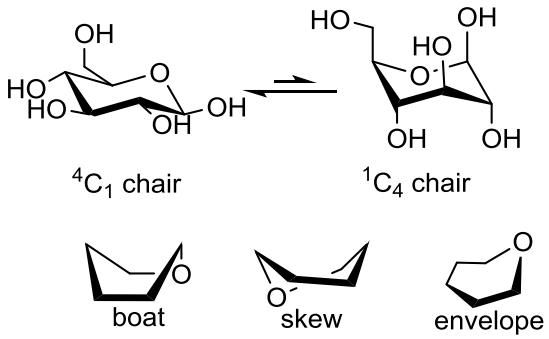


Figure 1.2. Conformational shapes of the monosaccharide β-D-glucopyranose. Top: regular chair conformations. Bottom: other prevalent conformers.

Regarding the conformation of the dihedral angle (ω) around C5-C6 in a pyranose ring (ω formed by O5C5C6O6, see Figure 1.1), a combination of solvation, steric, and stereoelectronic (*gauche effect*) effects drive the observed geometries. It has been monitored that for glucose-type sugars, a combination of *gauche-gauche* (*gg*) and *gauche-trans* (*gt*) rotamers coexists, while for galactose-type, ω prefers *gauche-trans* (*gt*) and *trans-gauche* (*tg*) geometries. This experimental behaviour has also been validated using computational protocols (Figure 1.3).[9]

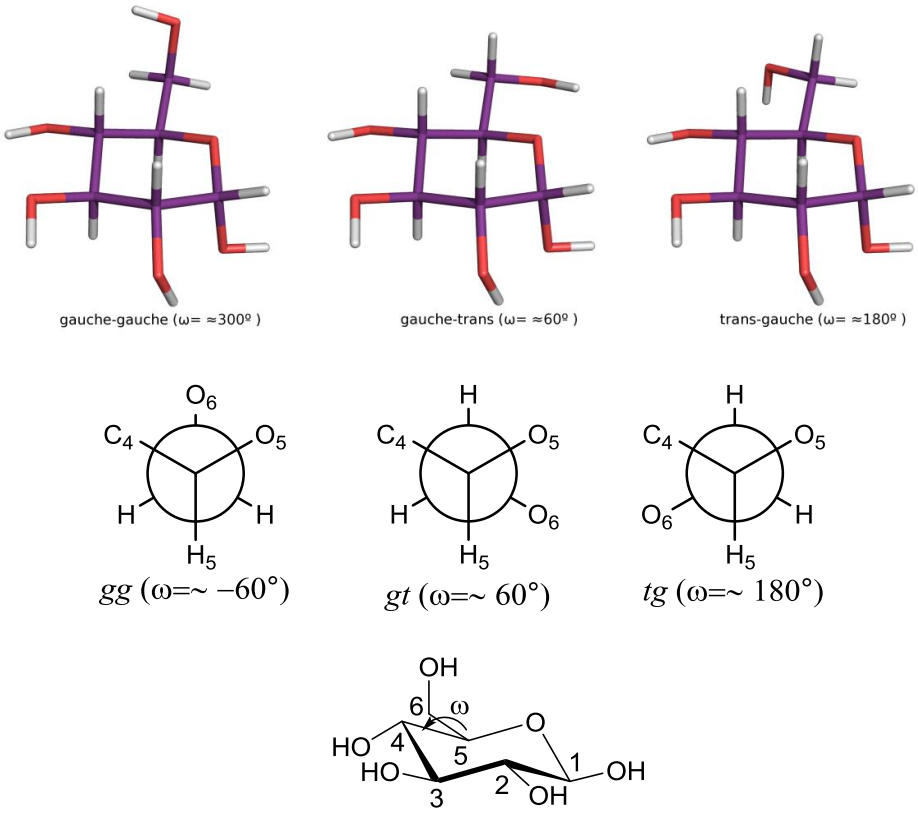


Figure 1.3. Possible conformation of ω (O5C5C6O6) in α-glucose.

1.1.2 Disaccharides

Monosaccharides are attached among them through glycosidic linkage. The anomeric oxygen of one particular monosaccharide is attached to a secondary (or primary) hydroxyl group of a second sugar moiety to build one disaccharide. The glycosidic linkage is an acetal linkage with particular chemical properties (sensitive to strong acid and basic conditions) and conformational features.

If the substituent at the anomeric carbon shows lone pair electrons, as an alkoxy group, retrodonation can proceed from the exocyclic oxygen to the sigma anti-bonding orbital of the intracyclic bond C-O. This is known as the *exo-anomeric effect*. In this case, the lone-pair orbital has to be anti-periplanar to the antibonding orbital of C-O, and consequently the alkyl substituent at the glycosidic oxygen adopts a *syn*-type orientation. This molecular orbital theory is able to predict also the modification of the bond length. In fact, for equatorially substituted pyranosides, the intracyclic C-O bond is longer than the exocyclic C-O bond (Figure 1.4).

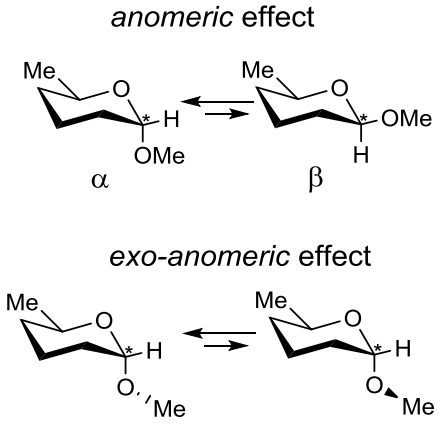


Figure 1.4. Representation of the anomeric effect and the exo-anomeric effect.

Actually, these preferences lead to a *syn*-type conformation, although it has been observed that a minor proportion of *anti*- conformation may co-exist in β-glycosides (e.g. in 1→2, 1→3 and 1→4 linkages) provided that the contiguous OH-2 displays an equatorial orientation (i.e., glucose or galactose). The *anti*-type geometry is also favoured by the *exo-anomeric effect*. Combined modelling/Nuclear Magnetic Resonance (NMR) approaches are mandatory to characterize the glycosidic linkage of a particular disaccharide or oligo-saccharide (either α and β, either 1→2, 1→3, 1→4, or 1→6).[9-10] All the particular features mentioned above for saccharides, besides their intrinsic dynamic properties require the synergy combination of rigorous experimental and theoretical protocols.[11]

1.1.3 Oligosaccharides, polysaccharides and others

Oligosaccharides are composed of a small number of monosaccharides. Their conformational properties, adaptability and recognition features have been, and still are, deeply analysed.[12] Chemical methods have been employed to provide stable chemical analogues of oligosaccharides, with increased resistance to glycosyl hydrolases. Typically, either the glycosidic (inter-unit) or the endocyclic oxygen within the ring is substituted by other chemical element. Therefore, *C*-,[13] *S*-,[14] *Se*-,[15] and *N*-[16] glycosyl analogues have been synthesized to provide chemical probes for interaction studies or to prepare enzyme inhibitors. Given the different chemical properties of these analogues, their conformational properties have been carefully analysed and compared to those of their parent *O*-glycosides.[17] Computational studies have been directed towards the analysis of the flexibility of these new linked glycosides, and the preferences of the binding proteins to select the proper "binding pose" from the conformational assemble, not always being the most stable conformer in terms of potential energy.[18-21]

1.1.4 Glycoproteins and glycolipids

In glycoproteins, oligosaccharides are present as *N*-glycans attached to an asparagine of the protein or as *O*-glycans attached to a threonine or a serine. The different substitution patterns make them to be named as displaying antennas. In fact, depending of the degree of branched substituents, they are bi-antennary, tri-antennary and even tetra-antennary. The different substitutions provide these molecules with different conformational and dynamic features,[22] which can be exploited for recognition. Additional core substitutions at certain positions might modify the dynamics of the different arms in a rather dramatic way. These aspects have been addressed in computational studies of *N*-glycans[22-23] in parallel with novel NMR methods to deduce their conformational behaviour.[24-27]

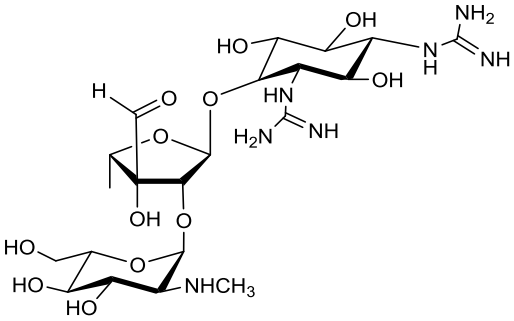
Glycolipids are amphiphilic molecules composed by a polar sugar part attached to lipophilic lipid chains. They display rather distinct structural and conformational properties. In this regard, the conformational properties of different lipooligosaccharides (LOS) and lipopolysaccharides (LPS) from different sources can be accessed by modelling protocols, in combination with NMR methods.[28-29] For example, nodulation (Nod) factors are lipochitooligosaccharides incorporing *N*-acetylglucosamines that mediate the recognition between *Rhizobium bacteria* and their legume symbionts. Further modifications, such as methylations, sulfations, and fatty acids, account for the specific recognition between the interacting partners. The structure of Nod factors and their recognition properties by receptors have been exhaustively modelled *in silico*.[14, 30-31]

Among the class of polysaccharides the most known are the glycosaminoglycans (GAGs). They are widely spread in nature and they encompass a variety of sulphated saccharides. GAGs are involved in several molecular recognition and signalling events. The relevancy of GAGs is owed to their recognition by different receptors, triggering a plethora of biological responses, including inflammation, cell adhesion, and regulation of cell growth and proliferation.[32-33] They are composed of different monosaccharide residues, either positively (rarely) or negatively charged (sulfate or carboxyl groups) or neutral and aminosugars, frequently acylated. Within this extensive family, heparin and heparan sulfate have been employed as therapeutic agents. Heparin and heparin sulfate are linear polysaccharides consisting of a glucuronic or iduronic acid, fragment that is 1→4 linked to a glucosamine moiety.[34] The structure of these molecules has been extensively studied using a variety of experimental (preeminently in the shape of NMR restraints) and modelling methods.[35-36] Special mention should be given to the "heparin pentasaccharide" one of the carbohydrate-based drugs in clinic, used as anti-thrombotic agent.[37-38] The study of the conformational properties of this molecule is one paradigmatic example of the employment of computational methods and NMR, both in its free and bound state to a variety of receptors.

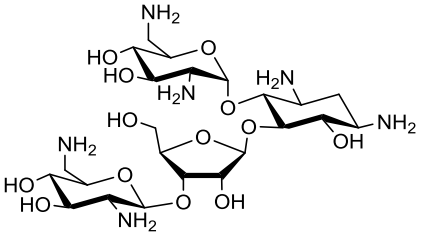
As a token, the application of both protein- and ligand-detected NMR methods, assisted by computational protocols have allowed for the assessment of the binding mode of the heparin pentasaccharide and related GAGs to fibroblast growth factors,[39] as well as the conformation of the GAGs in the bound state.[40] These studies permitted to identify the bioactive structure of the complex, and to reveal a receptor-driven conformational selection of particular iduronic acid ring puckers. Also plant polysaccharides have many possible geometries, shapes and therefore, physical and chemical properties. Many of them have been extensively studied, such as pectins, which contain L-rhamnose and arabinose and have complex branched structures,[41-42] and others, such as chitin $(GlcNAc\beta1\text{-}4GlcNAc)_n$, which display a linear shape.[43]

1.1.5 Glycans in medicinal chemistry

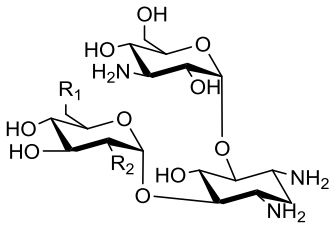
Several carbohydrate-based drugs are currently being used to treat several diseases, the most illustrative examples are antibiotics and antivirals.[44] For decades, aminoglycosides, inhibitors of bacterial protein synthesis, have been used as antibiotic especially for infections by aerobic Gram-negative bacteria.[45] Streptomycin (Figure 1.5) was the first to be discovered and it was used for the treatment of tuberculosis.[46] Chemically, it contains an aminocyclitol moiety that is linked to a carbohydrate skeleton. Other chemically related molecules have been discovered/synthesized, among them framycetin, employed for the treatment of ophthalmic infections, paromomycin used for the treatment of acute and chronic intestinal amebiasis, kanamycin and its derivative amikacin, for short-term treatment of serious infections, gentamicin,[47] tobramycin, specially indicated for *Pseudomonas aeruginosa* lung infections, septicemia or used in combination with other antibiotics and/or anti-inflammatory glucocorticoids for the treatment of different *Pseudomonas* spp., ophthalmic, and respiratory infections, including those associated with cystic fibrosis (Figure 1.5).

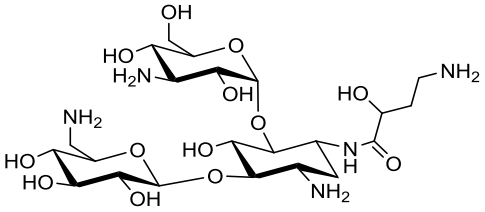


Streptomycin



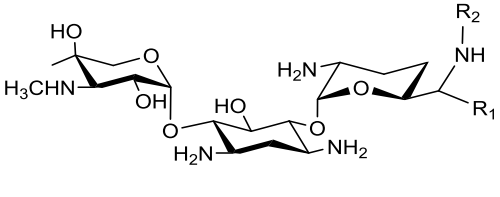
Framycetin

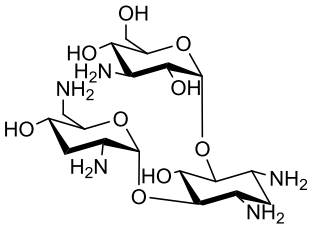




Amikacin

	R_1	R_2
Kanamycin A	NH_2	OH
Kanamycin B	NH_2	NH_2
Kanamycin C	OH	NH_2



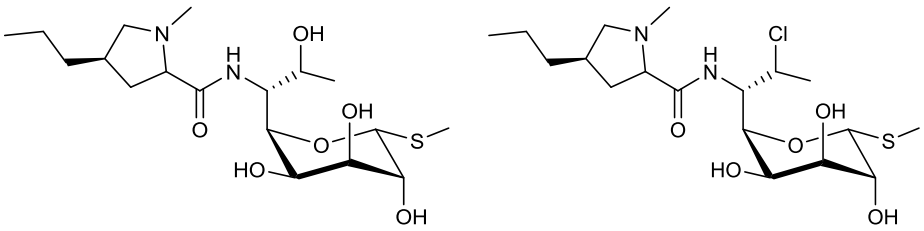


Tobramycin

	R_1	R_2
Gentamicin C1	CH_3	CH_3
Gentamicin C2	CH_3	H
Gentamicin C1a	H	H

Figure 1.5. Chemical structure of several aminoglycosides.

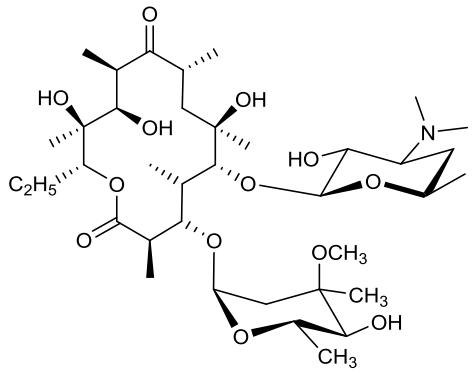
Also lincosamides contain a sugar moiety (1-thio-D-erythro-α-D-galactopyranoside). They are also inhibitors of protein synthesis of bacteria, binding to the 50S ribosomal subunit of the bacteria. The first lincosamide to be discovered was lincomycin (Figure 1.6), isolated from *Streptomyces lincolnensis*, but it was afterward replaced by its semisynthetic derivative, clindamycin (Figure 1.6). Clindamycin improved lincomycin side effect profile and gave the possibility of topical or systemic treatments. It is effective for the treatment of serious infections caused by susceptible anaerobic bacteria as well as in poly-microbic infections such as intra-abdominal or pelvic infections, respiratory infections, vaginosis, and topically to treat acne.[48]



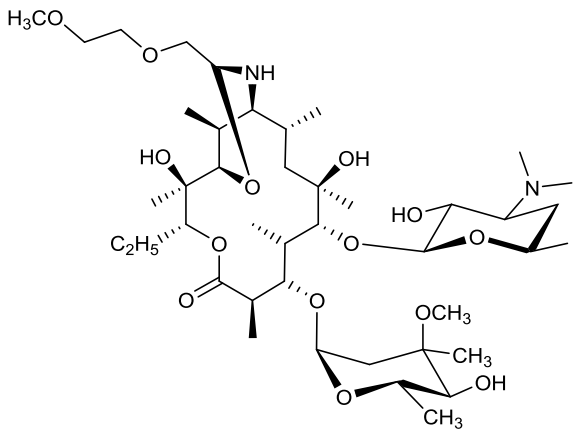
Lincomycin Clindamycin

Figure 1.6. Chemical structure of several lincosamides.

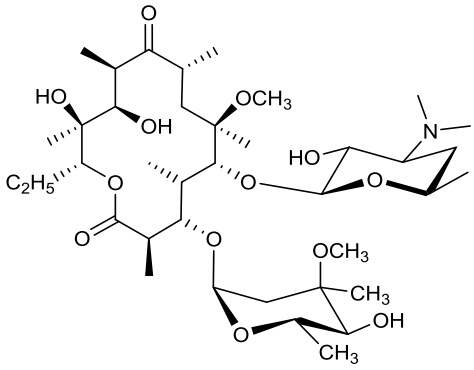
Although the chemical diversity, also macrolides contain one or more deoxy sugars, usually cladinose and desosamine, attached to a large macrocyclic lactone ring. These antibiotics also interfere with the proper translocation of the aminoacyl t-RNAs, thus inhibiting the bacterial synthesis of the protein by interfering with the proper translocation of the aminoacyl t-RNAs. Erythromycin, and its semi-synthetic derivatives dirithromycin, clarithromycin, and azithromycin, belong to this class of antibiotics (Figure 1.7). They possess a wide spectrum of activity including many Gram-positive and Gram-negative aerobic bacteria, and they are used to treat infections of the respiratory system.[49]



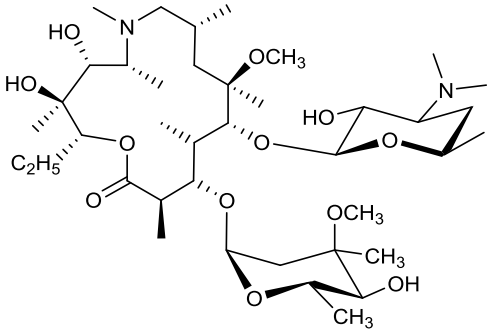
Erythromycin



Dirithromycin



Clarithromycin



Azithromycin

Figure 1.7. Chemical structure of several macrolides.

There also are glycan-based antifungal agents, like amphotericin B (Figure 1.8), used for the treatment of systemic mycosis. Amphotericin B is an amphipathic molecule: the hydrophobic macrolide is connected with a hydrophilic sugar portion (3-amino-3,6-dideoxy-β-D-mannopyranose). It binds ergosterol in the fungus cell membrane, creating a transmembrane channel. Due to the important alterations in the cell permeability, the fungal cell dies.[50-51]

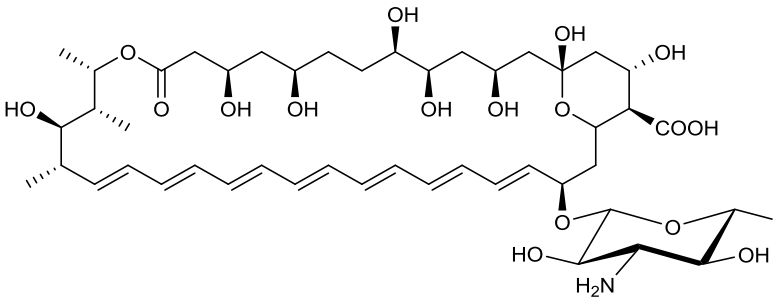
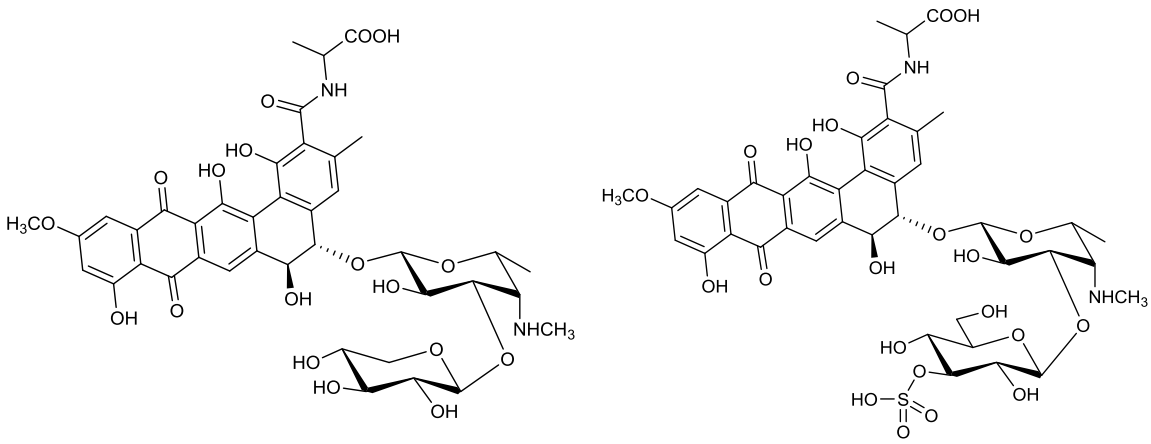


Figure 1.8. Chemical structure of amphotericin B.

The glycans also form part of the glycoproteins present on the viral surface and they are produced by the regular glycosylation machinery of the host cell, and thus trigger the immune system tolerance of the host. Human Immunodeficiency Virus (HIV) generally infects CD4+ cells employing a molecular mechanism in which the viral envelope gp120 glycoprotein interacts with a particular receptor (CD4) and a co-receptor (CXCR4 or CR5) located on the host cells. Then, the subsequent insertion of the viral envelope gp41 glycoprotein into the cell membrane causes the fusion and the entry of the virus.[52] Different approaches have been proposed to combat HIV.

Within the field of glycan-based drugs the discovery of a series of glycan-dependent broadly neutralizing antibodies (bNAbs) from HIV-infected individuals laid the foundations for designing better antigens for innovative HIV vaccines. In particular, the so called 2G12 monoclonal antibody (mAb) exquisitely recognizes a conserved and unusual dense cluster of oligomannose residues formed by terminal α-D-Manp-(1→2)-Man residues on the envelope glycoprotein gp120. Another approach could be the competition with the corresponding glycoproteins and antibodies,[53] using a variety of natural and glycomimetic ligands.

For instance, a family of small-size non peptide molecules, dubbed as carbohydrate-binding agents (CBAs), have been devised to target the highly mannosylated gp120 glycoprotein of the HIV envelope. These molecules have shown potent antiviral activities via interfering the interaction between this protein and CD4+.[54-55] Among CBAs, pradimicin A, a non-peptidic antibiotic, as well as its more soluble analogue, pradimicin S, also block the virus entry by inhibiting the fusion of the HIV virions with their target cells (Figure 1.9).[56-57]



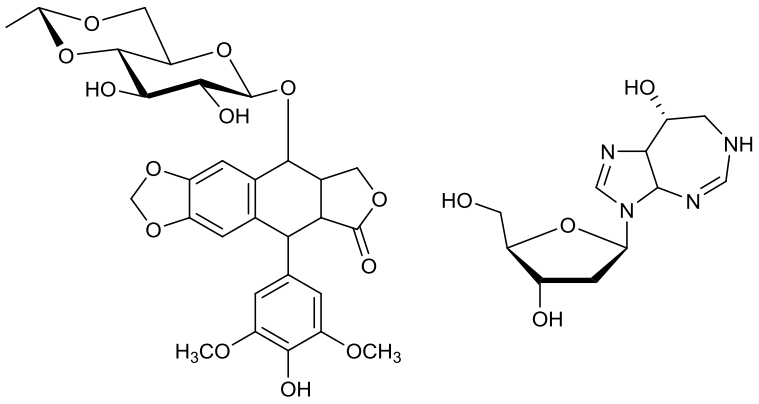
Pradimicin A Pradimicin S

Figure 1.9. Chemical structure of several carbohydrate-binding agents.

The knowledge of the capsular polysaccharides (CPS),[58] exposed antigens on bacteria surfaces, has allowed to develop glycan-based vaccine.[59-60] However, vaccines formulated exclusively with saccharides have regularly presented some limitations, such as the duration of the protection effect (one or two years), due to the absence of memory stimulating effects.[61-63] In fact, polysaccharides are able to stimulate specifically B cells, without the active participation of helper T-cell, and for this reason they are considered T-cell independent type 2 (TI-2) antigens. This fact precludes the existence of a significant immunity memory stimulation.[64-65]

Also cancer is not strange to glycans and aberrant glycosylation patterns either.[66] Abnormal protein/lipid glycosylation is diagnostic and a proper biomarker for a number of diseases. Recently, several glycosyl processing enzymes have been recognized as important targets for therapeutic intervention, leading to the design and development of distinct relevant classes of drugs.[67] In the early stages of cancer, there is an incomplete synthesis process of the complex glycans normally expressed in healthy epithelial cells. This event provides the biosynthesis of truncated structures, as sialyl Tn antigen (STn) expression in gastrointestinal and breast cancers.[68-69] On the other hand, in advanced stages of cancer, a neo-synthesis and expression of peculiar antigens take place, due to the cancer-associated induction of certain genes that control the expression of the carbohydrates.

This is, for instance, the case for the presence of sialyl Lewis a (SLea) and x (SLex) in many cancers.[70] In order to address a wide variety of tumours, different vaccines have been designed, aiming to target these different carbohydrate antigens, such as Globo-H, GM2, STn, TF, and Tn, on a particular peptide backbone. These molecules are then conjugated to a carrier protein, as the keyhole limpet hemocyanin (KLH) or the diphtheria toxin mutant (CRM197).[71] Also an overexpression of mucin glycoproteins, abnormal branching of *N*- and *O*-glycans, as well as significant modification of the fucosylation and sialylation levels have been identified.[72-74] For this reason, Tumour-Associated Carbohydrate Antigens (TACAs), such as the mucin-related epitopes, have been proposed as key targets for the development of anti-tumour vaccines.[67] To date, different TACAs-targeted vaccines are already in preclinical and early clinical settings,[75] such as VAXIL.[76] There are also anti-neoplasic drugs with a saccharide in their structure. For example, etoposide (Figure 1.10), a semi-synthetic β-D-glucopyranoside derivative of podophyllotoxin, inhibits DNA synthesis by forming a complex with topoisomerase II and DNA.[77] Another example is pentostatin (Figure 1.10), an inhibitor of adenosine deaminase and ribonucleotide reductase, particularly active for the treatment of lymphoproliferative malignancies.[78]



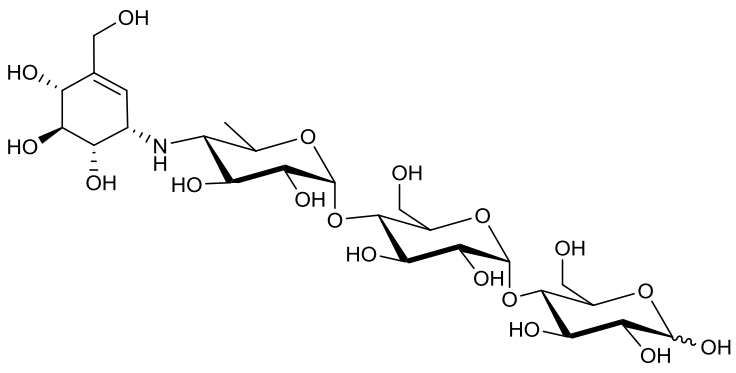
Etoposide Pentostatin

Figure 1.10. Chemical structure of several antineoplastic drugs.

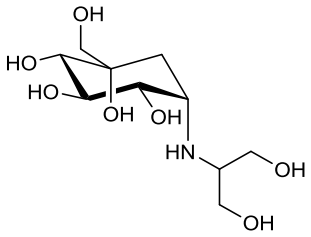
As mentioned above, malignancy is often related to glycosylation changes, and these variations can be used as cancer diagnosis and prognosis.

In fact, different antigens are used as diagnostic tools, such as the Carcino-Embryonic Antigen (CEA) in the colorectal cancer,[79] MUC16 (CA125) for the detection of ovarian tumor,[80-81] the prostate-specific antigen[82] for the prostate tumours, and the carbohydrate antigen 19-9 (CA19-9) for pancreatic, colorectal, gastric or biliary cancer.[83-84]

Glycans are also related to metabolic drugs. The metabolic syndrome (MetS) is associated with a variety of interconnected physiological, biochemical, clinical, and metabolic factors that directly increase the risk of a pro-inflammatory state that could end up into a chronic condition. The most known diseases associated to the MetS are atherosclerosis, cardiovascular disease, and type 2 diabetes mellitus. Since a wide variety of factors could cause the MetS, there are many treatments addressing the dysfunctions associated to the MetS.[85] Inhibitors of α-glucosidase have been extensively used to prevent the presence of high glucose concentrations in blood after food ingestion.[86-87] α-glucosidase is an enzyme present in the brush border of the small intestines, hydrolyzing oligosaccharides into glucose and other monosaccharides in the postprandial state. High glucose levels lead to the activation of kinases causing endothelial disfunction,[88] with cardiovascular effects. To prevent these events, some of the most used drugs are α-glucosidase inhibitors,[67] which have also been demonstrated to be useful in patients with diabetes mellitus type 1.[89-90] Among them, acarbose (Figure 1.11) is a synthetic tetrasaccharide glycomimetic, displaying one carbamino sugar moiety. This molecule has diverse activities, since not only inhibits membrane-bound α-glucosidases, but also blocks pancreatic α-amylase.[91] Acarbose is orally administered to non-insulin dependent diabetes mellitus patients, whenever diet modifications or oral hypoglycemic agents are not able to control their condition. In the context, voglibose (Figure 1.11), also a glycomimetic displaying one carbamino sugar fragment, is frequently used for the treatment of type II diabetes.[92] A variety of sugar mimics, including iminosugars and piperidine-containing saccharide mimetics, also inhibit various glycosidases in a reversible and competitive manner. In fact, there are mimics of the ground state conformation of the natural sugars, as a consequence because of their structural resemblance to the terminal saccharide moiety of the natural neutral substrates.



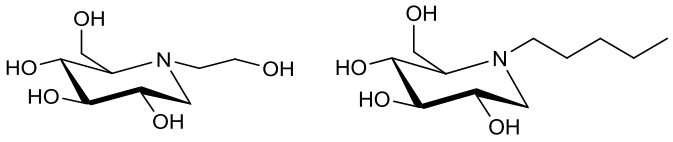
Acarbose



Voglibose

Figure 1.11. Chemical structure of several α-amilase inhibitors.

A large number of iminosugars have been discovered from plants and microorganisms.[93-95] For example, miglitol (Figure 1.12), an analogue of 1-deoxynojirimycin, is used for the treatment of type II diabetes. Iminosugars are also employed in the treatment of lysosomal storage disorders (LSDs), rare diseases with terrible consequences for normal life. Among them, the most important drug is miglustat (Figure 1.12), which is used for the treatment of Gaucher's disease. In this disease, the normal activity of the ceramide glucosyltransferase is altered, leading to the accumulation of glucosylceramide. This glycolipid causes the enlargement of the liver and the spleen, changes in the bone marrow and blood, and bone disease.[96] Miglustat has also been approved for the treatment of progressive neurological complications in patients with Niemann–Pick disease type C, another LSD characterized by the progressive accumulation of cholesterol in the lysosomes. Different basic research studies have permitted to disentangle the mechanism of action of these inhibitors. In particular, they are able to resemble the cationic hydrolase transition state in glycoside hydrolysis.[97] This knowledge can be exploited to develop new strategies based on molecular design coupled to the development of efficient screening protocols for the identification of novel chemical entities. The final aim would be to develop more selective inhibitors with minimized side effects versus the currently employed drugs.[98]



Miglitol Miglustat

Figure 1.12. Chemical structure of several iminosugars.

Nowadays, glycoproteins are also well introduced in the drug market as therapeutics. Recombinant expression of the glycoproteins may lead to isoforms with null or incomplete glycosylation patterns, which may modify their therapeutical properties. Since the efficacy of biopharmaceuticals proteins critically depends on their specific glycoforms, different synthetic, bioengineering, chemoenzymatic, and chemical biology-mixed strategies have been applied for the production of improved therapeutic glycoproteins. One of the most representative cases is the erythropoietin (EPO), a blockbuster anti-anemia drug. This hormone regulates the production of red blood cells, other human hormones, interferon-β, interleukin-6, and glycan-containing monoclonal antibodies.[99] Recombinant EPO can be safely discriminated from endogenous EPO due to the glycosylation pattern characteristic of mammalian cells. Nevertheless, the current success in the synthesis of glycosylated EPO poses a challenge for the discrimination of the endogeneous and the exogenous hormone.

Also glycolipids are very abundant in nature especially on the surface of Gram-negative bacteria (i.e., lipopolysaccharides, LPS, and lipooligosaccharides, LOS). In particular, LPS (Figure 1.13) are responsible for acute sepsis and septic shock,[100] and its chemical structure has served as inspiration for rational drug design studies. The LPS structure comprises a lipid A moiety, which contains a β-1→6 linked disaccharide of *N*-acetylglucosamine linked by either ester or amide bonds to five or six hydroxylated fatty acids (FA), plus a polysaccharidic core.[101-103] The toxic action of LPS is due to a complex cascade of extracellular protein-LPS and protein-protein interactions, leading finally to the activation of the membrane-bound Toll-like receptor 4 (TLR4),[104] with the subsequent intracellular signalling process, and the synthesis of pro-inflammatory cytokines.

TLR4 activation requires the participation of the myeloid differentiation factor 2 (MD-2), and the dimerization of two TLR4/MD-2 single complexes. LPSs are able to bind this TLR4/MD-2 system, in the so-called antagonist conformation, by inserting the FA chains into the lipophilic pocket of MD-2. The global 3D geometry is also stabilized by the establishment of polar interactions with both TLR4 partners.[105]

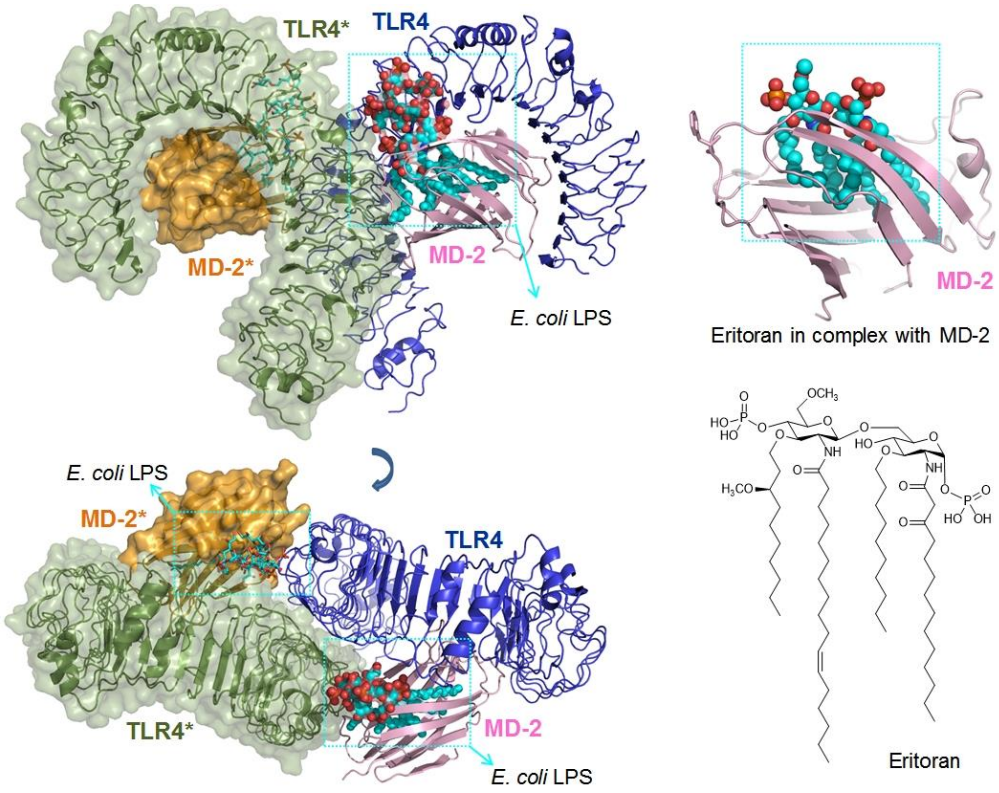


Figure 1.13. Representation of the X-ray crystallographic structure of TLR4/MD-2 in complex with Escherichia coli LPS (PDB ID 3FXI). On the right (top), detail of the X-ray crystallographic structure of MD-2 in complex with eritoran extracted from PDB ID 2Z65. On the right (bottom), the chemical structure of eritoran is depicted.

1.2 Carbohydrate-protein interaction

Proteins that bind carbohydrates are called GBPs (Glycan Binding Protein) and in this group lectins, receptors, toxins, microbial adhesins, antibodies and enzymes, among others[106] are included. In the framework of cellular recognition, non-enzymatic proteins are the most important proteins that bind carbohydrates without drastic conformational changes. Between them, lectins are widely distributed in nature since they can be found in almost all living organisms. The recognition process between a lectin and a carbohydrate usually occurs through:

- Hydrophobic interactions, such as CH-π interaction (between the sugar moiety and residues such as tryptophan, tyrosine and phenylalanine).
- A net of hydrogen bonds between the OH group of the sugar and the polar group of the protein. Frequently a water molecule is also involved in this net.
- Metal coordination (usually with calcium or manganese) present in the binding site of the lectins.

These characteristic interactions occur in a specific region of the protein normally named Carbohydrate Recognition Domain (CRD). Unconventional sugars binding sites have been observed, as for the *C*-type lectin-like receptor 2 (CLEC-2).[107] The specificity of a lectin to recognize selectively some epitopes is due to the diversity of the CRD. Actually lectins have been classified based on their CRD, the sugar specificity and their expression patterns. In particular, human lectins are divided into two classes: *C*-type (or calcium-dependent) lectins and S-type (sulfhydryl-dependent or calcium-independent) lectins. Their main functions are the cellular recognition, cellular adhesion regulation, as well as immunological functions such as recognition of sugars from pathogen organisms (Figure 1.14).[108] Lectins also play an essential role in the infection process. One of the most studied players in this event is the Dendritic-Cell Specific ICAM-3 Grabbing Nonintegrin (DC-SIGN). DC-SIGN is a *C*-type lectin that has been demonstrated to spread and evade the immune system by the HIV virus. Fittingly, this lectin seems to be involved in several infections by pathogens, such as *Ebola* virus and *Mycobacterium tuberculosis*.[46, 109]

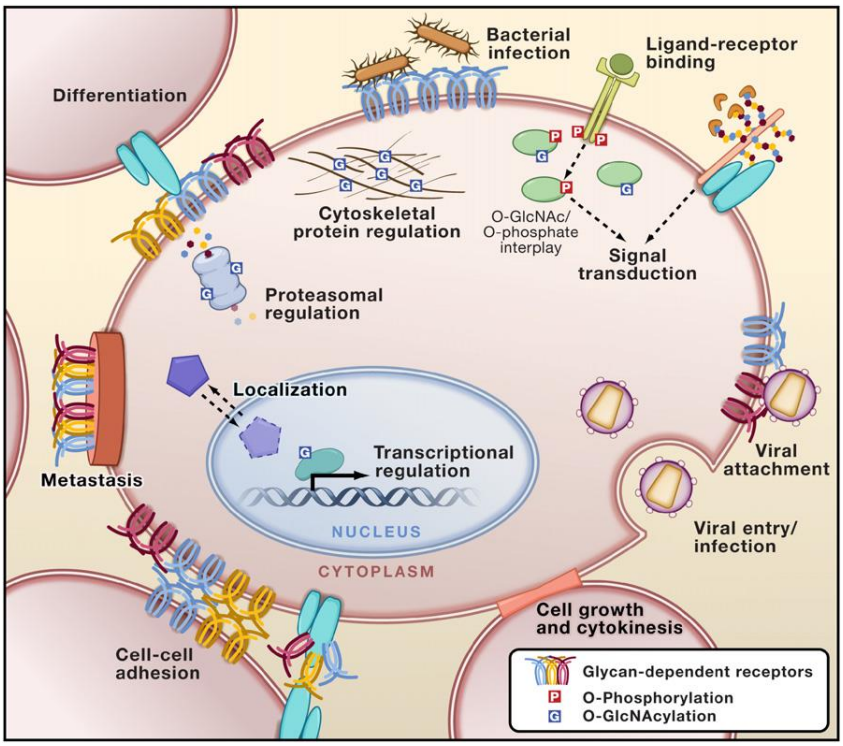


Figure 1.14. Schematic representation of the main function of glycans (figure from reference 110).

Plant lectins are accumulated in seeds or other tissues and mainly have a storage function but also they seem to play an important role in plant defence. Twelve families of lectins have been identified, although they commonly share some features. Plant lectins are usually found as dimers or tetramers, in which each monomer has a CRD with two metal binding sites for calcium and manganese. In the dimer there is a side-side protein-protein interaction forming 12-strandeed sheet. Plant lectins are very interesting because due to their easy availability are extensively used as spy molecules due to their high glycans specificity. For example, it is very well known that Pea lectin and *Concanavalin A* selectively bind mannose or glucose, while *Maackia amurensis* lectins bind sialic acids. To date, more than 132 crystal structures of legume lectins have from 18 different plants are available from the Protein Data Bank.[111]

1.2.1 Pea Lectin

Pisum sativum lectin (also named Pea lectin) is a dimeric protein mannose/glucose specific with agglutinating activity.[112] To date, there are five 3D structures of *Pea* lectin in complex with glucose, mannose and sucrose available at the PDB (Table 1.1). In chemical glycobiology, Pea lectin is usually used as a model lectin for the study of carbohydrate-protein interactions, especially in the case of studying mannose or glucose-based glycomimetics.[113]

Table 1.1. Pea lectin X-ray crystallographic structures available at the PDB.

PDB ID	Resolution (Å)	Ligand
1OFS	1.86	sucrose (α-D-glucopyranosyl-1-2-β-D-fructofuranoside)
1HKD	2.09	α-methyl-D-glucopyranoside
2BQP[114]	1.9	α/β D-glucopyranoside
1BQP[115]	2.1	α/β D-mannopyranose
1RIN[116]	2.6	trimannose

In the carbohydrate binding site, manganese and calcium ions are present. The specificity for mannose and glucose is due to the fact that the hydroxyl group present in the position 2 either in axial or in equatorial position of each sugar is not involved in any interaction (Figure 1.15).[116]

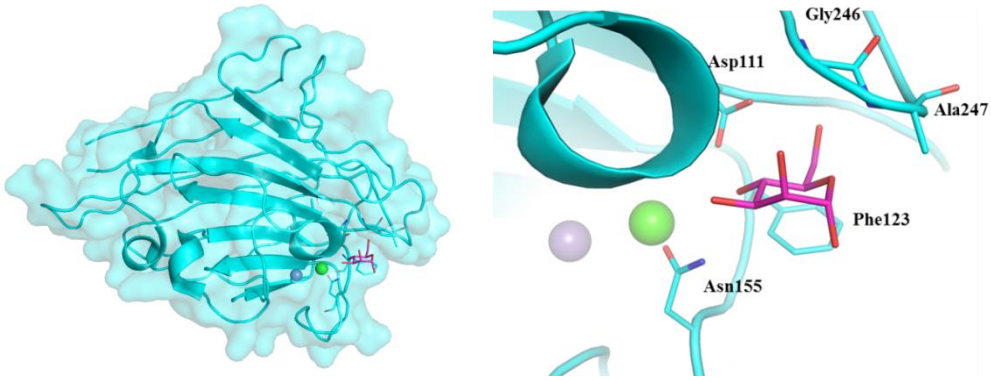
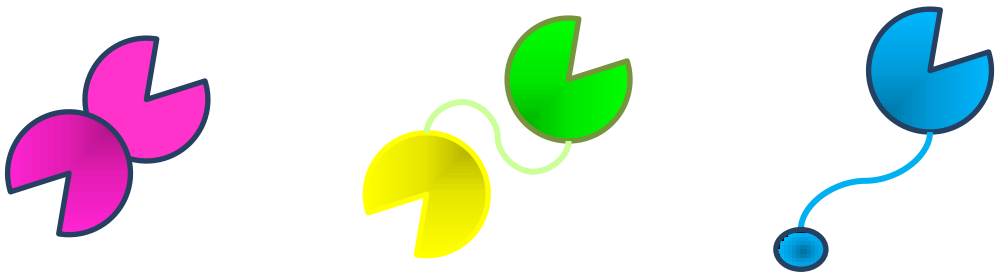


Figure 1.15. Left: X-ray crystallographic structure (PDB ID 1RIN) of the Pea lectin (cyan) in complex with the mannoside (magenta). Magnesium (purple) and calcium (green) are shown as spheres. Right: Details of the carbohydrate recognition domain. Aminoacids involved in the binding are depicted (shown in sticks).

1.2.2 Galectins

A special type of lectins evolutionary-conserved[117] is galectins (abbreviation of galactose binding lectins). Galectins constitute a family of β-D-galactoside binding proteins. They are localized in cytoplasm, nucleus, cell surface, and extracellular matrix,[118-119] and present a S-type sequence motif. To date 15 members of this family have been identified in mammals (gal-1 to -15), but only 10 of them are present in humans (gal-1, -2, -3, -4, -7, -8, -9, -10, -12, -13).[120] They are soluble proteins with a molecular weight around 14-36 kDa[121] and they lack of enzymatic activity.[6] Among the physiological processes in which galectins have an important role, there are regulation of the immune response, cell cycle, cell growth and apoptosis.[122-127] Due to these functions, an alteration in these proteins can trigger various pathological processes, depending on which galectin is affected. Thus, it has been demonstrated that galectins are involved in processes like cancer, metastasis, inflammation, hypersensitivity and atherosclerosis, among other pathologies.[128-137] Therefore, in the last years the interest in these proteins as therapeutic targets has increased since the development of selective galectins modulators could improve the prognosis of many diseases.[138-142]

Based on the structural features, mammalian galectins have been classified in **proto**, **tandem-repeat**, and **chimera** types (Figure 1.16). Proto-type galectins (gal-1, -2, -5, -7, -10, -11, -13, -14, and -15) are usually dimers containing only one type of CRD (homodimers) non-covalently joined. Tandem-repeat-type galectins (gal-4, -6, -8, -9, and -12) have two different CRDs, almost homologous, that are covalently joined. Chimera-type galectin (only gal-3) has a carboxyl-terminal CRD associated to an amino-terminal peptide which is rich in tyrosine, proline and glycine (collagen-like sequence),[143] forming a tail that intervenes in the oligomerization of this unique galectin.[144]



Proto
Two identical CRDs

Galectin-1, -2, -5, -7, -10, -11, -13, -14 and -15

Tandem-repeat
Two different CRDs

Galectin-4, -6, -8, -9 and -12

Chimera
One CRD and one non-lectin domain

Galectin-3

Figure 1.16. Classification of galectins depending on their CRD.

The activity of galectins is ligand-concentration dependent and it can be both stimulatory and inhibitory.[145] These proteins recognize carbohydrates through the CRD, which consists of 130 amino acids approximately,[146] and it folds into a β-sandwich structure comprising two anti-parallel β-sheets (the F-sheet and S-sheet).[147] The CRD that is different among the galectins family, establishes a net of interactions selectively with the β-galactoside moiety through the formation of hydrogen bonds and hydrophobic interactions with tryptophan (Trp), histidine (His), and arginine (Arg) residues (Figure 1.17).[148] Apart from the CRD, galectins differ in other aspects, such as chain length or number of domains in their structure. To date, more than one hundred X-ray structures of human galectins are available at the Protein Data Bank (PDB).[111] For the studies performed for this Thesis, we selected one or two X-ray crystallographic structures for each human galectin (PDB ID codes are shown in Table 1.2).

Table 1.2. X-ray crystallographic structures of human galectins used in this Thesis.

Galectin	PDB ID	Resolution (Å)	Ligand
Galectin-1	1GZW[149]	1.7	lactose
Galectin-2	5DG2[150]	1.61	lactose
Galectin-3	3ZSJ[151] 1KJR[152]	0.86 1.55	lactose 2,3,5,6-tetrafluoro-4-methoxy-lactose
Galectin-4	4XZP[153]	1.48	no ligand
Galectin-7	4GAL[154]	1.95	lactose
Galectin-8	4BMB[155]	1.35	lactose
Galectin-9	3NV4[156]	1.99	sialyllactose
Galectin-10	1G86[157]	1.8	*N*-ethylmaleimide

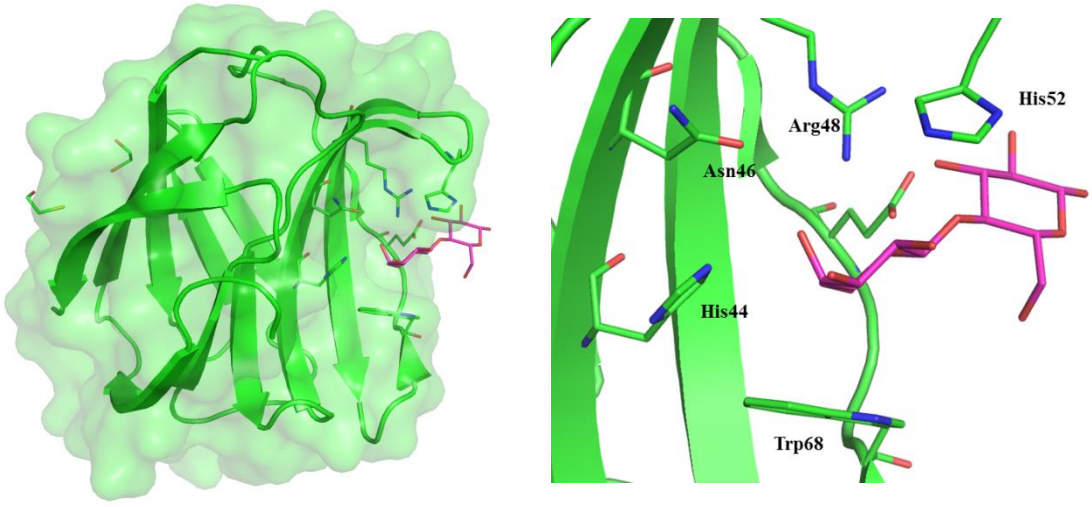


Figure 1.17. Left: X-ray crystallographic structure (PDB ID 1GZW) of human galectin-1 (green) in complex with lactose (magenta). Details of the carbohydrate recognition domain. Aminoacids involved in the binding are depicted (shown in sticks).

1.2.3 *Maackia amurensis lectins*

In the early 1970,[158] it was described the hemoagglutinating activity from the seeds from *Maackia amurensis*, and it was discovered that it was due to the presence of two different lectins able to agglutinate different blood cell types: *Maackia amurensis* leukoagglutinin (MAL) and *Maackia amurensis* hemagglutinin (MAH). These lectins selectively recognize sialylated oligosaccharides. In particular, MAL preferentially binds sialylactosamine (αNeu5Ac-(2-3)βGal1-4βGlcNAc) and *N*-linked glycans and MAH disialylated tetrasaccharide (αNeu5Ac2-3βGal1-3(αNeu5Ac2-6)αGalNAc) and *O*-linked glycans. Although the high identity sequence (more than 86%), the presence of an alanine (Ala219) in MAH instead of a tyrosine (Tyr250) in MAL in the binding site determines the different recognition process. The X-ray crystallographic structure of MAL is available at the PDB (PDB ID 1DBN) and for MAH a homology model has been already reported.[159] The binding site of the MAL consists in a deep pocket that can accommodate up to three sugar moieties. The peculiarity of these lectins are related to the fact that normally all the legume lectins conserve polar interactions with one aspartate, one glycine and one asparagine. Instead, MAL and MAH have one lysine (Lys136 in MAH and Lys105 in MAH) instead of the glycine, and one aspartate (Asp166 in MAL and Asp135 in MAH) instead of the asparagine (Figure 1.18). Mutagenesis studies on MAH demonstrate that the replacement of the Lys105 with a glycine, or the Asp135 with an asparagine, abolish the ability to bind the ligand.[160]

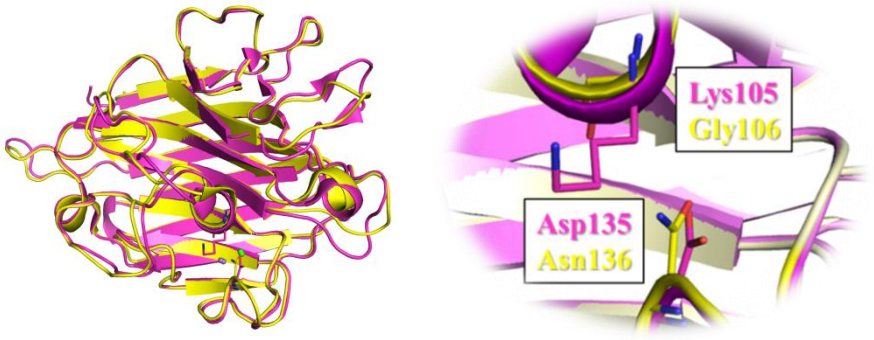


Figure 1.18. Left: Superimposition of the X-ray crystal structure of MAL (in magenta, PDB ID 1DBN), and the X-ray crystal structure of a lectin from Ulex europaeus (in yellow, PDB ID 1QNW) as representative lectin. Right: Differences in the aminoacids between a model legume lectin and MAL are shown.

1.2.4 Glycosyltransferases

Glycosyltransferases (GTs) are key enzymes responsible of the incorporation of carbohydrates into a variety of acceptor biomolecules, including proteins, lipids, oligosaccharides and different metabolites.[161-163] The resulting glycoconjugates mediate a wide range of functions from structure and storage to signalling and, in consequence, they are related with important diseases, non-alcoholic fatty liver disease.[164] Therefore, chemical manipulation of GTs activity could lead to the development of useful therapeutic drugs.[165] Because of it, considerable synthetic efforts have been directed toward the preparation of efficient GT inhibitors.[166-172] Transfer of the sugar residue occurs from an anionic nucleotide sugar donor to the acceptor substrate; it can take place with retention or inversion of configuration at the anomeric centre of the sugar residue.[173] Only nine sugar donors are known to be involved in protein glycosylation in mammal organisms,[174] which is the most abundant post-translational modification in nature. Six of these sugar donors (glucose, glucosamine, glucuronic acid, galactose, galactosamine, xylitol) contain the uridine moiety that is in line with the existence of GTs employing UDP sugars as the most predominant in nature.[175] In particular, GalNAc-T2 a member of the *N*-acetylgalactosaminyl transferase family (GalNAcTs E.C. 2.4.1.41) is involved in mucin biosynthesis,[176-177] so it is involved in aberrant glycosilation in pathological condition such as cancer.[178] To date, there are 20 X-ray crystallographic structures of GalNAc-T2 available at the PDB. For the present Thesis, we used the structures shown in Table 1.3. The peculiarity of this protein is the conformational change associated to a loop which determines the binding of the substrate. For this reason, in this Thesis two different crystal structures were used, one in closed conformation (PDB ID 4D0Z) and one in open conformation (PDB ID 2FFV).

Table 1.3. X-ray crystallographic structures of GalNAc-T2 used for this Thesis.

PDB ID	Resolution (Å)	Ligand	Conformation
2FFV	2.75	UDP	Open
4D0Z	2.2	UDP-5SGalNAc	Closed
5FV9	2.07	Compound 16d	Open

1.3 Computational techniques in drug design

Given the difficulty of getting proper crystals of carbohydrates beyond a particular size, molecular modelling has proven to be very useful for the elucidation of their structural and conformational characteristics, either alone or in combination with experimental techniques, usually solution state NMR spectroscopy methods. In fact, the combination of modelling/NMR protocols has been extremely useful to deduce the conformational and dynamic properties of free and bound carbohydrate molecules.[13, 179-180] Molecular modelling of carbohydrates can be performed at different levels of complexity. Specific structural properties of carbohydrates can be extracted from quantum mechanics (QM) methods, especially those related to electronic features, to study chemical reactions (for instance, by using hybrid quantum mechanics/molecular mechanics QM/MM approaches), and to calculate force constants and atom charges to be used as force field parameters.[181-183] These methods can nowadays be employed even to deal with molecular recognition details. The most employed protocols for the study of carbohydrates are based on force fields, to perform molecular mechanics (MM) and molecular dynamics (MD) simulations (Figure 1.19). Research in force field development has been very active in recent years, yielding several high quality force fields able to reproduce dynamic and electronic properties of carbohydrates. In particular, MD simulations are one of the most useful tools to have insights into the conformational dynamics of the system, combined, if available, with NMR experimental information. The interaction of carbohydrates with their receptors can also be performed at different levels of complexity, also using force-field based methods (MM and MD simulations) or quantum mechanics (QM) methods (Figure 1.19).[184-185] Choosing the appropriate computational method is a key decision in a given study, and it will depend on the properties under investigation, and also on the starting experimental information we have in our hands.

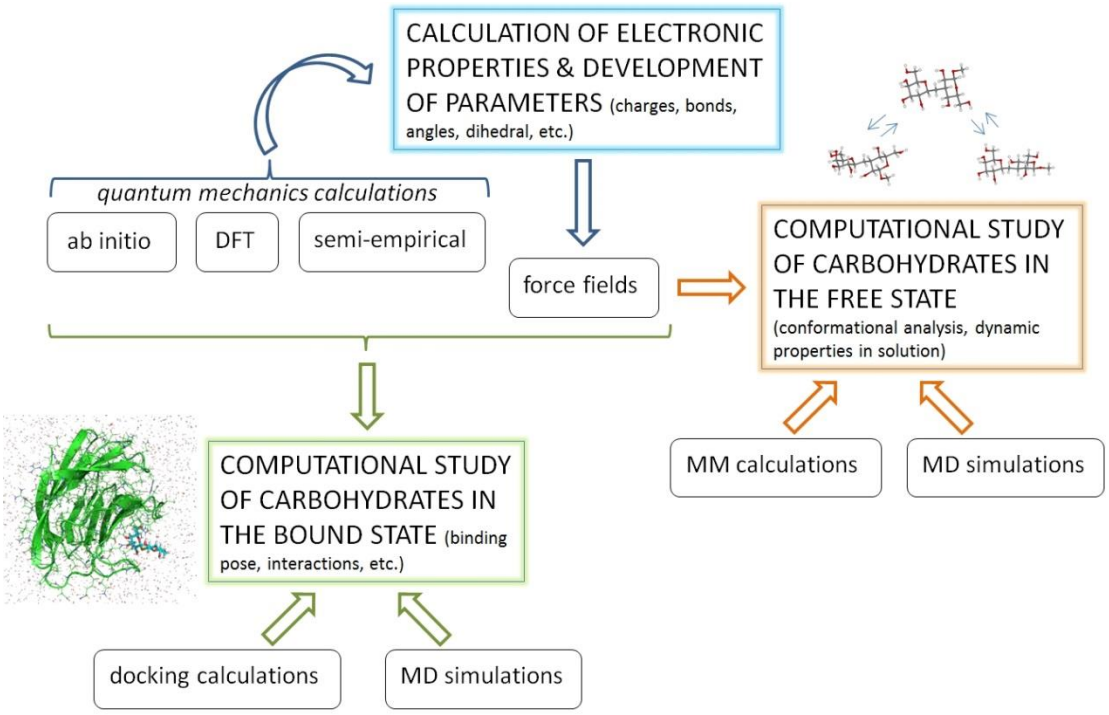


Figure 1.19. Computational methodologies employed for the study of carbohydrates.

1.3.1 Ab initio calculations and semi-empirical approaches

Ab initio quantum mechanical (QM) calculation is a very useful tool to investigate molecular properties or describe chemical reaction. This calculation is not based on experimental data but it takes information from the electronic properties of each atom. Depending on the size of a molecule or a system (receptor-ligand) this type of calculation has a very high computational cost, reason why with this technique it is not possible to analyse system with more than few hundred atoms. QM calculations are based on the solution of Schrödinger equation (Equation 1).[186]

$$\left\{-\frac{\hbar}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V\right\}\Psi(r,t) = i\hbar\frac{\partial\Psi(r,t)}{\partial t} \qquad \text{(Eq. 1)}$$

This equation describes the probability of a single electron of mass *m* that is moving in a three-dimensional space with a vector r = x+y+z during the time t. V represents an external field such as the electrostatic potential due to the presence of the nuclei in a molecule. i is the unity imaginary number (the square root of -1) and ħ is the Plank constant divided by 2π. Several mathematical approximations have been proposed to approximate a solution to the Schrödinger equation.

The most known approximation is the Hartree Fock (HF) method, in which electronic correlation (the Coulombic electron-electron repulsion) is not taken in consideration. The limitation of this method is that the resulted energy is usually higher than the exact one. Moreover, in HF calculations electrons must be described by some functional form, which is only known for a single-electron system.[187] The set of one particle functions employed to assemble molecular orbital are defined as basis sets. Among them the linear combinations of Slater type orbitals (STO) or Gaussian type orbitals (GTO) are the most popular. Also other functions can be added such as the polarization function, which takes in consideration the possible shift of an atom's orbital, and the diffuse function, in which the electron is considered far away from the nucleus. The diffuse function is very important for anions and van der Waals complexes.

Quantum Monte Carlo (QMC) is a very time consuming method, which avoids HF limitations. This method explicitly correlates wave functions and evaluates integrals numerically using a Monte Carlo integration. QMC is probably the most accurate method known today.[188]

Density Functional Theory (DFT) is another QM method, in which the total energy is related to the total electron density instead of the wavefunction. In the Chemical Biology field, DFT methods are very widely used since they combine low computational cost with high accuracy.[189]

Semiempirical calculations are based on HF assumption but the parameters or numbers are fitted with experimental data. The advantages of this method are the lower computational cost, compared with *ab initio* calculations and the velocity of the calculation. The disadvantage of this technique is related to the lack of experimental data for particular systems.

Widely used semi-empirical methods include AM1, PM3, and MNDO.[190] More recently, also PM5 and PM6 have been used for carbohydrates.[191] DFT has also been employed for conformational studies of carbohydrates using different basis sets,[192-195] and also solvation models.[196] To tear down the computational costs a reasonable approach can consist on preliminary calculations by means of semi-empirical methods followed by more precise studies at *ab initio* level or DFT calculation.[197-198]

1.3.2 Docking

Molecular docking is the most common method applied in structure-based design. It provides a model of interaction between two proteins or between a small molecule and a protein at the atomic level. Computational docking protocols are usually adopted to explore the possible binding poses (and therefore the putative ligand-receptor interactions) of a given ligand within a particular receptor, and to estimate the binding affinity of the resulting complex. The availability of the 3D structure of the receptor (by means of, commonly, X-ray crystallography, NMR spectroscopy, electron microscopy or homology modelling) is required and previous information about the binding site is preferable. The knowledge of the binding site could be deduced by comparison of other related target proteins or through complementary computational studies. For decades, the interaction between protein and ligand has been thought as key-lock system. This model is very easily reproducible with a rigid docking in which both ligand and protein are treated as rigid. With the advent of the induced-fit theory by Koshland, in which a mutual conformation change of both ligand and protein occurs at the moment of the interaction, the global vision of this model evolved.[199] However, these two theories are now better substituted by the concept of conformational ensembles. In a dynamic context, a biomolecule indeed does not have only a fixed conformation. The real three-dimensional structure will be represented by an ensemble of structures which interconvert through dynamic processes. Thus, the study of dynamics and its impact in conformation and molecular recognition is of paramount importance within this field.

So, it has to take in consideration that ligands are usually faced with several receptor microstates rather than with a unique structure, as is usually assumed for simplification.

Karush defines this phenomenon as a configurational adaptability in which the best-fitting configuration for a biomolecule after interacting with a ligand would become selected from the whole structural ensemble.[200] Moreover, Weber suggested that the binding with a ligand shifts the conformational equilibrium of the receptor in favour of those conformers in the dynamic ensemble that are most complementary to the ligand.[201]

In order to reproduce this event, flexible docking is needed. Nowadays, current docking methods incorporate ligand flexibility with high efficiency, and receptor flexibility is becoming incorporated in several docking programs, although with limitations.[202] The flexibility of a protein can be included into the binding site before the docking or afterward, for example, it is possible to generate different ensembles structures from MD simulation or from Normal Mode Analysis (NMA). Some programs use different strategy to include the flexibility in the receptor. For example, AutoDock4 program allows the selection of some bonds of the side chain of several residues as rotatable bonds when performing the docking of the ligand.[203] Another strategy is to perform a conformational search and a refinement of some side chains of selected residues or in the region of the binding site while docking the ligand (*e.g.* protocol used by Glide program).[204-205] Obviously this technique has an higher computational cost than considering the receptor as rigid.[206]

A docking calculation is basically characterized by two main steps: in the first step, a conformational search is performed to predict possible conformations of the small molecule (ligand); in the subsequent second step, for the different binding poses, the ligand binding energy is calculated by applying a scoring function so the predicted ligand-receptor complexes are therefore scored and ranked. The main differences among the different docking programs are the search algorithms that are used in the conformational search, and the scoring function employed to rank the different binding poses. The accuracy and the computational cost of this step depend also on the number of rotable bonds present in a ligand. Each program has a threshold of the maximum number of rotable bonds because numerous rotable bonds increase exponentially the computational cost. Three types of search algorithm can be applied: shape matching, systematic search and stochastic algorithms. The more time-efficient algorithms are the stochastic or non-deterministic ones.

In the stochastic algorithm random changes in the ligand are executed. This change will be accepted or rejected according to probabilistic criteria. Four types of stochastic algorithms are known: Monte Carlo (MC) methods, Evolutionary Algorithms (EA), Tabu search methods and Swarm Optimization (SO) methods.

In MC methods Boltzmann probability function is applied (Equation 2).

$$P = exp\left[\frac{-(E_1 - E_0)}{k_B T}\right] \qquad \text{(Eq. 2)}$$

In (Equation 2) E_0 and E_1 represent the energy score before and after the change of the conformer respectively, k_B the Boltzmann constant and T the absolute temperature of the system.

Evolutionary algorithms search for the correct ligand binding mode using ideas based on genetics and evolutionary process in biological systems. A "natural protocol" is applied, which consists in the first generation of individuals and the evaluation of the fitness. According to the Darwin theory, the best-fit individuals for reproduction are selected, and after crossover and mutations new individuals are generated and again the fitness of them will be evaluated. Thus, the optimization is based on a selection process that mimics biological evolution. One of the most popular is the Genetic Algorithm (GA), in which an analogy with evolutionary selection is applied. Basically, the translation, orientation, and conformation of a small molecule describe the gene, each state (conformation) corresponds to the genotype, and the atomic coordinates correspond to the phenotype. The best solutions will survive and will reproduce (with crossover and mutation), whereas the worse ones will be discarded. Other genetic algorithms have been proposed in order to represent more complex situations.[207] For example, an improvement of the genetic algorithm can be obtained by introducing a local search (LS) method.[208] A LS method performs energy minimization, without requiring gradient information about the local energy landscape. Moreover it is possible to define the step size changing the number of consecutive failures, or increases in energy. The hybrid of the GA method with the adaptive LS method is named Lamarckian genetic algorithm.

Another hybrid algorithm is the simulated annealing, in which a GA and LS methods are applied depending on the temperature.

The advantage of this method is the possibility to overcome energy barriers separating energetic valleys by means of high temperature.

In order to evaluate the accuracy of an algorithm a scoring function plays a key role. The scoring function could be of three types:

- Force Field (FF) based scoring functions (Equation 3),[209] where a classic force field is employed to compute individual interaction terms such as van der Waals and electrostatic energies, stretching/bending/torsional energies. Several disadvantages are related to this scoring function. One of the disadvantages is the solvent effect, which is finally defined by a distance-dependent dielectric constant. Another problem related to the FF scoring function is how to treat the water. The Poisson-Boltzmann/Surface Area (PB/SA) model and the Generalized-Born/Surface Area (GB/SA) use implicit solvent models to overcome this problem. But the more severe challenge still remains the entropic effect.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \qquad \text{(Eq. 3)}$$

  In (Equation 3) r_{ij} represents the distance between an atom *i* of the protein and an atom *j* of the ligand. *Aij* and *Bij* are the van der Walls parameters, and q_i and q_j are the atomic charges of the atom i and j respectively. $\varepsilon(r_{ij})$ reflects the screening effect of water on electrostatic interactions and it is usually set to is usually set to $4r_{ij}$.

- Empirical scoring function (Equation 4)[210-211] calculates the overall binding free energy, by summing up a set of weighted empirical energy terms, including hydrogen bond (H-bond) and hydrophobic interactions. Compared to the FF scoring functions, it is more computationally efficient. Its applicability depends on the training set (known experimental data for a set of molecules) and the fitting to known binding affinities. Glide Score is one of the examples of empirical scoring function.

$$\Delta G = \sum_i W_i \Delta G_i \qquad \text{(Eq. 4)}$$

ΔG_i stands for the individual empirical energy terms and the corresponding coefficients W_i.

- Knowledge-Based (KB) scoring functions (Equation 5)[212-213] are based on the sum of distance-dependent statistical potentials between the ligand and the target. The structural information of the complex formed by the protein and the ligand is needed.

$$\omega(r) = -k_B T ln\left[\frac{\rho(r)}{\rho^*(r)}\right] \qquad \text{(Eq. 5)}$$

In (Equation 5) the scoring function is dependent on the density of the protein-ligand atom pair at the distance r ($\rho(r)$) in the training set, and the pair density in a reference state ($\rho^*(r)$) where there are no interatomic interactions at the absolute temperature T. k_B is the Boltzmann constant.

In this Thesis, three different docking protocols were used: AutoDock4,[214] Glide[204-205] and VINA.[215] Since docking programs are not specifically developed to treat with carbohydrates, further refinement of the proposed docked poses has been accomplished by means of MD simulations in water. If available, experimental information (i.e., NMR/X-ray data, binding/affinity data, etc.) has been taken into account in the calculations.

1.3.2.1 AutoDock4

AutoDock4[214] is a freeware molecular docking program. The conformational search is performed with Lamarckian genetic method, but also a simulated annealing search method and a traditional genetic algorithm are available. A grid-based method is applied in order to quickly evaluate the binding energy of putative binding poses.

The ranking of the docked binding poses are scored by a semi-empirical function. The binding free energy is calculated as the difference between the potential energy of ligand and the protein in the complex (bound state) and the potential energy of the ligand and protein in a "free" or unbound state (Equation 6).

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + (V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}) \quad \text{(Eq. 6)}$$

Each potential energy contribution is given by the sum of van der Walls, hydrogen bonds, electrostatic and solvation term (Equation 7).

$$V = W_{vdW} \sum_{i\ j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbound} \sum_{i\ j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec} \sum_{i\ j} \left(\frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}}\right) + W_{sol} \sum_{i\ j} \left(S_i V_j + S_j V_i\right) e^{\left(-\frac{r_{ij}^2}{2\sigma^2}\right)} \quad \text{(Eq. 7)}$$

In van der Waals term (W_{vdW}) A and B are parameters taken from the Amber force field,[216] while in the hydrogen-bond term (W_{hbound}) C and D are parameters depending on the distance (r) between the atom i and j. Both of the terms are potential based (6/12 in the van der Waals and 10/12 in the hydrogen bond).[217] The electronic term (W_{elec}) takes into account the partial charges of the atoms (q_i and q_j) in the solvent ($\varepsilon(r_{ij})$). The desolvation potential is based on the volume (V) of the atoms around a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance (r_{ij}) with a distance weighting factor σ of 3.5 Å.[218]

In the final ranking and clustering-based scoring methods also the loss of entropy (ΔS_{conf}) associated to the binding (Equation 8) is considered. This contribution is strictly related to the number of rotatable bonds (N_{tors}) in the ligand.

$$\Delta S_{conf} = W_{conf} N_{tors} \quad \text{(Eq. 8)}$$

Autodock4 is very versatile and has different utilities. It is possible to perform flexible docking, specifying which residues of the protein should be considered flexible during the docking.

1.3.2.2 Glide

Glide is a commercial molecular docking program provided by Schrödinger.[204-205, 219] Possible binding sites of the receptor are hierarchical filtered, based on the shape and properties of the receptor that are represented on a grid by different sets of fields. A certain number of conformations is kept depending on the minima found in the torsion-angle space of the ligand and these conformers are screened over the binding region or the whole space. This step is very crucial because strictly depends on initial geometry of the ligand. The "survived" conformations are then minimized using a standard molecular force field (OPLS-AA[220]) applying a distance-dependent dielectric model to threat the solvent effect. Eventually, the three to six lowest-energy poses are optimized by Monte Carlo algorithm. The predicting binding affinity and the ranking of the binding pose is performed using GlideScore function (Equation 9) based on ChemScore function.[211]

$$\begin{aligned}
&\Delta G_{bind} \\
&= C_{lipo-lipo} \sum f(r_{lr}) \\
&+ C_{hbond-neut-neut} \sum g(\Delta r) h(\Delta\alpha) \\
&+ C_{hbond-neut-charged} \sum g(\Delta r) h(\Delta\alpha) \\
&+ C_{hbond-charged-charged} \sum g(\Delta r) h(\Delta\alpha) \\
&+ C_{max-metal-ion} \sum f(r_{lm}) + C_{rotb} H_{rotb} \\
&+ C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} \\
&+ solvation\ terms
\end{aligned} \qquad \text{(Eq. 9)}$$

The lipophilic-lipophilic term ($C_{lipo\text{-}lipo}\ \Sigma f(r_{lr})$), hydrogen-bonding terms ($C_{hbond\text{-}neut\text{-}neut}\Sigma g(\Delta r)h(\Delta\alpha)$, $C_{hbond\text{-}neut\text{-}charged}\ \Sigma g(\Delta r)h(\Delta\alpha)$), charged-charged term ($C_{hbond\text{-}charged\text{-}charged}\ \Sigma g(\Delta r)h(\Delta\alpha)$) and metal-ligand interaction term ($C_{max\text{-}matal\text{-}ion}\ \Sigma f(r_{lm})$) are pretty similar defined as in ChemScore. Coulomb ($C_{coul}E_{coul}$), vdW ($C_{vdW}E_{vdW}$) and solvation terms are added.

The polar-hydrophobic term ($C_{polar-phob}V_{polar-phob}$) comes from Schrödinger's active site mapping facility, and represents the contribution of a polar but non-hydrogen-bonding atom is present in a hydrophobic region. Glide uses two forms of GlideScore, the standard precision (SP) and the extra precision (XP). The main differences between SP and XP docking are related to the chemical-physical properties. Actually XP docking applies large desolvation penalties to polar groups of both ligand and protein and takes in consideration specific motif which can enhance binding affinity.[204-205]

1.3.2.3 VINA

AutoDock VINA (acronym of Vina Is Not AutoDock) is one of the most popular molecular docking programs.[215] A search space has to be defined by the user and maximum twenty binding poses per ligand are generated. The generated conformations are ranked by a hybrid scoring function in which empirical information from the conformational preferences of the complex (receptor-ligand) is extracted and used to compare with experimental affinity measurements. The applied scoring function is based on the number of interactions that each conformation can form, following the formula (Equation 10):

$$c = \sum_{i<j} f_{t_i t_{ij}}(r_{ij}) \qquad \text{(Eq. 10)}$$

where i and j represent two atoms and t_i and t_j their types and r_{ij} the interatomic distance, excluding the 1-4 interaction. A symmetric set of interaction functions f_{titj} should be defined. Basically, this sum corresponds to the sum of intermolecular and intramolecular contributions. Finally, the optimization algorithm used is the Iterated Local Search global optimizer.[221-222]

1.3.3 Molecular Mechanics and Molecular Dynamics Simulations

Molecular Mechanics (MM) or classical mechanics describes molecules as a set of bonded atoms whose interactions can be modelled using standard Newton mechanics instead of QM (see 1.3.1 section). The interactions between atoms are modelled with simple parameterized functions based on experimental values or *ab initio* calculations. These functions define the so-called force fields, which are used to easily calculate the energetic function of a system. The molecules are described as an ensemble of spheres representing the atoms and the nuclei positions have a fixed electron distribution, thus simplifying the complexity of the calculations. A force field consists of both the set of equations used to calculate the potential energy and forces from particle coordinates, as well as a collection of parameters used in the equations. The potential energy function (Equation 11) could be divided into two main terms: bonded potential energy (Equation 12) and non-bonded (Equation13).

$$V(R)_{total} = V(R)_{bonded} + V(R)_{non-bonded} \qquad \text{(Eq. 11)}$$

$$V(R)_{bonded} = \sum_{bonds} K_b\,(b - b_o)^2 + \sum_{angles} K_\theta\,(\theta - \theta_o)^2 + \sum_{dihedrals} K_\phi\,[1 + \cos(n\phi - \gamma)] \qquad \text{(Eq. 12)}$$

$$V(R)_{non-bonded} = \sum_{\substack{nonbonded \\ atomparis}} \left(\varepsilon_{ij} \left[\left(\frac{R_{\min ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\varepsilon_D r_{ij}} \right) \qquad \text{(Eq. 13)}$$

The bonded potential energy (Equation 12) describes intramolecular or local contributions (bond stretching, angle bending, and dihedral and improper torsions) to the total energy, and the non-bonded one represents the repulsive and van der Waals interactions (12-6 Lennard-Jones potential) and the Coulombic interactions. The equilibrium values (b_0, θ_0, χ) are typically obtained either from *ab initio* or semi-empirical QM calculations or by fitting to experimental data. In a classical force field, molecules are defined as group of atoms linked by elastic (harmonic) forces, according to the Hook's law. Bond stretching and angle-bending are treated as harmonical function around the equilibrium value (b_0, θ_0) applying a force constant (K_b, K_θ). An oscillatory function is applied to the torsional term with a periodicity (ϕ) between 0 and 2π and a phase (γ) with a force constant (K_χ) (Figure 1.20). The force constant depends on the energy barrier to pass from two local minimum energies. In general, to keep the planarity of certain systems, improper general term is introduced. Improper terms have similar quadratic forms of angles, but with smaller spring constants.

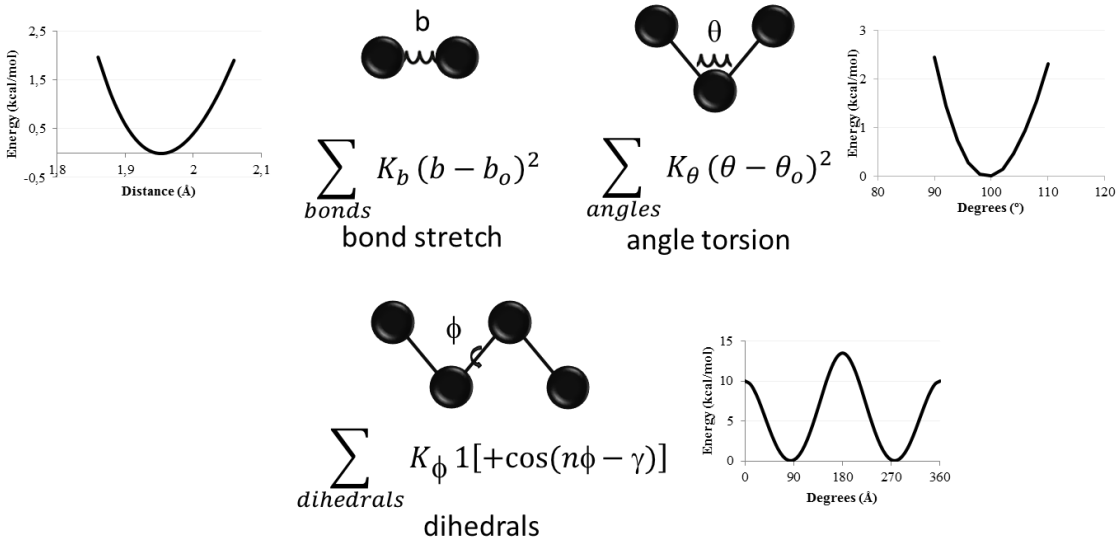


Figure 1.20. Representation of the bonded terms (bonds, angles and dihedrals term) of the force field equation. The atoms are represented as spheres and a bond as a spring.

The non-bonded interactions (Equation 13) consist of Lennard-Jones repulsion and dispersion $\varepsilon_{ij}\left[\left(\frac{R_{\min ij}}{r_{ij}}\right)^{12}-\left(\frac{R_{\min ij}}{r_{ij}}\right)^{6}\right]$ as well as Coulomb electrostatics $\frac{q_i q_j}{\varepsilon_D r_{ij}}$ (Figure 1.21). The 12-6 Lennard-Jones (LJ) potential is used to describe the repulsion between two atoms (*i* and *j*) in which the overlap of the electron clouds of both atoms induces dipoles generating an attractive component. The corresponding well depth (ε_{ij}) (Equation 15) for the interaction between two atoms *i* and *j* is given by the geometric mean of each well depth for atom of the same type (Equation 14).

$$R_{\min ij}=\frac{\left(R_{\min i}+R_{\min ij}\right)}{2} \qquad \text{(Eq.14)}$$

$$\varepsilon_{ij}=\sqrt{\left(\varepsilon_i \varepsilon_j\right)} \qquad \text{(Eq. 15)}$$

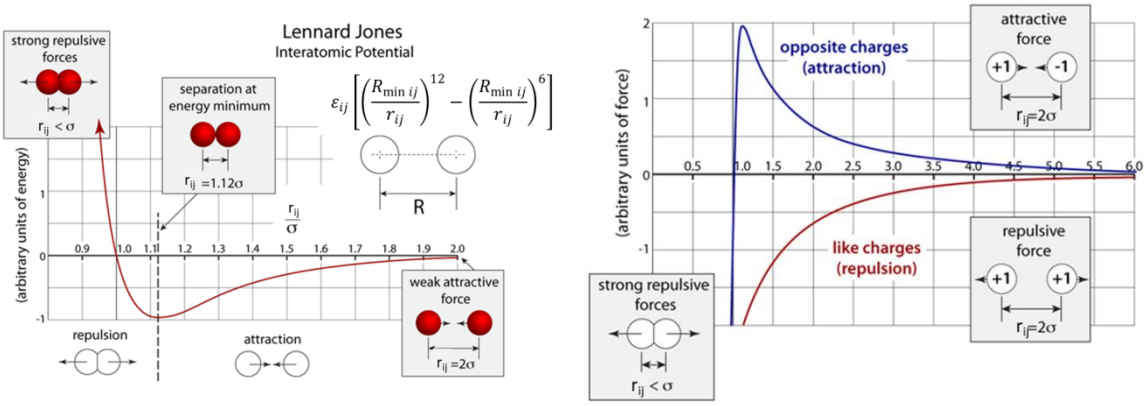


Figure 1.21. Left: Lennard-Jones potential for two atoms Right: Representation of Coulombic electrostatic term (extracted and modified from http://atomsinmotion.com/book/chapter5/md).

The Coulombic term adds the electrostatic contribution due to the partial charges of each atom (q_i and q_j) in distance-dependent manner (r_{ij}), taking into account the permittivity of the solvent (ε_D). The current development of force fields for MD simulations are facing exciting challenges; for example, the development of polarizable force fields[223-224], or the simulation of the behaviour of intrinsically disordered proteins.[225-230] To study biological systems the most accurate approach would use QM calculations, but they are very time consuming.

In this context, MD simulations have become a very important tool to have insights into the conformational dynamics of the system, combined, if available, with NMR experimental information. In particular, MD simulations allow to study at different timescale and in solution, molecules and complexes. This technique has been used to study molecular properties, but also dynamic phenomena such as the interaction between ligand and protein. In MD simulations Newton's second law of motion (Equation 16) is applied to monitor molecular systems over time using empirically derived functions.

$$\frac{d^2x_i}{dt^2} = \frac{F_{xi}}{m_i} \qquad \text{(Eq. 16)}$$

x_i is the Cartesian position of the atom I and F_{xi} represents the force acting on the atom with a mass m_i and t the time. A classical MD simulation consists in different steps. The first step is an energy minimization, in which the search of a minimum in the energy landscaper of a system is performed, starting from higher initial energy state. Two different methods are applied in MD, the steepest gradient method and the conjugate method. In the steepest gradient method, an optimization of the geometry is performed until reaching the local minimum. Different parameters can be customized if the local minimum is not reached. This method depends on the starting geometry and is suitable for quicker optimization. The conjugate gradient method uses subsequent minimization steps by taking information by the previous step.[231] In this phase the solvent environment is very important. Three different methods can be applied to treat the solvent context. In *in vacuo* simulation, a distance-dependent solvent is used. If the solvent is considered implicit, two different models can be applied, Generalized Born (GB) and Poisson-Boltzmann (PB) (see section 1.3.4). With explicit water, the simulation of complex takes place into a box with solvent molecules, under periodic boundary conditions (PBC) to avoid surface artefacts. Different water models have been proposed, but the most used is the TIP3P model in which the HOH angle is fixed to 104.5 and fixed partial charges are assigned (-0.834 for the O and +0.147 for the H).[232] To calculate the infinite electrostatic interactions, particle mesh Ewald summation (PME) is used in which the summation into short- and long-range parts is split.

In the second step, bad contacts between solvent and solute are removed. In this phase, the velocity of the atoms is increased by heating and the velocities are calculated with standard temperature-dependent Maxwell-Boltzmann distribution. Successively there is the equilibrium step, in which the system is relaxed and energy, temperature, volume and pressure are controlled and monitored. Eventually, the production step will prompt the trajectories.

Depending on the problem to solve, the choice of the correct force field is a crucial step. It is clear that the lack of unique force field for all the systems could be seen as a disadvantage but on the other side the variety of force fields for specific systems heads to a high accuracy. Amber package[233] in fact contains several force field specific for lipids (to date Lipid14), for sugars (GLYCAM06), for proteins, nucleic acids and water molecules (ff14SB) and for general organic molecules (GAFF). Those force fields have been regularly used, in combination with structural experimental information, to gain insights into the dynamic and energetic behaviour of molecules and receptor-ligand complexes.

1.3.3.1 AMBER ff14SB

Among the biomolecule systems, proteins are one of the most studied. The computational studies of a protein are strictly related with the parameters set for the system. The most efficient parameters usually are from QM, but the computational cost will be very high. So different force fields for protein have been developed, and Amber ff14SB[234] and its previous versions (ff99SB, ff99 and ff94) are widely used. Different problems are related with the development of a force field for proteins. Amber force field has fixed charges, that are usually less accurate than polarizable force field, but they are still quite reasonable. Another problem related to the simulation of a protein is the prediction and the reproducibility of the protein's secondary structure. This limitation is strictly connected with the definition of the dihedral terms and the energy profile of the dihedrals terms. Parameters are usually calculated by *ab initio* methods and then fitted and validated with experimental results (especially NMR). In particular, the study of the backbone is dictated by two dihedral angles, φ (C−N−Cα−C) and ψ (N−Cα−C−N) (Figure 1.22).

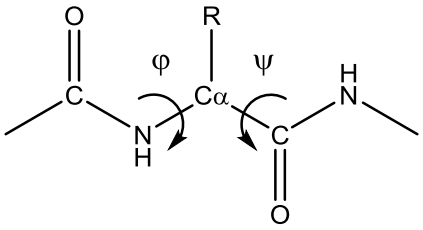


Figure 1.22. Representation of the backbone. φ and ψ are the dihedral angles which are defined in a force field.

1.3.3.2 GLYCAM06

GLYCAM06[181] is the most widely used force field designed for carbohydrates. QM data were employed to compute all torsion terms, derived in a hierarchical manner using fitting valence parameters. This is the only force field in which the same atom type (Cg) at the anomeric carbon (C1) is assigned in both anomers α and β. The anomeric carbon shows a different partial charge, calculated with ensemble-averaged partial charge sets, in α or β anomer. The feature to have just one C atom type facilitates the simulation of ring-flipping, having equilibrium between conformers with axial and equatorial substituents at the anomeric centre. GLYCAM06 is suitable for mono and oligosaccharides, for D and L enantiomers and for all glycosidic linkage possibilities and it is also the only force field to include parameter for *N*-glycosidic linkages. Great caution should be taken before attempt into mix the GLYCAM parameters for carbohydrates with non-AMBER parameter sets. GLYCAM turns the 1-4 non-bonded interactions off to correctly reproduce the rotation of the ω-angle (see 1.1.2 section). This choice aims to simulate the correct population present in each rotamer of this angle (*gg*, *gt*, *tg*), without being influenced by the possibility of O6 to interact with either O4 or O5. GLYCAM is consistent in TIP3P water model. Parameter set was developed for the explicit solvent simulation of hexopyranose-based carbohydrates. This force field has been optimized to enhance the stability of the $^4C_1$ chair conformation and to reproduce the *gauche effect*. In principle, GLYCAM was created to cover a large number of carbohydrates included furanoses[235] and glycolipids in membranes.[236]

1.3.4 MM-PBSA/MM-GBSA

Molecular Mechanics-Poisson-Boltzmann Surface Area (MM-PBSA) and the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) are methods designed to quantitative estimate the binding free energy associated to a complex with implicit solvent. MM-PBSA uses the Poisson-Boltzmann equation to compute the electrostatic contribution to the free-energy; MM-GBSA uses the Generalized Born approximation, which is an approximate and faster treatment of the Poisson-Boltzmann equation.[237]

These two methods combine MM calculations (for the conformational energy term) and continuum solvation model (for the solvation free-energy term) to calculate the binding energy (ΔG) associated to a system. To calculate ΔG, Gibbs free-energy formula is applied (Equation 17)

$$\Delta G = \Delta H - T\Delta S \qquad \text{(Eq. 17)}$$

where ΔG is the variation of the free energy in a system, ΔH the variation of the enthalpy, ΔS the variation of the entropy in a system and T the temperature of the system. Initially the receptor and the ligand are solvated by water molecules. In the bound state, assuming that no covalent bonds are formed between the ligand and the receptor, non–bonded interactions are formed, and an entropic change is correlated with this process. The variation of the entropy is linked with the decrease of the conformational freedom of the ligand and the rearrangement of the water molecules that were surrounding the ligand and the receptor. The AMBER suit of programs allows to calculate the absolute free binding energies by both MM-PBSA and MM-GBSA approaches (Figure 1.23 represents the calculations are made). Basically, the ΔG of the binding is calculated as the difference between the ΔG of the solvated complex (ΔG^{C}_{solv}) and the sum of the ΔG of the solvated ligand and receptor ($\Delta G^{R}_{solv} + \Delta G^{L}_{solv}$), taking also into account the ΔG associated to the complex in vacuum ($\Delta G_{bind,vac}$)(Equation 18).

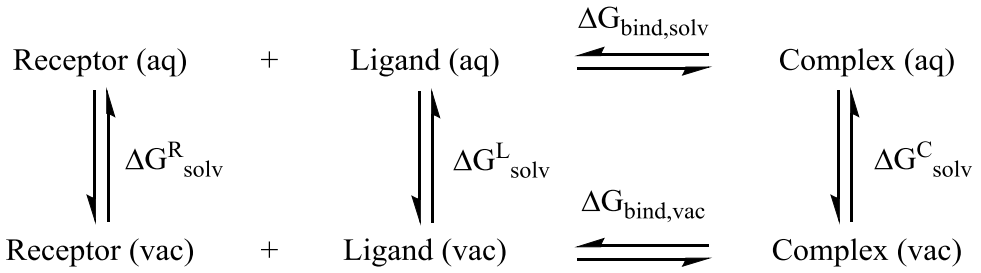


Figure 1.23. Representation of the cycle of the binding free energy calculations for a protein−ligand complex.

$$\Delta G_{bind,solv} = \Delta G_{bind,vac} + \Delta G^{C}_{solv} - (\Delta G^{R}_{solv} + \Delta G^{L}_{solv}) \qquad \text{(Eq. 18)}$$

$\Delta G_{bind,vac}$ takes into account the energy differences associated to bonds, angles and dihedrals upon binding and the electrostatic and van der Waals terms (ΔE_{MM}) but also the conformational entropy change ($T\Delta S$) that can be computed by normal mode analysis (Equation 19).

$$\Delta G_{bind,vac} = \Delta E_{MM} - T\Delta S \qquad \text{(Eq.19)}$$

The solvation free energy (ΔG_{solv}) can be split into two terms, the polar (or electrostatic), and the non-polar (or non-electrostatic) contributions (Equation 20). The polar contribution ($\Delta G_{PB/GB}$) varies in function of the solute-solvent interaction and its contribution can be calculated either with the GB and PB model. The non-polar contribution (ΔG_{SA}) is assessed by solvent accessible surface area (SASA).

$$\Delta G_{solv} = \Delta G_{PB/GB} + \Delta G_{SA} \qquad \text{(Eq.20)}$$

A common strategy to reduce the computational cost and to reduce errors MM-PBSA/MM-GBSA can be run using different snapshots of a MD simulation of the complex (receptor-ligand). In principle PB is more theoretically rigorous than GB, although GB is faster.[237]

1.3.5 Virtual Screening

Virtual screening (VS) is a technique used in ligand-based and structure-based drug discovery field to search hits for a specific target or to find new lead compounds. In this Thesis, we have used a receptor-based virtual screening (RBVS) approach, where the availability of the 3D coordinates of the target is mandatory, either from X-ray crystallography, NMR or homology modelling. Prior knowledge about the ligand binding site may help in the identification of proper binders although, in some approaches, the search for novel binding pockets can be an additional interesting -and challenging-element in the drug discovery process.[238-239] Virtual screening aims to reduce the number of compounds that are experimentally screened and to increase the probability to find new active hits. In a structure-based virtual screening there is a sensor which reflects the likelihood of a ligand to bind in the binding site. Than the chemical structures are compared to the sensor and a scoring function is applied to rank them. The comparison of the sensor and the structures can be based on chemical similarity, on pharmacophore and on shape. FLAP and Glide programs have been used to carry out the virtual screening of fragments data base (Chapter 2) and are briefly outlined below.

1.3.5.1 FLAP

FLAP (Fingerprints for Ligands And Proteins) is a VS program based on molecular fingerprints.[240] In general, for virtual screening, two different approaches can be followed: ligand-based VS (LBVS) or structure-based VS (SBVS). The choice of the approach depends on the information available for the case of study. GRID Molecular Interaction Fields[217] (MIFs) defines the molecular fingerprints, mapping the interaction between a molecule and a probe, a receptor in the SBVS or a ligand in LBSV. MIFs have been widely used in several computational techniques such as 3D-QSAR, ADME and pharmacokinetic modelling. Given that the 3D structure of the receptor is available, the putative binding site can be user-defined by a gridbox in which known aminoacids involved in the binding are included or, alternatively, a pocket-finding algorithm can be applied. The grid points are assigned depending on the chemical structure of the receptor (probe). The grid probes evaluate for example the possibility to have hydrogen bonds or lipophilic interactions between protein and ligand.

The initial number of grid points is reduced to representatives one based on a weighted energy function. Then, four points "hotspots" are generated and the distances between the points, named also *Common Reference Framework*, are calculated. The hotspots will define the pharmacophore fingerprint. After generating the MIFs of the receptor, during the screening, FLAP compares the MIFs of the binding site with the MIFs of the ligands. A data set of ligands will be screened, and molecules will be discarded or not depending on the overlapping with the hotspots. To quantify the similarity between the MIFs, a scoring function based on Tanimoto coefficient is applied. The output will give information about the scoring of each MIF, the sum of them (Glob-Sum) or the product (Glop-Prod) for each ligand. Once the initial searching and alignment is performed, only the best-scored conformer for each molecule is maintained.

1.3.5.2 Glide

Glide[204-205] uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. It has a systematic method to treat ligand flexibility, with an exhaustive search algorithm. The Glide protocol is intuitive and relies on four steps: the ligands and protein preparation, the receptor grid generation, and the docking process. Before launching the docking step, Glide has to generate a grid that represents the shape and the properties of the receptor, using several different sets of fields that provide progressively more accurate scoring of ligand poses. The grid permits to dock only the relevant region of the receptor, thus saving calculation time. Regarding the last point, the full docking VS workflow includes three docking stages: HTVS (High Throughput Virtual Screening), SP (Standard Precision) and XP (Extra Precision). The first stage performs HTVS docking. It is intended for rapid screening of a very large number of ligands and has much more restricted conformational sampling than SP docking. The second stage performs SP docking. It is appropriate for screening ligands of unknown quality in large numbers. The third stage is the XP docking and scoring. It is a more powerful and discriminating procedure using an implementation of a modified and expanded version of the ChemScore scoring function, called GlideScore and categorized as an empirical scoring function.

Glide can be used to perform virtual screening, accurate binding mode precision and, furthermore, Glide exhibits excellent docking accuracy and high enrichment across a diverse range of receptor types.

1.3.6 Conformational analysis

Molecular properties are deeply linked with the three-dimensional structure, or conformation. So the conformational analysis is very important to determine the characteristic related to a molecule. Especially in carbohydrate field this technique helps to understand sugar behaviour in water solution or in complex with a protein. This method could be used either alone or in combination with experimental techniques, usually solution state NMR spectroscopy methods. In fact, the combination of modelling/NMR protocols has been extremely useful to deduce the conformational and dynamic properties of free and bound carbohydrate molecules.[13, 179-180] In this sense, very useful are carbohydrate databases (such as http://glycosciences.de) that collect available NMR data and X-ray experiment of particular dihedral angles, chair conformations and glycosidic linkages.

The conformational search aims to find the minima local energy for a given molecule in order to correlate a molecule with its behaviour. The conformations are usually defined as a rearrangement of the atoms in space simply rotating single bonds. So the minimization of energy is a crucial step in this protocol. Different techniques can be used to achieve the lowest energy conformer from *ab initio* calculations, semi-empirical methods or MD simulations. Moreover, this technique is also applied for docking studies, in fact it represents the first step of the docking protocol (see Section 1.3.2). In this Thesis all the three methods were used. In particular, for *ab initio* calculation DFT methods were applied either in combination with semi-empirical method (in most of the cases MM3*) or/and MD simulation.

1.3.7 Homology modelling

The availability of the 3D structure of a given protein is a limiting step in the drug design and development process. Usually, the 3D structures of the macromolecules are elucidated by NMR or X-ray crystallography techniques, and they are deposited in the public database Protein Data Bank (PDB).[111] When the 3D structure for a protein of interest has not been resolved by experimental techniques, homology modelling is very powerful to predict the structure of a protein. In order to build a homology model, the availability of the experimental 3D structure (template) or multiple structures is required. Starting from the protein sequence (target) available on the *Universal Protein Resource* database (UNIPROT),[241] a sequence alignment with the template is performed. This procedure is online accessible by means of *Basic Local Alignment Search Tool* (BLAST).[242] The scoring of BLAST provides to estimate the quality of the sequence alignment using the BLOSUM62 (BLOck SUbstitution Matrix).[243] In the score assignment, different parameters are taken into account such as the percentage of identity between the searched protein and the template (a good model has minimum 30% of the percentage identity)[244] and the percentage of query cover. To build a homology model three different protocols can be applied: *ab initio* calculations (see Section 1.3.1), fold recognition and comparative modelling. With *ab initio* is possible to build small peptides. Fold recognition (or threading) protocol recognizes specific sequences whose folding motif is already know from experimentally elucidated 3D structures. With comparative modelling, the templates from the PDB are assumed to be homologous to the target and the templates guide the building of the three-dimensional model. In this Thesis, we uses the Phyre2 web server,[245] which uses an ensemble fold recognition method with an in-house algorithm.[246] Once the model is built, a subsequent refinement step, especially for the side chain, is necessary. This step can be carry out by MD simulation, restraining the Cα of the backbone.

Validation of the model can be performed by using different tools. In our studies, we have used the SAVES tool[247] which validates the homology modelling prediction analysing different parameters. Overall structure geometry and the statistics of non-bonded interactions is analysed and compared with highly refined structures. Moreover it inspects the compatibility between the predicted secondary structure for a protein with its own amino acid sequence. Also atomic volumes and the possibly similar structures are also checked and compared with existing crystal structures in PDB. SAVES is able to identify the presence of residue conformations that are not the expected ones. For example, the representation of the Ramachandran Plot is a useful tool to check backbone angles of the protein structure by comparing them with observed/experimental angles. A good homology model usually has a score of more than 90% meaning that at least 90% of the residues are in the most favourable energy region.

1.4 Objectives

In this Thesis, the general objective is to elucidate the carbohydrate-protein interactions at the atomic level through computational techniques. In particular, the following systems will be studied: human galectins -1, -3, and -7, *Pisum sativum* lectin, *Maackia amurensis* seed lectin, and glycosyltransferase GalNAc-T2. The final aim is to provide new insights for the understanding of the molecular recognition events underlying the biological functions of these proteins. These studies will be carried out by addressing the following specific objectives:

Chapter 2: Elucidation of the binding modes of fluorinated glycans with a model lectin from *Pisum sativum*. We seek the elucidation of the binding modes of a fluorinated trimannoside α1→3 and α1→6 linked and its fluorinated disaccharide derivatives. The molecular modelling studies will take advantage of ^{19}F-NMR experimental data to unequivocally elucidate the binding mode of these ligands in particular for the non-fluorinated natural mannosides that has remained unresolved.

Chapter 3: Computational study of the binding mode of several β-galactoside ligands towards human galectin-3. These ligands are Lac, LacdiNAc, and and three glycans containing the LacNAc motif: blood group A type II tetrasaccharide, and two compounds derived from the important xenoantigen α-Gal epitope. We seek the characterization of the key interactions required for the distinct recognition by galectin-3, which has been related to immune responses.

Chapter 4: Computational design of novel selective galectin binders with improved affinity. By means of virtual screening protocols, the vicinal pocket close to the carbohydrate recognition domain will be explored to identify appropriate moieties able to be anchored to this pocket. We seek the design of modified OMe-Lac derivatives able to bind the galectins with improved affinity and with selectivity towards the different galectin subtypes. Selected candidates will be synthesized by collaborators and their affinity will be measured by biophysical techniques.

Chapter 5: Computational study of the molecular recognition process of podoplanin by *Maackia amurensis* seed lectin (MASL). We explore the binding site of the two molecular species of the MASL, *Maackia amurensis* leukoagglutinin and *Maackia amurensis* hemagglutinin, and the possible recognition by the epitope of the glycoprotein podoplanin as a plausible mechanism for the protective activity of this lectin in osteoarthritis.

Chapter 6: Computational study of the binding mode of a series of synthetic nucleoside mimetics as inhibitors of a model glycosyltransferase (GalNAc-T2). These less-polar nucleotide sugar analogues derived from uridine only containing the β-phosphate could be efficient ligands for the enzyme, as deduced from the experimental data. We seek the proposal of a computational 3D model for the binding mode of these ligands to provide further insights into the mechanism of the catalytic cycle of this family of enzyme, useful for the further design of more potent inhibitors.

Bibliography

1. Gabius, H.-J.; Siebert, H. C.; André, S.; Jiménez-Barbero, J.; Rudiger, H., Chemical biology of the sugar code. *ChemBioChem* **2004,** *5* (6), 740-764.
2. Fraser-Reid, B. O.; Tatsuta, K.; Thiem, J., Glycoscience. Springer: 2008; Vol. 2nd Edition.
3. Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., *Essential of Glycobiology*. Cold Spring Harbor Laboratory Press: 2009; Vol. 2nd Edition.
4. Gabius, H.-J.; André, S.; Kaltner, H.; Siebert, H. C., The sugar code: functional lectinomics. *Biochim. Biophys. Acta* **2002,** *1572* (2-3), 165-177.
5. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009,** *5* (10), 1087-1104.
6. Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **2011,** *36* (6), 298-313.
7. Mallajosyula, S. S.; MacKerell, A. D., Jr., Influence of solvent and intramolecular hydrogen bonding on the conformational properties of O-linked glycopeptides. *J. Phys. Chem. B* **2011,** *115* (38), 11215-11229.
8. Fadda, E.; Woods, R. J., Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* **2010,** *15* (15-16), 596-609.
9. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (19), 10541-10545.
10. Dabrowski, J.; Kozar, T.; Grosskurth, H.; Nifant'ev, N. E., Conformational mobility of oligosaccharides: experimental evidence for the existence of an "Anti" conformer of the Gal-beta-1-3-Glc-beta-1-OMe disaccharide. *J. Am. Chem. Soc.* **1995,** *117* (20), 5534-5539.
11. Asensio, J. L.; Jiménez-Barbero, J., The use of the AMBER force field in conformational analysis of carbohydrate molecules: determination of the solution conformation of methyl alpha-lactoside by NMR spectroscopy, assisted by molecular mechanics and dynamics calculations. *Biopolymers* **1995,** *35* (1), 55-73.
12. Widmalm, G., A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr. Res.* **2013,** *378*, 123-132.
13. Vidal, P.; Roldos, V.; Fernandez-Alonso Mdel, C.; Vauzeilles, B.; Bleriot, Y.; Cañada, F. J.; André, S.; Gabius, H.-J.; Jiménez-Barbero, J.; Espinosa, J. F.; Martin-Santamaria, S., Conformational selection in glycomimetics: human galectin-1 only recognizes syn-Psi-type conformations of beta-1,3-linked lactose and its C-glycosyl derivative. *Chem. Eur. J.* **2013,** *19* (43), 14581-14590.
14. Martin-Santamaria, S.; André, S.; Buzamet, E.; Caraballo, R.; Fernandez-Cureses, G.; Morando, M.; Ribeiro, J. P.; Ramirez-Gualito, K.; de Pascual-Teresa, B.; Cañada, F. J.; Menendez, M.; Ramstrom, O.; Jiménez-Barbero, J.; Solis, D.; Gabius, H.-J., Symmetric dithiodigalactoside: strategic combination of binding studies and detection of selectivity between a plant toxin and human lectins. *Org. Biomol. Chem.* **2011,** *9* (15), 5445-5455.

15. Suzuki, T.; Makyio, H.; Ando, H.; Komura, N.; Menjo, M.; Yamada, Y.; Imamura, A.; Ishida, H.; Wakatsuki, S.; Kato, R.; Kiso, M., Expanded potential of seleno-carbohydrates as a molecular tool for X-ray structural determination of a carbohydrate-protein complex with single/multi-wavelength anomalous dispersion phasing. *Biorg. Med. Chem.* **2014,** *22* (7), 2090-2101.
16. Sanchez-Fernandez, E. M.; Risquez-Cuadro, R.; Aguilar-Moncayo, M.; Garcia-Moreno, M. I.; Ortiz Mellet, C.; Garcia Fernandez, J. M., Generalized anomeric effect in gem-diamines: stereoselective synthesis of alpha-N-linked disaccharide mimics. *Org. Lett.* **2009,** *11* (15), 3306-3309.
17. Asensio, J. L.; Siebert, H. C.; von Der Lieth, C. W.; Laynez, J.; Bruix, M.; Soedjanaamadja, U. M.; Beintema, J. J.; Cañada, F. J.; Gabius, H.-J.; Jiménez-Barbero, J., NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacetylchitotriose. *Proteins* **2000,** *40* (2), 218-236.
18. Espinosa, J.-F.; Cañada, F. J.; Asensio, J. L.; Martín-Pastor, M.; Dietrich, H.; Martín-Lomas, M.; Schmidt, R. R.; Jiménez-Barbero, J., Experimental evidence of conformational differences between C-glycosides and O-glycosides in solution and in the protein-bound State: the C-lactose/O-lactose case. *J. Am. Chem. Soc.* **1996,** *118* (44), 10862-10871.
19. Garcia-Herrero, A.; Montero, E.; Munoz, J. L.; Espinosa, J. F.; Vian, A.; Garcia, J. L.; Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J., Conformational selection of glycomimetics at enzyme catalytic sites: experimental demonstration of the binding of distinct high-energy distorted conformations of C-, S-, and O-glycosides by E. Coli beta-galactosidases. *J. Am. Chem. Soc.* **2002,** *124* (17), 4804-4810.
20. Gonzalez, L.; Asensio, J. L.; Ariosa-Alvarez, A.; Verez-Bencomo, V.; Jiménez-Barbero, J., Solution conformation and dynamics of the trisaccharide fragments of the O-antigen of Vibrio cholerae O1, serotypes Inaba and Ogawa. *Carbohydr. Res.* **1999,** *321* (1-2), 88-95.
21. Montero, E.; Vallmitjana, M.; Perez-Pons, J. A.; Querol, E.; Jiménez-Barbero, J.; Cañada, F. J., NMR studies of the conformation of thiocellobiose bound to a beta-glucosidase from Streptomyces sp. *FEBS Lett.* **1998,** *421* (3), 243-248.
22. Sattelle, B. M.; Almond, A., Shaping up for structural glycomics: a predictive protocol for oligosaccharide conformational analysis applied to N-linked glycans. *Carbohydr. Res.* **2014,** *383*, 34-42.
23. Almond, A.; Duus, J. O., Quantitative conformational analysis of the core region of N-glycans using residual dipolar couplings, aqueous molecular dynamics, and steric alignment. *J. Biomol. NMR* **2001,** *20* (4), 351-363.
24. Rinnbauer, M.; Ernst, B.; Wagner, B.; Magnani, J.; Benie, A. J.; Peters, T., Epitope mapping of sialyl Lewis(x) bound to E-selectin using saturation transfer difference NMR experiments. *Glycobiology* **2003,** *13* (6), 435-443.
25. Sayers, E. W.; Prestegard, J. H., Solution conformations of a trimannoside from nuclear magnetic resonance and molecular dynamics simulations. *Biophys. J.* **2000,** *79* (6), 3313-3329.
26. Canales, A.; Mallagaray, A.; Perez-Castells, J.; Boos, I.; Unverzagt, C.; André, S.; Gabius, H.-J.; Cañada, F. J.; Jiménez-Barbero, J., Breaking pseudo-symmetry in

multiantennary complex N-glycans using lanthanide-binding tags and NMR pseudo-contact shifts. *Angew. Chem. Int. Ed.* **2013,** *52* (51), 13789-13793.
27. Canales, A.; Mallagaray, A.; Berbis, M. A.; Navarro-Vazquez, A.; Dominguez, G.; Cañada, F. J.; André, S.; Gabius, H.-J.; Perez-Castells, J.; Jiménez-Barbero, J., Lanthanide-chelating carbohydrate conjugates are useful tools to characterize carbohydrate conformation in solution and sensitive sensors to detect carbohydrate-protein interactions. *J. Am. Chem. Soc.* **2014,** *136* (22), 8011-8017.
28. Cighetti, R.; Ciaramelli, C.; Sestito, S. E.; Zanoni, I.; Kubik, Ł.; Ardá-Freire, A.; Calabrese, V.; Granucci, F.; Jerala, R.; Martín-Santamaría, S., Modulation of CD14 and TLR4· MD-2 Activities by a synthetic lipid A mimetic. *ChemBioChem* **2014,** *15* (2), 250-258.
29. Marchetti, R.; Canales, A.; Lanzetta, R.; Nilsson, I.; Vogel, C.; Reed, D. E.; Aucoin, D. P.; Jiménez-Barbero, J.; Molinaro, A.; Silipo, A., Unraveling the interaction between the LPS O-antigen of Burkholderia anthina and the 5D8 monoclonal antibody by using a multidisciplinary chemical approach, with synthesis, NMR, and molecular modeling methods. *ChemBioChem* **2013,** *14* (12), 1485-1493.
30. Mulder, L.; Lefebvre, B.; Cullimore, J.; Imberty, A., LysM domains of Medicago truncatula NFP protein involved in Nod factor perception. Glycosylation state, molecular modeling and docking of chitooligosaccharides and Nod factors. *Glycobiology* **2006,** *16* (9), 801-809.
31. Rouge, P.; Nerinckx, W.; Gough, C.; Bono, J. J.; Barre, A., Docking of chitin oligomers and Nod factors on lectin domains of the LysM-RLK receptors in the Medicago-Rhizobium symbiosis. *Adv. Exp. Med. Biol.* **2011,** *705*, 511-521.
32. Esko, J. D.; Kimata, K.; Lindahl, U., Proteoglycans and sulfated glycosaminoglycans. In *Essentials of Glycobiology*, Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press The Consortium of Glycobiology Editors, La Jolla, California.: Cold Spring Harbor (NY), 2009.
33. Sasisekharan, R.; Raman, R.; Prabhakar, V., Glycomics approach to structure-function relationships of glycosaminoglycans. *Annu. Rev. Biomed. Eng.* **2006,** *8*, 181-231.
34. Gargiulo, V.; Morando, M. A.; Silipo, A.; Nurisso, A.; Perez, S.; Imberty, A.; Cañada, F. J.; Parrilli, M.; Jiménez-Barbero, J.; De Castro, C., Insights on the conformational properties of hyaluronic acid by using NMR residual dipolar couplings and MD simulations. *Glycobiology* **2010,** *20* (10), 1208-1216.
35. Pomin, V. H., Solution NMR conformation of glycosaminoglycans. *Prog. Biophys. Mol. Biol.* **2014,** *114* (2), 61-68.
36. Rudd, T. R.; Skidmore, M. A.; Guerrini, M.; Hricovini, M.; Powell, A. K.; Siligardi, G.; Yates, E. A., The conformation and structure of GAGs: recent progress and perspectives. *Curr. Opin. Struct. Biol.* **2010,** *20* (5), 567-574.
37. Gallus, A. S.; Coghlan, D. W., Heparin pentasaccharide. *Curr. Opin. Hematol.* **2002,** *9* (5), 422-429.
38. Reverter, J. C., Fondaparinux sodium. *Drugs Today (Barc)* **2002,** *38* (3), 185-194.
39. Nieto, L.; Canales, Á.; Fernández, I. S.; Santillana, E.; González-Corrochano, R.; Redondo-Horcajo, M.; Cañada, F. J.; Nieto, P.; Martín-Lomas, M.; Giménez-Gallego, G., Heparin modulates the mitogenic activity of fibroblast growth factor by inducing

dimerization of its receptor. A 3D view by using NMR. *ChemBioChem* **2013,** *14* (14), 1732-1744.
40. Nieto, L.; Canales, A.; Gimenez-Gallego, G.; Nieto, P. M.; Jiménez-Barbero, J., Conformational selection of the AGA*IA(M) heparin pentasaccharide when bound to the fibroblast growth factor receptor. *Chem. Eur. J.* **2011,** *17* (40), 11204-11209.
41. Choi, J. K.; Lee, B. H.; Chae, C. H.; Shin, W., Computer modeling of the rhamnogalacturonase-"hairy" pectin complex. *Proteins* **2004,** *55* (1), 22-33.
42. Wu, Y.; Ai, L.; Wu, J.; Cui, S. W., Structural analysis of a pectic polysaccharide from boat-fruited sterculia seeds. *Int. J. Biol. Macromol.* **2013,** *56*, 76-82.
43. Franca, E. F.; Lins, R. D.; Freitas, L. C. G.; Straatsma, T. P., Characterization of chitin and chitosan molecular structure in aqueous solution. *J. Chem. Theory Comput.* **2008,** *4* (12), 2141-2149.
44. Lacetera, A.; Galante, S.; Jiménez-Barbero, J.; Martín-Santamaría, S., Glycans in medicinal chemistry. Reedijk, J. E. E. R. M. i. C., Molecular Sciences and Chemical Engineering. Waltham, MA, Ed. 2016.
45. Mingeot-Leclercq, M.-P.; Glupczynski, Y.; Tulkens, P. M., Aminoglycosides: activity and resistance. *Antimicrob. Agents Chemother.* **1999,** *43* (4), 727-737.
46. Jones, D.; Metzger, H.; Schatz, A.; Waksman, S. A., Control of gram-negative bacteria in experimental animals by streptomycin. *Science* **1944,** *100* (2588), 103-105.
47. Magnet, S.; Blanchard, J. S., Molecular insights into aminoglycoside action and resistance. *Chem. Rev.* **2005,** *105* (2), 477-498.
48. Spízek, J.; Novotná, J.; Rezanka, T., Lincosamides: chemical structure, biosynthesis, mechanism of action, resistance, and applications. *Adv. Appl. Microbiol.* **2004,** *56*, 121.
49. Liang, J.-H.; Han, X., Structure-activity relationships and mechanism of action of macrolides derived from erythromycin as antibacterial agents. *Curr. Top. Med. Chem.* **2013,** *13* (24), 3131-3164.
50. Zielińska, J.; Wieczór, M.; Bączek, T.; Gruszecki, M.; Czub, J., Thermodynamics and kinetics of amphotericin B self-association in aqueous solution characterized in molecular detail. *Sci. Rep.* **2016,** *6*.
51. Gray, K. C.; Palacios, D. S.; Dailey, I.; Endo, M. M.; Uno, B. E.; Wilcock, B. C.; Burke, M. D., Amphotericin primarily kills yeast by simply binding ergosterol. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (7), 2234-2239.
52. Zwick, M. B.; Saphire, E. O.; Burton, D. R., gp41: HIV's shy protein. *Nature Medicine* **2004,** *10* (2), 133-134.
53. Bernardi, A.; Jiménez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K.-E.; Lahmann, M., Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev.* **2013,** *42* (11), 4709-4727.
54. François, K.; Balzarini, J., Potential of carbohydrate-binding agents as therapeutics against enveloped viruses. *Med. Res. Rev.* **2012,** *32* (2), 349-387.
55. Kilby, J. M.; Eron, J. J., Novel therapies based on mechanisms of HIV-1 cell entry. *N. Engl. J. Med.* **2003,** *348* (22), 2228-2238.
56. Balzarini, J.; Van Laethem, K.; Daelemans, D.; Hatse, S.; Bugatti, A.; Rusnati, M.; Igarashi, Y.; Oki, T.; Schols, D., Pradimicin A, a carbohydrate-binding nonpeptidic lead compound for treatment of infections with viruses with highly glycosylated envelopes, such as human immunodeficiency virus. *J. Virol.* **2007,** *81* (1), 362-373.

57. Shahzad-ul-Hussan, S.; Ghirlando, R.; Dogo-Isonagie, C. I.; Igarashi, Y.; Balzarini, J.; Bewley, C. A., Characterization and carbohydrate specificity of pradimicin S. *J. Am. Chem. Soc.* **2012,** *134* (30), 12346-12349.
58. Roy, R., New trends in carbohydrate-based vaccines. *Drug Discov. Today Technol.* **2004,** *1* (3), 327-336.
59. Jin, S.-D.; Kim, Y.-M.; Kang, H.-K.; Jung, S.-J.; Kim, D., Optimization of capsular polysaccharide production by Streptococcus pneumoniae type 3. *J. Microbiol. Biotechnol.* **2009,** *19* (11), 1374-1378.
60. Bertolo, L.; Ewing, C. P.; Maue, A.; Poly, F.; Guerry, P.; Monteiro, M. A., The design of a capsule polysaccharide conjugate vaccine against Campylobacter jejuni serotype HS15. *Carbohydr. Res.* **2013,** *366*, 45-49.
61. Morelli, L.; Poletti, L.; Lay, L., Carbohydrates and immunology: synthetic oligosaccharide antigens for vaccine formulation. *Eur. J. Org. Chem.* **2011,** *2011* (29), 5723-5777.
62. Pozsgay, V., Recent developments in synthetic oligosaccharide-based bacterial vaccines. *Curr. Top. Med. Chem.* **2008,** *8* (2), 126-140.
63. Seeberger, P. H.; Werz, D. B., Synthesis and medical applications of oligosaccharides. *Nature* **2007,** *446* (7139), 1046-1051.
64. Segal, S.; Pollard, A., Vaccines against bacterial meningitis. *Br. Med. Bull.* **2004,** *72* (1), 65-81.
65. González-Fernández, Á.; Faro, J.; Fernández, C., Immune responses to polysaccharides: lessons from humans and mice. *Vaccine* **2008,** *26* (3), 292-300.
66. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **2015,** *15* (9), 540-555.
67. Cipolla, L.; Peri, F., Carbohydrate-based bioactive compounds for medicinal chemistry applications. *Mini. Rev. Med. Chem.* **2011,** *11* (1), 39-54.
68. Marcos, N. T.; Bennett, E. P.; Gomes, J.; Magalhaes, A.; Gomes, C.; David, L.; Dar, I.; Jeanneau, C.; DeFrees, S.; Krustrup, D., ST6GalNAc-I controls expression of sialyl-Tn antigen in gastrointestinal tissues. *Front. Biosci. (Elite Ed)* **2011,** *3*, 1443-1455.
69. Julien, S.; Adriaenssens, E.; Ottenberg, K.; Furlan, A.; Courtand, G.; Vercoutter-Edouart, A.-S.; Hanisch, F.-G.; Delannoy, P.; Le Bourhis, X., ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumourigenicity. *Glycobiology* **2006,** *16* (1), 54-64.
70. Kannagi, R.; Yin, J.; Miyazaki, K.; Izawa, M., Current relevance of incomplete synthesis and neo-synthesis for cancer-associated alteration of carbohydrate determinants—Hakomori's concepts revisited. *BBA. General subjects* **2008,** *1780* (3), 525-531.
71. Zhu, J.; Wan, Q.; Lee, D.; Yang, G.; Spassova, M. K.; Ouerfelli, O.; Ragupathi, G.; Damani, P.; Livingston, P. O.; Danishefsky, S. J., From synthesis to biologics: preclinical data on a chemistry derived anticancer vaccine. *J. Am. Chem. Soc.* **2009,** *131* (26), 9298-9303.
72. Bennett, E. P.; Mandel, U.; Clausen, H.; Gerken, T. A.; Fritz, T. A.; Tabak, L. A., Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. *Glycobiology* **2012,** *22* (6), 736-756.
73. Hakomori, S., Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (16), 10231-10233.

74. Arnold, J. N.; Saldova, R.; Hamid, U. M. A.; Rudd, P. M., Evaluation of the serum n-linked glycome for the diagnosis of cancer and chronic inflammation. *Proteomics* **2008,** *8* (16), 3284-3293.
75. Wilson, R. M.; Danishefsky, S. J., A vision for vaccines built from fully synthetic tumor-associated antigens: from the laboratory to the clinic. *J. Am. Chem. Soc.* **2013,** *135* (39), 14462-14472.
76. Kovjazin, R.; Horn, G.; Smorodinsky, N. I.; Shapira, M. Y.; Carmon, L., Cell surface-associated anti-MUC1-derived signal peptide antibodies: implications for cancer diagnostics and therapy. *PloS one* **2014,** *9* (1), e85400.
77. Kaufmann, S. H., Induction of endonucleolytic DNA cleavage in human acute myelogenous leukemia cells by etoposide, camptothecin, and other cytotoxic anticancer drugs: a cautionary note. *Cancer Res.* **1989,** *49* (21), 5870-5878.
78. Kay, N. E.; Geyer, S. M.; Call, T. G.; Shanafelt, T. D.; Zent, C. S.; Jelinek, D. F.; Tschumper, R.; Bone, N. D.; Dewald, G. W.; Lin, T. S., Combination chemoimmunotherapy with pentostatin, cyclophosphamide, and rituximab shows significant clinical activity with low accompanying toxicity in previously untreated B chronic lymphocytic leukemia. *Blood* **2007,** *109* (2), 405-411.
79. Goldstein, M. J.; Mitchell, E. P., Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. *Cancer Invest.* **2005,** *23* (4), 338-351.
80. Felder, M.; Kapur, A.; Gonzalez-Bosquet, J.; Horibata, S.; Heintz, J.; Albrecht, R.; Fass, L.; Kaur, J.; Hu, K.; Shojaei, H., MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress. *Mol. Cancer* **2014,** *13* (1), 1.
81. Wong, N. K.; Easton, R. L.; Panico, M.; Sutton-Smith, M.; Morrison, J. C.; Lattanzio, F. A.; Morris, H. R.; Clark, G. F.; Dell, A.; Patankar, M. S., Characterization of the oligosaccharides associated with the human ovarian tumor marker CA125. *J. Biol. Chem.* **2003,** *278* (31), 28619-28634.
82. Peracaula, R.; Tabarés, G.; Royle, L.; Harvey, D. J.; Dwek, R. A.; Rudd, P. M.; de Llorens, R., Altered glycosylation pattern allows the distinction between prostate-specific antigen (PSA) from normal and tumor origins. *Glycobiology* **2003,** *13* (6), 457-470.
83. Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L., Alterations in glycosylation as biomarkers for cancer detection. *J. Clin. Pathol.* **2010,** *63* (4), 322-329.
84. Locker, G. Y.; Hamilton, S.; Harris, J.; Jessup, J. M.; Kemeny, N.; Macdonald, J. S.; Somerfield, M. R.; Hayes, D. F.; Bast, R. C., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J. Clin. Oncol.* **2006,** *24* (33), 5313-5327.
85. Kaur, J., A comprehensive review on metabolic syndrome. *Cardiol. Res. Pract.* **2014,** *2014*.
86. Mooradian, A. D.; Thurman, J. E., Drug therapy of postprandial hyperglycaemia. *Drugs* **1999,** *57* (1), 19-29.
87. Phillips, L. K.; Deane, A. M.; Jones, K. L.; Rayner, C. K.; Horowitz, M., Gastric emptying and glycaemia in health and diabetes mellitus. *Nat. Rev. Endocrinol.* **2015,** *11* (2), 112-128.
88. Kato, T.; Node, K., Therapeutic potential of α-glucosidase inhibitors to prevent postprandial endothelial dysfunction. *Int. Heart J.* **2014,** *55* (5), 386-390.
89. Scheen, A., Clinical efficacy of acarbose in diabetes mellitus: a critical review of controlled trials. *Diabetes Metab.* **1998,** *24* (4), 311-320.

90. Campo, V. L.; Aragao-Leoneti, V.; Carvalho, I., Glycosidases and diabetes: metabolic changes, mode of action and therapeutic perspectives. *Carbohydr. Chem.* **2013,** *39*, 181-203.
91. Chiasson, J.-L.; Josse, R. G.; Gomis, R.; Hanefeld, M.; Karasik, A.; Laakso, M.; Group, S.-N. T. R., Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *The Lancet* **2002,** *359* (9323), 2072-2077.
92. Dabhi, A. S.; Bhatt, N. R.; Shah, M. J., Voglibose: an alpha glucosidase inhibitor. *J. Clin. Diagn. Res.* **2013,** *7* (12), 3023.
93. Asano, N.; Nash, R. J.; Molyneux, R. J.; Fleet, G. W., Sugar-mimic glycosidase inhibitors: natural occurrence, biological activity and prospects for therapeutic application. *Tetrahedron: Asymmetry* **2000,** *11* (8), 1645-1680.
94. Watson, A. A.; Fleet, G. W.; Asano, N.; Molyneux, R. J.; Nash, R. J., Polyhydroxylated alkaloids-natural occurrence and therapeutic applications. *Phytochemistry* **2001,** *56* (3), 265-295.
95. Chen, Z.; Hao, J.; Wang, L.; Wang, Y.; Kong, F.; Zhu, W., New α-glucosidase inhibitors from marine algae-derived Streptomyces sp. OUCMDZ-3434. *Sci. Rep.* **2016,** *6*, 20004.
96. Platt, F. M., Sphingolipid lysosomal storage disorders. *Nature* **2014,** *510* (7503), 68-75.
97. Gloster, T. M.; Vocadlo, D. J., Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology. *Nat. Chem. Biol.* **2012,** *8* (8), 683-694.
98. Bernardi, A.; Sattin, S., Carbohydrates: a phenol sandwich fights diabetes. *Nat. Chem. Biol.* **2015,** *11* (9), 635-636.
99. Doores, K. J., The HIV glycan shield as a target for broadly neutralizing antibodies. *FEBS J.* **2015,** *282* (24), 4679-4691.
100. Cipolla, L.; Fernandes, M. R.; Gregori, M.; Airoldi, C.; Nicotra, F., Synthesis and biological evaluation of a small library of nojirimycin-derived bicyclic iminosugars. *Carbohydr. Res.* **2007,** *342* (12), 1813-1830.
101. Molinaro, A.; Holst, O.; Di Lorenzo, F.; Callaghan, M.; Nurisso, A.; D'Errico, G.; Zamyatina, A.; Peri, F.; Berisio, R.; Jerala, R., Chemistry of lipid A: at the heart of innate immunity. *Chem. Eur. J.* **2015,** *21* (2), 500-519.
102. Jerala, R., Structural biology of the LPS recognition. *Int. J. Med. Microbiol.* **2007,** *297* (5), 353-363.
103. Tan, Y.; Kagan, J. C., A cross-disciplinary perspective on the innate immune responses to bacterial lipopolysaccharide. *Mol. Cell.* **2014,** *54* (2), 212-223.
104. Beutler, B., TLR4: central component of the sole mammalian LPS sensor. *Curr. Opin. Immunol.* **2000,** *12* (1), 20-26.
105. Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O., The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **2009,** *458* (7242), 1191-1195.
106. Garner, O. B.; Baum, L. G., Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling. *Biochem. Soc. Trans.* **2008,** *36* (Pt 6), 1472-1477.
107. Fukuhara, H.; Furukawa, A.; Maenaka, K., New binding face of C-type lectin-like domains. *Structure (London, England : 1993)* **2014,** *22* (12), 1694-1696.
108. Sharon, N.; Lis, H., Lectins: cell-agglutinating and sugar-specific proteins. *Science* **1972,** *177* (4053), 949-959.

109. van Kooyk, Y.; Geijtenbeek, T. B., DC-SIGN: escape mechanism for pathogens. *Nat. Rev. Immunol.* **2003,** *3* (9), 697-709.
110. Hart, G. W.; Copeland, R. J., Glycomics hits the big time. *Cell* **2010,** *143* (5), 672-676.
111. Protein Data Bank. http://www.rcsb.org (accessed March 29, 2017).
112. Paulova, M.; Ticha, M.; Entlicher, G.; Kostir, J.; Kocourek, J., Relationship between red blood cell agglutination and polysaccharide precipitation by phytohemagglutinin of Pisum sativum L. *FEBS Lett.* **1970,** *9* (6), 345-347.
113. Rüdiger, H.; Gabius, H.-J., Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconjugate J.* **2001,** *18* (8), 589-613.
114. Pletnev, V.; Ruzheinikov, S.; Tsygannik, I.; Mikhailova, I.; Duax, W.; Ghosh, D.; Pangborn, W., The structure of Pea Lectin-D-glucopyranose complex at a 1.9 resolution. *uss. J. Bioorg. Chem.* **1997,** *23*, 436-445.
115. Ruzheinikov, S.; Mikhailova, I.; Tsygannik, I.; Pangborn, W.; Duax, W.; Pletnev, V., The Structure of the Pea Lectin-D-mannopyranose Complex at a 2.1 Resolution. *Bioorg. Khim.* **1998,** *24*, 277-279.
116. Rini, J. M.; Hardman, K. D.; Einspahr, H.; Suddath, F. L.; Carver, J. P., X-ray crystal structure of a Pea lectin-trimannoside complex at 2.6 A resolution. *J. Biol. Chem.* **1993,** *268* (14), 10126-10132.
117. Kuwabara, N.; Hu, D.; Tateno, H.; Makyio, H.; Hirabayashi, J.; Kato, R., Conformational change of a unique sequence in a fungal galectin from Agrocybe cylindracea controls glycan ligand-binding specificity. *FEBS Lett.* **2013,** *587* (22), 3620-3625.
118. Raz, A.; Nakahara, S., Biological modulation by lectins and their ligands in tumor progression and metastasis. *Anticancer Agents Med. Chem.* **2008,** *8* (1), 22-36.
119. Lannoo, N.; Van Damme, E. J., Nucleocytoplasmic plant lectins. *BBA. General subjects* **2010,** *1800* (2), 190-201.
120. Yoshioka, K.; Sato, Y.; Murakami, T.; Tanaka, M.; Niwa, O., One-step detection of galectins on hybrid monolayer surface with protruding lactoside. *Anal. Chem.* **2010,** *82* (4), 1175-1178.
121. Murakami, T.; Yoshioka, K.; Sato, Y.; Tanaka, M.; Niwa, O.; Yabuki, S., Synthesis and galectin-binding activities of mercaptododecyl glycosides containing a terminal β-galactosyl group. *Biorg. Med. Chem. Lett.* **2011,** *21* (4), 1265-1269.
122. Yang, R.-Y.; Hsu, D. K.; LIu, F.-T., Expression of galectin-3 modulates T-cell growth and apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93* (13), 6737-6742.
123. Liu, F.-T.; Patterson, R. J.; Wang, J. L., Intracellular functions of galectins. *BBA. General subjects* **2002,** *1572* (2), 263-273.
124. Yang, R.-Y.; Rabinovich, G. A.; Liu, F.-T., Galectins: structure, function and therapeutic potential. *Expert Rev. Mol. Med.* **2008,** *10*, e17.
125. Delacour, D.; Koch, A.; Jacob, R., The role of galectins in protein trafficking. *Traffic* **2009,** *10* (10), 1405-1413.
126. Sato, S.; St-Pierre, C.; Bhaumik, P.; Nieminen, J., Galectins in innate immunity: dual functions of host soluble β-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). *Immunol. Rev.* **2009,** *230* (1), 172-187.
127. Elola, M. T.; Ferragut, F.; Cárdenas, D. V.; Nugnes, L. G.; Gentilini, L.; Laderach, D.; Troncoso, M. F.; Compagno, D.; Wolfenstein-Todel, C.; Rabinovich, G.

A., Expression, localization and function of galectin-8, a tandem-repeat lectin, in human tumors. *Histol. Histopathol.* **2014,** *29* (9), 1093-1105.
128. Ochieng, J.; Green, B.; Evans, S.; James, O.; Warfield, P., Modulation of the biological functions of galectin-3 by matrix metalloproteinases. *BBA. General subjects* **1998,** *1379* (1), 97-106.
129. Vasta, G. R., Roles of galectins in infection. *Nat. Rev. Microbiol.* **2009,** *7* (6), 424-438.
130. Tuğçe, Ç., Immunhistochemical expression of galectin-3 in cancer: a review of the literature. *Turk. Patoloji Derg.* **2012,** *28* (1), 1-10.
131. Radosavljevic, G.; Volarevic, V.; Jovanovic, I.; Milovanovic, M.; Pejnovic, N.; Arsenijevic, N.; Hsu, D. K.; Lukic, M. L., The roles of Galectin-3 in autoimmunity and tumor progression. *Immunol. Res.* **2012,** *52* (1-2), 100-110.
132. Fujihara, S.; Mori, H.; Kobara, H.; Rafiq, K.; Niki, T.; Hirashima, M.; Masaki, T., Galectin-9 in cancer therapy. *Recent Pat. Endocr. Metab. Immune Drug Discov.* **2013,** *7* (2), 130-137.
133. Heusschen, R.; Griffioen, A. W.; Thijssen, V. L., Galectin-9 in tumor biology: a jack of multiple trades. *BBA Reviews on cancer* **2013,** *1836* (1), 177-185.
134. Shin, T., The pleiotropic effects of galectin-3 in neuroinflammation: a review. *Acta Histochem.* **2013,** *115* (5), 407-411.
135. Duray, A.; De Maesschalck, T.; Decaestecker, C.; Remmelink, M.; Chantrain, G.; Neiveyans, J.; Horoi, M.; Leroy, X.; Gabius, H.-J.; Saussez, S., Galectin fingerprinting in naso-sinusal diseases. *Oncol. Rep.* **2014,** *32* (1), 23-32.
136. Kopitz, J.; Vértesy, S.; André, S.; Fiedler, S.; Schnölzer, M.; Gabius, H.-J., Human chimera-type galectin-3: defining the critical tail length for high-affinity glycoprotein/cell surface binding and functional competition with galectin-1 in neuroblastoma cell growth regulation. *Biochimie* **2014,** *104*, 90-99.
137. Thijssen, V. L.; Griffioen, A. W., Galectin-1 and-9 in angiogenesis: a sweet couple. *Glycobiology* **2014**, cwu048.
138. Ito, K.; Stannard, K.; Gabutero, E.; Clark, A. M.; Neo, S.-Y.; Onturk, S.; Blanchard, H.; Ralph, S. J., Galectin-1 as a potent target for cancer therapy: role in the tumor microenvironment. *Cancer Metastasis Rev.* **2012,** *31* (3-4), 763-778.
139. St-Pierre, C.; Ouellet, M.; Giguère, D.; Ohtake, R.; Roy, R.; Sato, S.; Tremblay, M. J., Galectin-1-specific inhibitors as a new class of compounds to treat HIV-1 infection. *Antimicrob. Agents Chemother.* **2012,** *56* (1), 154-162.
140. Zucchetti, M.; Bonezzi, K.; Frapolli, R.; Sala, F.; Borsotti, P.; Zangarini, M.; Cvitkovic, E.; Noel, K.; Ubezio, P.; Giavazzi, R., Pharmacokinetics and antineoplastic activity of galectin-1-targeting OTX008 in combination with sunitinib. *Cancer Chemother. Pharmacol.* **2013,** *72* (4), 879-887.
141. Astorgues-Xerri, L.; Riveiro, M. E.; Tijeras-Raballand, A.; Serova, M.; Neuzillet, C.; Albert, S.; Raymond, E.; Faivre, S., Unraveling galectin-1 as a novel therapeutic target for cancer. *Cancer Treat. Rev.* **2014,** *40* (2), 307-319.
142. Funasaka, T.; Raz, A.; Nangia-Makker, P. In *Nuclear transport of galectin-3 and its therapeutic implications*, Seminars in Cancer Biology, Elsevier: 2014; pp 30-38.
143. André, S.; Kaltner, H.; Manning, J. C.; Murphy, P. V.; Gabius, H.-J., Lectins: getting familiar with translators of the sugar code. *Molecules* **2015,** *20* (2), 1788-1823.
144. Ahmad, N.; Gabius, H.-J.; André, S.; Kaltner, H.; Sabesan, S.; Roy, R.; Liu, B.; Macaluso, F.; Brewer, C. F., Galectin-3 precipitates as a pentamer with synthetic

multivalent carbohydrates and forms heterogeneous cross-linked complexes. *J. Biol. Chem.* **2004,** *279* (12), 10841-10847.
145. Thijssen, V. L.; Barkan, B.; Shoji, H.; Aries, I. M.; Mathieu, V.; Deltour, L.; Hackeng, T. M.; Kiss, R.; Kloog, Y.; Poirier, F., Tumor cells secrete galectin-1 to enhance endothelial cell activity. *Cancer Res.* **2010,** *70* (15), 6216-6224.
146. Collins, P. M.; Bum-Erdene, K.; Yu, X.; Blanchard, H., Galectin-3 interactions with glycosphingolipids. *J. Mol. Biol.* **2014,** *426* (7), 1439-1451.
147. Salomonsson, E.; Carlsson, M. C.; Osla, V.; Hendus-Altenburger, R.; Kahl-Knutson, B.; Öberg, C. T.; Sundin, A.; Nilsson, R.; Nordberg-Karlsson, E.; Nilsson, U. J., Mutational tuning of galectin-3 specificity and biological function. *J. Biol. Chem.* **2010,** *285* (45), 35079-35091.
148. Bourne, Y.; Bolgiano, B.; Liao, D.-I.; Strecker, G.; Cantau, P.; Herzberg, O.; Feizi, T.; Cambillau, C., Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. *Nat. Struct. Mol. Biol.* **1994,** *1* (12), 863-870.
149. Lopez-Lucendo, M. F.; Solis, D.; Andre, S.; Hirabayashi, J.; Kasai, K.; Kaltner, H.; Gabius, H.-J.; Romero, A., Growth-regulatory human galectin-1: crystallographic characterisation of the structural changes induced by single-site mutations and their impact on the thermodynamics of ligand binding. *J. Mol. Biol.* **2004,** *343* (4), 957-970.
150. Si, Y.; Feng, S.; Gao, J.; Wang, Y.; Zhang, Z.; Meng, Y.; Zhou, Y.; Tai, G.; Su, J., Human galectin-2 interacts with carbohydrates and peptides non-classically: new insight from X-ray crystallography and hemagglutination. *Acta Biochim. Biophys. Sinica* **2016,** *48* (10), 939-947.
151. Saraboji, K.; Håkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; Logan, D. T., The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* **2012,** *51* (1), 296-306.
152. Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J., Structural and thermodynamic studies on cation-Pi interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc.* **2005,** *127* (6), 1737-1743.
153. Rustiguel, J. K.; Soares, R. O.; Meisburger, S. P.; Davis, K. M.; Malzbender, K. L.; Ando, N.; Dias-Baruffi, M.; Nonato, M. C., Full-length model of the human galectin-4 and insights into dynamics of inter-domain communication. *Sci. Rep.* **2016,** *6*, 33633.
154. Leonidas, D. D.; Vatzaki, E. H.; Vorum, H.; Celis, J. E.; Madsen, P.; Acharya, K. R., Structural basis for the recognition of carbohydrates by human galectin-7. *Biochemistry* **1998,** *37* (40), 13930-13940.
155. Ruiz, F. M.; Scholz, B. A.; Buzamet, E.; Kopitz, J.; Andre, S.; Menendez, M.; Romero, A.; Solis, D.; Gabius, H. J., Natural single amino acid polymorphism (F19Y) in human galectin-8: detection of structural alterations and increased growth-regulatory activity on tumor cells. *The FEBS journal* **2014,** *281* (5), 1446-64.
156. Yoshida, H.; Teraoka, M.; Nishi, N.; Nakakita, S.; Nakamura, T.; Hirashima, M.; Kamitori, S., X-ray structures of human galectin-9 C-terminal domain in complexes with a biantennary oligosaccharide and sialyllactose. *J. Biol. Chem.* **2010,** *285* (47), 36969-36976.
157. Ackerman, S. J.; Liu, L.; Kwatia, M. A.; Savage, M. P.; Leonidas, D. D.; Swaminathan, G. J.; Acharya, K. R., Charcot-Leyden crystal protein (galectin-10) is not

a dual function galectin with lysophospholipase activity but binds a lysophospholipase inhibitor in a novel structural fashion. *J. Biol. Chem.* **2002,** *277* (17), 14859-14868.
158. Kawaguchi, T.; Matsumoto, I.; Osawa, T., Studies on hemagglutinins from Maackia amurensis seeds. *J. Biol. Chem.* **1974,** *249* (9), 2786-2792.
159. Imberty, A.; Gautier, C.; Lescar, J.; Pérez, S.; Wyns, L.; Loris, R., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **2000,** *275* (23), 17541-17548.
160. Yamamoto, K.; Ishida, C.; Saito, M.; Konami, Y.; Osawa, T.; Irimura, T., Cloning and sequence analysis of the Maackia amurensis haemagglutinin cDNA. *Glycoconjugate J.* **1994,** *11* (6), 572-575.
161. Voglmeir, J.; Flitsch, S. L., Glycosyltransferases. *Science of Synthesis, Biocatalysis in Organic Synthesis* **2015,** *1*, 507-542.
162. Gloster, T. M., Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **2014,** *28*, 131-141.
163. Roychoudhury, R.; Pohl, N. L. B., New structures, chemical functions, and inhibitors for glycosyltransferases. *Curr. Opin. Chem. Biol.* **2010,** *14* (2), 168-173.
164. Zhan, Y.-T.; Su, H.-Y.; An, W., Glycosyltransferases and non-alcoholic fatty liver disease. *World J. Gastroenterol.* **2016,** *22* (8), 2483-2493.
165. Chen, X., Fermenting next generation glycosylated therapeutics. *ACS Chem. Biol.* **2011,** *6* (1), 14-17.
166. Wang, S.; Vidal, S., Recent design of glycosyltransferase inhibitors. *Carbohydr. Chem.* **2013,** *39*, 78-101.
167. Merino, P.; Tejero, T.; Delso, I.; Hurtado-Guerrero, R.; Gomez-SanJuan, A.; Sadaba, D., Recent progress on fucosyltransferase inhibitors. *Mini-Rev. Med. Chem.* **2012,** *12*, 1455-1464.
168. Kajimoto, T.; Node, M., Synthesis of glycosyltransferase inhibitors. *Synthesis* **2009,** (19), 3179-3210.
169. Whalen, L. J.; Greenberg, W. A.; Mitchell, M. L.; Wong, C.-H., Iminosugar-based glycosyltransferase inhibitors. *Iminosugars* **2007**, 153-176.
170. Jung, K.-H.; Schmidt, R. R., Glycosyltransferase inhibitors. *Carbohydrate-Based Drug Discovery* **2003,** *2*, 609-659.
171. Qian, X.; Palcic, M. M., Glycosyltransferase inhibitors. *Carbohydrates in Chemistry and Biology* **2000,** *3*, 293-312.
172. Tedaldi, L.; Wagner, G. K., Beyond substrate analogues: new inhibitor chemotypes for glycosyltransferases. *Medchemcomm* **2014,** *5* (8), 1106-1125.
173. Liang, D.-M.; Liu, J.-H.; Wu, H.; Wang, B.-B.; Zhu, H.-J.; Qiao, J.-J., Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.* **2015,** *44* (22), 8350-8374.
174. Rini, J.; Esko, J.; Varki, A., Glycosyltransferases and glycan-processing enzymes. In *Essentials of Glicobiology. 2nd edition*, Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009; p Available from: http://www.ncbi.nlm.nih.gov/books/NBK1921/.
175. Malik, V.; Black Gary, W., Structural, functional, and mutagenesis studies of UDP-glycosyltransferases. *Adv. Protein Chem. Struct. Biol.* **2012,** *87*, 87-115.

176. Lira-Navarrete, E.; de las Rivas, M.; Companon, I.; Pallares, M. C.; Kong, Y.; Iglesias-Fernandez, J.; Bernardes, G. J. L.; Peregrina, J. M.; Rovira, C.; Bernado, P.; Bruscolini, P.; Clausen, H.; Lostao, A.; Corzana, F.; Hurtado-Guerrero, R., Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates proteinO-glycosylation. *Nat. Commun.* **2015,** *6*, 6937.
177. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The carbohydrate-active enzymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **2009,** *37*, D233-D238.
178. Chugh, S.; Gnanapragassam, V. S.; Jain, M.; Rachagani, S.; Ponnusamy, M. P.; Batra, S. K., Pathobiological implications of mucin glycans in cancer: sweet poison and novel targets. *Biochim. Biophys. Acta* **2015,** *1856* (2), 211-225.
179. Martin-Santamaria, S.; Jiménez-Barbero, J.; Gabius, H.-J., Structural studies on the interaction of saccharides and glycomimetics with galectin-1: a 3D perspective using a combined molecular modeling and NMR approach. *Pure Appl. Chem.* **2012,** *84* (1), 49-64.
180. Sapay, N.; Nurisso, A.; Imberty, A., Simulation of carbohydrates, from molecular docking to dynamics in water. *Methods Mol. Biol.* **2013,** *924*, 469-483.
181. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
182. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009,** *5* (9), 2353-2370.
183. Lins, R. D.; Hunenberger, P. H., A new GROMOS force field for hexopyranose-based carbohydrates. *J. Comput. Chem.* **2005,** *26* (13), 1400-1412.
184. Ishida, T., Computational modeling of carbohydrate-recognition process in E-selectin complex: structural mapping of sialyl Lewis X onto ab initio QM/MM free energy surface. *J. Phys. Chem. B* **2010,** *114* (11), 3950-3964.
185. Miller, M. C.; Ribeiro, J. P.; Roldos, V.; Martin-Santamaria, S.; Cañada, F. J.; Nesmelova, I. A.; André, S.; Pang, M.; Klyosov, A. A.; Baum, L. G.; Jiménez-Barbero, J.; Gabius, H.-J.; Mayo, K. H., Structural aspects of binding of alpha-linked digalactosides to human galectin-1. *Glycobiology* **2011,** *21* (12), 1627-1641.
186. Schrödinger, E., An undulatory theory of the mechanics of atoms and molecules. *Physical Review* **1926,** *28* (6), 1049.
187. Fischer, C. F., General hartree-fock program. *Comput. Phys. Commun.* **1987,** *43* (3), 355-365.
188. Hammond, B. L.; Lester, W. A.; Reynolds, P. J., *Monte Carlo methods in ab initio quantum chemistry*. World Scientific: 1994; Vol. 1.
189. Gross, E. K.; Dreizler, R. M., *Density functional theory*. Springer Science & Business Media: 2013; Vol. 337.
190. Rahal-Sekkal, M.; Sekkal, N.; Kleb, D. C.; Bleckmann, P., Structures and energies of D-galactose and galabiose conformers as calculated by ab initio and semiempirical methods. *J. Comput. Chem.* **2003,** *24* (7), 806-818.
191. Fanfrlik, J.; Bronowska, A. K.; Rezac, J.; Prenosil, O.; Konvalinka, J.; Hobza, P., A reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 protease with 22 ligands. *J. Phys. Chem. B* **2010,** *114* (39), 12666-12678.

192. Appell, M.; Strati, G.; Willett, J. L.; Momany, F. A., B3LYP/6-311++G** study of alpha- and beta-D-glucopyranose and 1,5-anhydro-D-glucitol: 4C1 and 1C4 chairs, (3,O)B and B(3,O) boats, and skew-boat conformations. *Carbohydr. Res.* **2004,** *339* (3), 537-551.
193. Nowacki, A.; Liberek, B., Methyl 3-amino-2,3,6,-trideoxy-L-hexopyranosides in DFT level theory conformational studies. *J. Phys. Chem. A* **2008,** *112* (30), 7072-7079.
194. Nowacki, A.; Walczak, D.; Liberek, B., Fully acetylated 1,5-anhydro-2-deoxypent-1-enitols and 1,5-anhydro-2,6-dideoxyhex-1-enitols in DFT level theory conformational studies. *Carbohydr. Res.* **2012,** *352*, 177-185.
195. Tvaroska, I.; Taravel, F. R.; Utille, J. P.; Carver, J. P., Quantum mechanical and NMR spectroscopy studies on the conformations of the hydroxymethyl and methoxymethyl groups in aldohexosides. *Carbohydr. Res.* **2002,** *337* (4), 353-367.
196. Momany, F. A.; Schnupf, U., DFTMD studies of beta-cellobiose: conformational preference using implicit solvent. *Carbohydr. Res.* **2011,** *346* (5), 619-630.
197. Levine, I. N., Quantum chemistry. 5. th. New York, Prentice Hall: 1991.
198. Parr, R. G., Density functional theory of atoms and molecules. In *Horizons of Quantum Chemistry*, Springer: 1980; pp 5-15.
199. Koshland, D. E., The key–lock theory and the induced fit theory. *Angew. Chem. Int. Ed.* **1995,** *33* (23-24), 2375-2378.
200. Karush, F., Heterogeneity of the binding sites of bovine serum albumin1. *J. Am. Chem. Soc.* **1950,** *72* (6), 2705-2713.
201. Weber, G., Ligand binding and internal equilibria in proteins. *Biochemistry* **1972,** *11* (5), 864-878.
202. Grant, O. C.; Woods, R. J., Recent advances in employing molecular modelling to determine the specificity of glycan-binding proteins. *Curr. Opin. Struct. Biol.* **2014,** *28c*, 47-55.
203. Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J., Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput. Aided Mol. Des.* **1996,** *10* (4), 293-304.
204. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-1749.
205. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004,** *47* (7), 1750-1759.
206. Lexa, K. W.; Carlson, H. A., Protein flexibility in docking and surface mapping. *Q. Rev. Biophys.* **2012,** *45* (3), 301-343.
207. Michalewicz, Z., Evolutionary programming and genetic programming. In *Genetic algorithms+ data structures= evolution programs*, Springer: 1996; pp 283-287.
208. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998,** *19* (14), 1639-1662.
209. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015,** *10* (5), 449-461.

210. Böhm, H.-J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994,** *8* (3), 243-256.
211. Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997,** *11* (5), 425-445.
212. Verkhivker, G.; Appelt, K.; Freer, S.; Villafranca, J., Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **1995,** *8* (7), 677-691.
213. Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000,** *295* (2), 337-356.
214. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
215. Trott, O.; Olson, A. J., AutoDock VINA: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
216. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P., A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984,** *106* (3), 765-784.
217. Goodford, P. J., A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985,** *28* (7), 849-857.
218. Stouten, P. F.; Frömmel, C.; Nakamura, H.; Sander, C., An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.* **1993,** *10* (2-6), 97-120.
219. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** *49* (21), 6177-6196.
220. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L., Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001,** *105* (28), 6474-6487.
221. Baxter, J., Local optima avoidance in depot location. *J. Oper. Res. Soc.* **1981,** *32* (9), 815-819.
222. Andrea, R.; Blesa, M.; Blum, C.; Michael, S., Hybrid metaheuristics-an emerging approach to optimization. Springer: 2008.
223. Patel, S.; Brooks, C. L., CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004,** *25* (1), 1-16.
224. Patel, S.; Mackerell, A. D.; Brooks, C. L., CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004,** *25* (12), 1504-1514.

225. Ponder, J. W.; Case, D. A., Force fields for protein simulations. *Adv. Protein Chem.* **2003,** *66*, 27-85.
226. MacKerell, A. D.; Feig, M.; Brooks, C. L., Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004,** *25* (11), 1400-1415.
227. Dunker, A. K.; Silman, I.; Uversky, V. N.; Sussman, J. L., Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **2008,** *18* (6), 756-764.
228. Bourhis, J. M.; Canard, B.; Longhi, S., Predicting protein disorder and induced folding: from theoretical principles to practical applications. *Current Protein and Peptide Science* **2007,** *8* (2), 135-149.
229. Bartlett, A. I.; Radford, S. E., An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* **2009,** *16* (6), 582-588.
230. Eliezer, D., Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009,** *19* (1), 23-30.
231. Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T.; Joannopoulos, J., Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **1992,** *64* (4), 1045.
232. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
233. *D.A. Case, T.A. Darden, 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. Götz, 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, AMBER 12, University of California, San Francisco*, 2012.
234. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
235. Wang, X.; Woods, R. J., Insights into furanose solution conformations: beyond the two-state model. *J. Biomol. NMR* **2016,** *64* (4), 291-305.
236. DeMarco, M. L., Molecular dynamics simulations of membrane- and protein-bound glycolipids using GLYCAM. *Methods Mol. Biol.* **2015,** *1273*, 379-390.
237. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011,** *51* (1), 69-82.
238. Pérez-Regidor, L.; Zarioh, M.; Ortega, L.; Martín-Santamaría, S., Virtual screening approaches towards the discovery of Toll-like receptor modulators. *Int. J. Mol. Sci.* **2016,** *17* (9), 1508.
239. Haga, J. H.; Ichikawa, K.; Date, S., Virtual screening techniques and current computational infrastructures. *Curr. Pharm. Des.* **2016,** *22* (23), 3576-3584.
240. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S., A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for

ligands and proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** *47* (2), 279-294.
241. Universal Protein Resource (UNIPROT). http://www.uniprot.org (accessed March 12, 2017).
242. Basic Local Alignment Search Tool (BLAST). https://blast.ncbi.nlm.nih.gov.
243. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (17), 3389-3402.
244. Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T., Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **2009,** *4* (1), 1-13.
245. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E., The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protocols* **2015,** *10* (6), 845-858.
246. Bennett-Lovsey, R. M.; Herbert, A. D.; Sternberg, M. J.; Kelley, L. A., Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: structure, function, and Bioinformatics* **2008,** *70* (3), 611-625.
247. SAVES. https://services.mbi.ucla.edu/SAVES.

CHAPTER 2

MOLECULAR RECOGNITION STUDIES OF ^{19}F GLYCOMIMETICS TARGETING PEA LECTIN. A COMBINED NMR & COMPUTATIONAL APPROXIMATION

2.1. Introduction

Molecular recognition processes are the pillars of biochemical communication. Subtle differences in chemical substitutions, geometry and structure in the molecular partners guarantee the effectiveness and specificity of the resulting response. Of all natural biomolecules, glycans are unequalled in their capacity for biological coding and signal translation into cellular messages.[1-3] This power derives from the tremendous structural variability of glycans with respect to saccharide sequence, chemical modification, conformation and shape.[4] For instance, branched and linear glycans present their saccharide moieties in different spatial orientations, translating into distinct biomolecular messages and effects even for identical saccharide composition.[5] It is therefore critical to thoroughly analyse the structural details of glycan recognition, and NMR spectroscopy/molecular modelling are key techniques for this difficult task.

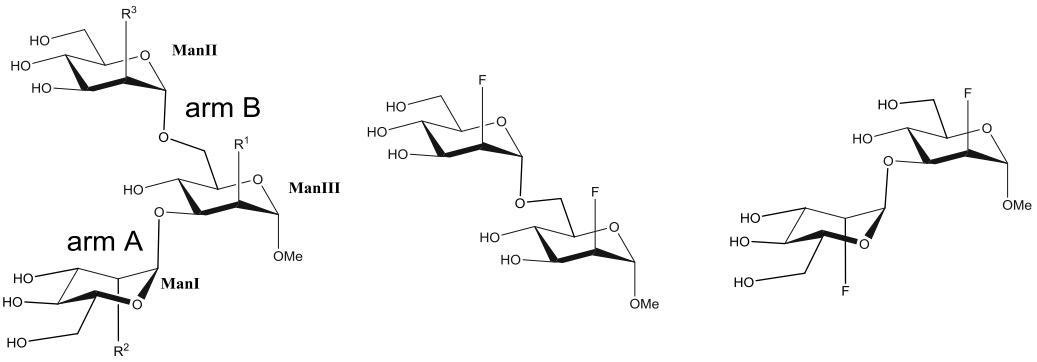
Nevertheless, the distinction of homologous monosaccharidic units in complex oligo- and poly-saccharides remains a challenge by conventional NMR methods, given the low dispersion and high overlap of the ^{1}H signals. Introduction of non-endogenous NMR active heteronuclei may greatly help to resolve such problems, and fluorine seems to be a good suited candidate.[6] Moreover, fluorine atom has been extensively used in medicinal chemistry for bioisosteric replacement of H and OH, so similar biological properties to the parent compound should be expected.

Thus, deoxy-fluoroglycans have widely been used to detect molecular interactions and deduce the enthalpic contribution of specific sugar hydroxyl groups to receptor binding.[7-8] From the NMR point of view, the sensitivity of the ^{19}F nucleus is almost comparable to that of ^{1}H, while its chemical shift dispersion is significantly larger. Thus, ^{19}F is extremely sensitive to variations in the chemical environment. ^{19}F NMR spectrum therefore fairly generally allows for immediate distinction of the pertaining saccharide moiety, in stark contrast to the often highly degenerate ^{1}H spectrum with extensive signal overlap and multiplet splitting. All ^{19}F NMR parameters (e.g., chemical shift, linewidth, J-couplings and relaxation times) are particularly useful in providing structural and dynamic information complementary to that provided by ^{1}H and ^{13}C NMR spectroscopy.[9] For these reasons, fluorinated glycomimetics have received great interest,[10] and have allowed to extend the use of ^{19}F NMR methods also to the fields of medicinal chemistry and chemical biology.[11]

This work aims to combine the use of fluorinated glycomimetics with novel ^{19}F-based STD experiments (performed by Jiménez-Barbero's group) and computational tools, to demonstrate the usefulness of such approach to assess the binding mode of complex glycans to lectins. As proof-of-principle, we have employed the trifluorinated glycomimetic of the ubiquitous trimannoside present in the inner core of *N*-glycans (Chart 2.1) and a model lectin. The plant lectin from pea, *Pisum sativum* (PSA), serves as a common model for studying glycan detection and epitope recognition in the quest for biomarkers. Reported X-ray and molecular modelling studies disagreed in providing a clear picture of the trimannoside epitope and the ligand binding mode into PSA.[12-13]

Substitution of the hydroxyl group on C2 in each of the three mannose (Man) moieties by fluorine has allowed us to characterise the otherwise rather elusive glycan binding epitope by robust ^{19}F NMR experiments, assisted by molecular modelling methods. This trisaccharide has been chosen for its intrinsic complexity from the NMR perspective (three similar Man units). We have compared the molecular recognition features of the natural trimannoside (**1**) with those of its trifluorinated analogue (**2**) and its two fluorinated dimannoside fragments (**3**, **4**).

The experimental ^{19}F and ^{1}H NMR data have been complemented with computational methods to provide a 3D model for the interaction with the PSA, and to confirm, indeed, the coexistence of multiple binding modes.



1 $R^1 = R^2 = R^3 = OH$
2 $R^1 = R^2 = R^3 = F$

3

4

*Chart 2.1. Structure of methyl 3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside (**1**), its disaccharide components, and its fluorine containing glycomimetics (**2**-**4**). The non-reducing terminal residues are labelled as arm A (1→3 linkage, ManI) and B (1→6 linkage, ManII). The reducing (branched) residues in the trisaccharides are dubbed ManIII.*

2.2. Results

Previously reported X-ray and molecular modelling studies on the PSA/**1** system could not provide definitive answers on the ligand binding mode and sugar epitope.[12] In fact, possibly due to the high flexibility of the glycan, X-ray diffraction data (PDB ID 1RIN, Figure 2.1, left)[12] could not identify the binding mode and showed only one mannose residue inside the lectin's recognition site, with no electron density observable for the other two sugar units. This result could be explained by the existence of an exchange process between alternative binding modes, differing in the particular Man unit bound at the recognition site. Another plausible explanation would be a single binding mode involving the 1→6 linkage (arm B), where its intrinsic flexibility would lead to conformational averaging and a blurring of the electron density for the central and terminal 1→3 linked Man residues. In fact, early molecular modelling studies suggested this (1→6)-mode flexible interaction with the lectin,[13] but more experimental data is needed to further clarify the binding mode.

Previous studies have also shown that the dissociation constant of trimannoside **1** from PSA is moderate, with a K_D = 0.19±0.07 mM, and three times stronger than for the monomer, α-methyl-mannopyranoside.[14] Thus, the two additional Man moieties contribute positively to the trimannoside binding, although their effect is small and the role of enthalpic/entropic compensation is not evident. PSA recognizes glucose (Glc) with virtually the same affinity as Man, which strongly suggests that the hydroxyl group at position C2 of the sugar moiety is not directly involved in key interactions. As mentioned above, the X-ray structure of the PSA/triMan complex (PDB ID 1RIN)[12] showed only one Man residue at the PSA binding site, with all its hydroxyl groups except O2 involved in intermolecular hydrogen bonds with different polar amino acid residues, mainly Asp111, Asn155, Gly159, Glu248, and Ala247. The very same binding site is occupied by Glc from sucrose in the X-ray complex with PSA (PDB ID 1OFS) (Figure 2.1 right). Therefore, 2-deoxy-fluoro substitution within all Man residues should not alter the binding features substantially (compound **2** in Chart 2.1**)**.

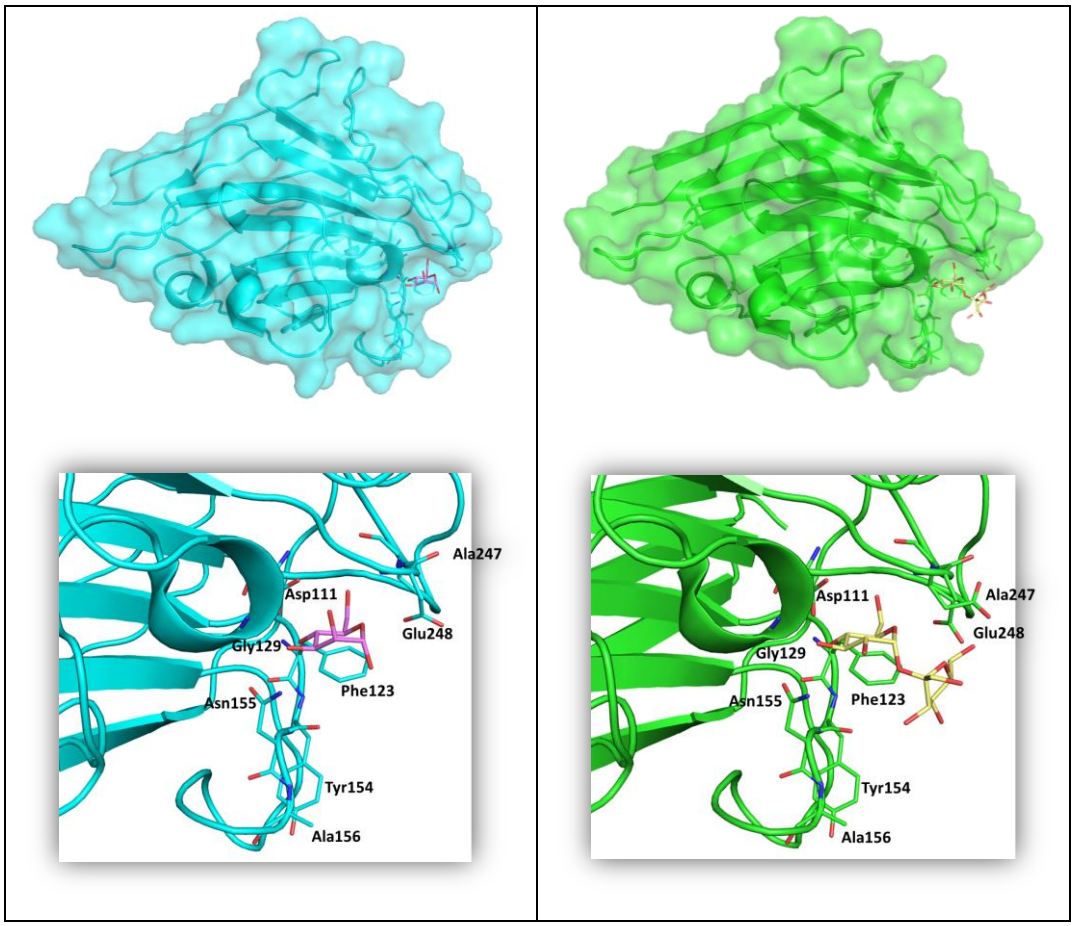


Figure 2.1. Left: Pea lectin (cyan, PDB ID 1RIN) in complex with a mannoside moiety (magenta). Right: Pea lectin (green, PDB ID 1OFS) in complex with sucrose (yellow). The main residues involved in the binding are shown in sticks.

In principle, the sugar moiety recognized by the lectin could correspond to any of the three Man residues and we therefore included both alternative fluorinated dimannosides into the NMR and computational studies, Manα1→6Man and Manα1→3Man (compound **3** and **4,** Chart 2.1). The binding epitopes were then analysed by using 1D STDreF experiments previously described by the Jiménez-Barbero's group and novel 2D and pseudo 3D extensions of the concept.[15]

2.2.1 The free state: conformational analysis by MM3 calculations and ab initio calculation, Molecular Dynamics simulations, and comparison to the NMR experimental data*

The conformation of the natural analogue **1** and related glycomimetics, including *C*-glycosyl compounds, have been extensively analysed.[16] The NMR data obtained for the trifluorinated analogue **2** were complemented by molecular mechanics and quantum mechanics (DFT) calculations. The use of both, the standard molecular mechanics (MM) protocol with the MM3* force field[17] included in Macromodel,[18] and the DFT calculations at the B3LYP level, allowed to predict a conformational distribution for **2** fairly similar to that found for **1.**[16] Six conformers that especially differ around the ψ and ω angles of the Manα1→6Man linkage and the ψ angle of the Manα1→3Man linkage have been identified. Then, the angles and distances deduced from the NMR experiments (analysis of the $^{3}J_{HH}$ coupling constants, $^{1}H,^{1}H$-NOESY and $^{1}H,^{19}F$-HOESY, for more information see Figure 2.A1-2.A4 and Table 2.A1-2.A2 in Annex II) were compared to the corresponding geometry parameters deduced from the theoretical calculations. Previous studies on the trimannoside **1** led to the conclusion that the conformational behaviour of this flexible saccharide is well described by a model in which six major conformations contribute.[19] The two highly populated structures are those in which the Manα1→6Man linkage is defined as *exo-syn*ϕ/*anti*ψ in back-folded *gg* conformer. Similarly for compound **2**, the NMR experiments show that the ϕ and ψ angles of the Manα1→3Man linkage correspond to a concerted population of two states resulting in two major conformations: (ϕ, ψ) = (-60, -60) and (-40, 20), **2a** and **2b** respectively (Figure 2.2). MM and DFT calculations suggest the predominance of the conformer **2b** (Figure 2.3).

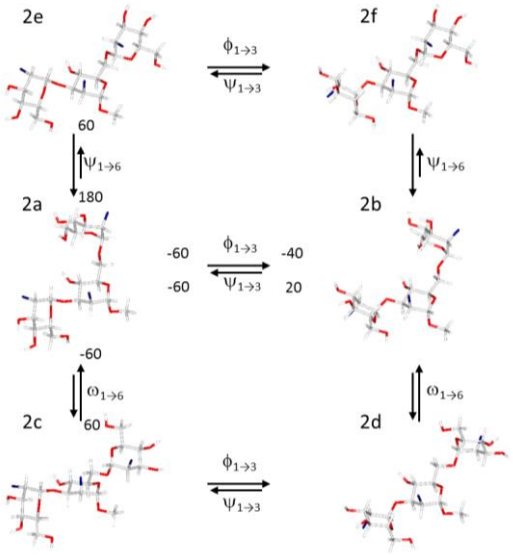


*Figure 2.2. Structural model for the trimannoside analogue **2** in solution. The various transitions correspond to those described for natural compound **1** and analogues.[16] The suffixes **a-f** design the six discussed conformers. Conformers **2a** and **2b** are the major ones in solution, followed by **2c** and **2d**. The contribution of **2e** and **2f** is minor, although still significant.*

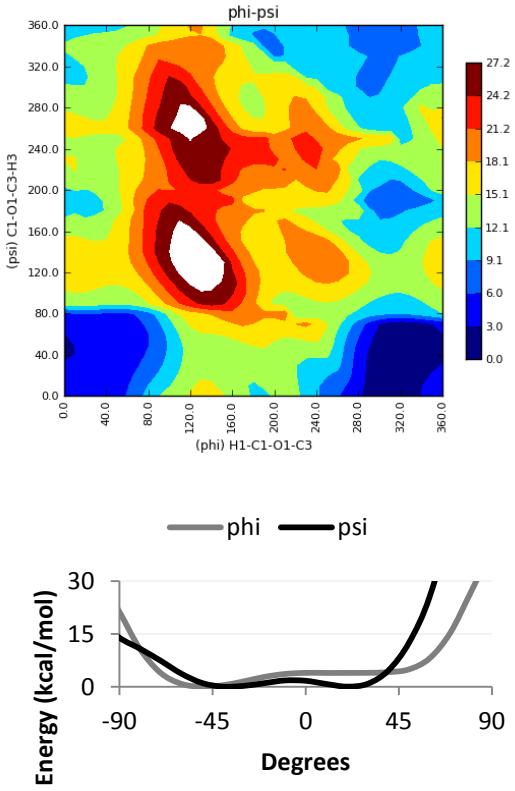


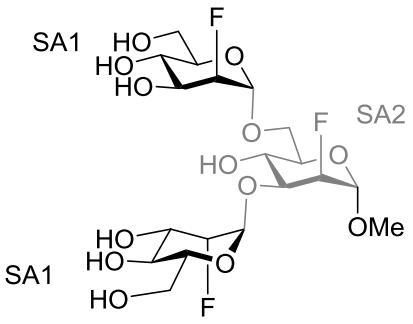
Figure 2.3. Top: adiabatic curve obtained by molecular mechanics calculation (MM3). Legend of relative energy scale is shown in kcal/mol. The blue areas are the most favourable regions for ϕ and ψ and in red the less one. Bottom: energy profile torsion of ϕ and ψ obtained by DFT calculations (B3LYP/3-21G). Only values of ϕ and ψ between -90 and 90 degrees are shown.*

The two conformers with ω torsion angle at the Manα1→6Man linkage in the extended *gt* conformation, resulting in roughly linear molecules (**2c, 2d**), are less populated. Finally, the two minor conformers **2e** and **2f** differ from the major conformers **2a** and **2b** in the dihedral angles ψ about the Manα1→6Man linkage. Thus, the Manα1→6Man linkage is defined as *exo-syn*ϕ/*syn*+ψ and results in a slightly bent structure. Our study is based on the comparison between the structural analysis reported for the natural analogue **1**[19] and those obtained for the mimic **2**. Then, the comparison of the angles and distances deduced from the NMR analysis with the corresponding geometry parameters deduced from computational calculations permitted to have an estimation of the actual population distribution of the different conformers.

The contribution of the two major geometries was also estimated through the analysis of interatomic distances using the calculated structures (Figure 2.2) compared to the experimental ones. A good agreement between the experimentally derived interatomic distances and those calculated for an average population distribution was reached. Molecular dynamics simulations (20 ns) were also performed to assess the conformational stability of the different conformers.

For the bonds, angles and torsions involving F atom, parameters were taken from gaff force field.[20] Charges were calculated with DFT and adapted for GLYCAM force field[21] (Table 2.1).

*Table 2.1. Partial charges assigned to each atom of compound **2**. SA1 residue is the α-2-F-mannose substituted in position 1 (ManI and ManII). SA2 (ManIII) residue is the α-2F-mannose substituted in positions 1, 3 and 6. The parameters for the OMe group present in position **1** of the central Man were taken from GLYCAM force field.*[21]



Atom Type	SA1	SA2
C1	0,400000	0,400000
H1	0,000000	0,000000
C2	0,287172	0,199808
H2	0,000000	0,000000
C3	0,212473	0,203358
H3	0,000000	0,000000
C4	0,277689	0,256190
H4	0,000000	0,000000
C5	0,175601	0,248887
H5	0,000000	0,000000
C6	0,249863	0,227038
H61	0,000000	0,000000
H62	0,000000	0,000000
O6	-0,738010	-0,395602
H6O	0,473381	
O5	-0,354505	-0,427901
O4	-0,776756	-0,693712
H4O	0,481578	0,425474
O3	-0,706108	-0,381428
H3O	0,466263	
F2	-0,254641	-0,256112

Angles (ϕ, ψ and ω of the Manα1→6Man and the Manα1→3Man linkages) were monitored, leading to the identification of different conformers. Agreement with the previous conformational analysis was detected, thus validating the parameters and charges employed, also useful for the MD simulations of the complexes. According to the MM calculations, all glycosidic torsion angles ϕ are governed by the *exo-anomeric* effect that leads to exclusive population of the *exo-syn* conformer for glycosidic linkages. Different values of the ψ dihedral and ω torsion angles of the Manα1→6Man linkage in extended *gt* and back-folded *gg* conformers were found (Figure 2.4).

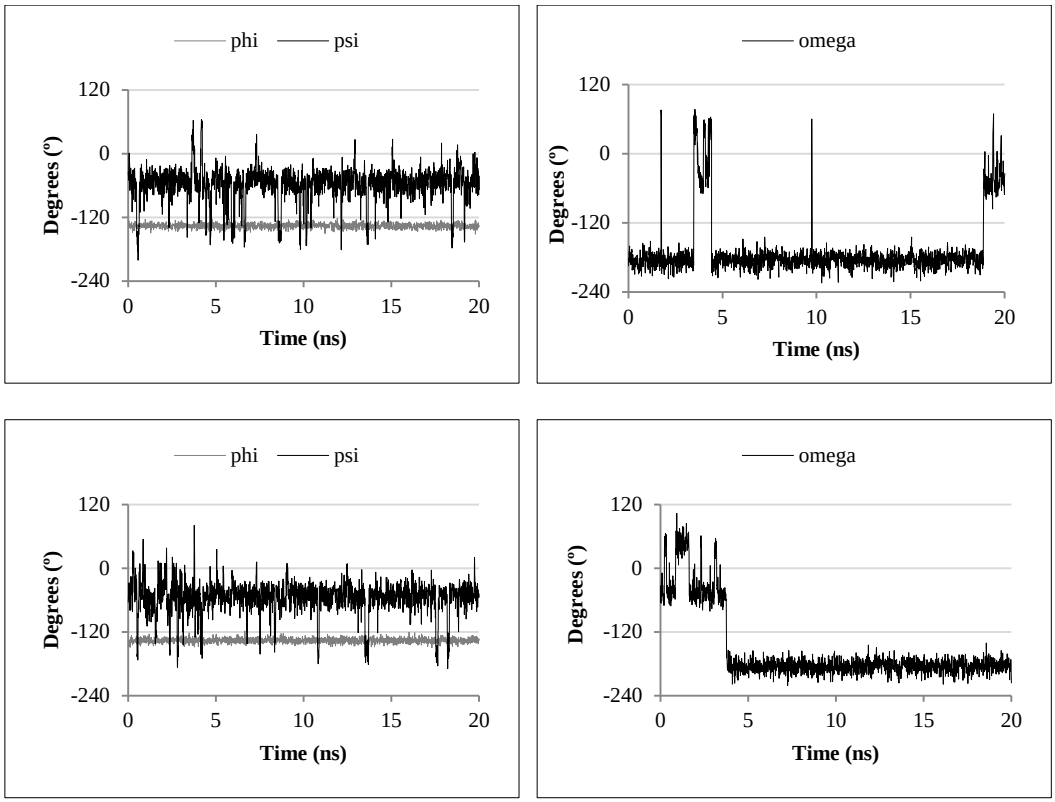


*Figure 2.4. Top: plot obtained from 20 ns of MD simulation of compound **2** in the free state, starting from the gg (ω=~1800) rotamer. The exo-anomeric conformer is unique at the glycosidic linkage, together with certain flexibility at ψ torsion. The gg rotamer is strongly predominant.*
*Bottom: plot obtained from 20 ns of MD simulation of compound **2** in the free state, starting from the gt (ω=~-60) rotamer. Trajectories of the ϕ, ψ and ω torsion angles at the Mana1→6Man linkage are shown. After a few ns, the gg rotamer becomes strongly predominant.*

Therefore, we can conclude that the substitution of the hydroxyl groups at position 2 with fluorine atoms not only does not influence the ring conformation but also has no consequences on the conformational behaviour around the glycosidic linkages. Fittingly, these conformational analysis and MD simulations permitted to conclude that the ensemble average of conformers in the free state for the trifluorinated trimannoside mimic **2** is basically the same than that for the natural trimannoside **1**.

2.2.2 The bound state: integrating NMR results, docking calculations and MD simulations

Interestingly, in the X-ray structure of concanavalin A, another model lectin for the study of glycan binding, in complex with the trimannoside **1** (PDB ID 1CVN), the αMan 1→6 linked provides the major contacts with the lectin recognition site.[22] Analogously, in order to assess the actual Man unit that is effectively bound to the PSA, we then conducted docking calculations of ligands **1-4** into the PSA structure followed by MD simulations of the resulting complexes. The X-ray structure of PSA in complex with trimannoside **1** (PDB ID 1RIN), with only one terminal Man elucidated, was used as 3D geometry of the lectin for the docking calculations. Additional docking calculations were performed with trimannoside **2** with PSA geometry from 1OFS (PSA in complex with sucrose). Fittingly, all the docking models of the PSA/trimannoside complexes discarded the possibility of the reducing central residue Man to be accommodated into the lectin binding site. In fact, there would be important steric clashes between the protein backbone and the α1→3 and α1→6 branches. Our calculations suggested that both A and B arms could be readily accommodated into the binding site (Figure 2.5).

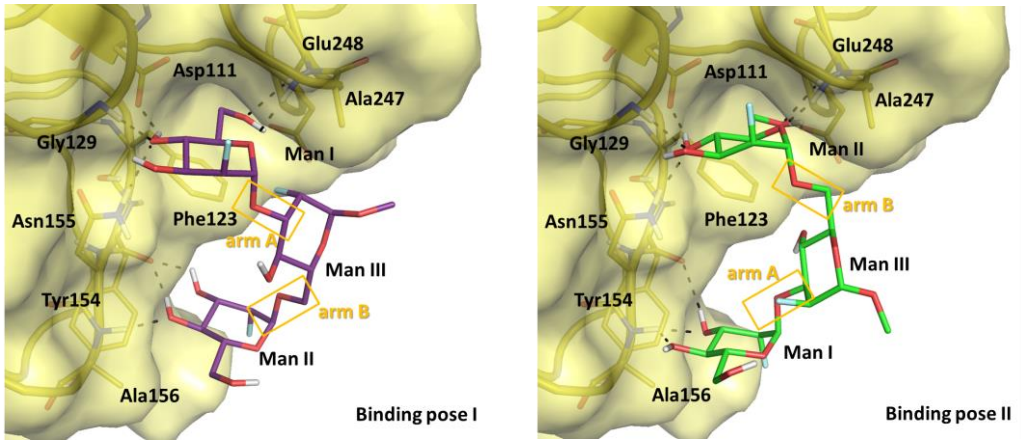


*Figure 2.5. Docked binding poses of compound **2** into the X-ray crystallographic structure of PSA (yellow) from PDB ID 1RIN. Left: binding pose "I": compound **2** (purple) is bound to the PSA through the arm A. Right: binding pose "II": compound **2** (green) is bound to the PSA through the arm B. Only in the case of binding pose I, the fluorine atom on ManIII is close to the PSA.*

Particularly for compound **2**, the docking calculations, with both AutoDock4 and VINA, led to two main predicted binding poses dubbed I and II (Figure 2.5), with only 1 kcal mol^{-1} of difference in the predicted binding energies in favour of the binding pose I. For both cases, we performed the MD simulations of the best docked complexes and analysed the ligand-protein interactions. In particular, we selected the docked poses from AutoDock4. In binding pose I, the arm A (1→3 linkage) is accommodated in the binding site; in the alternative pose II, the arm B (1→6 linkage) provides the major interactions with the lectin binding site. In fact, in both cases, equivalent interactions to those described in the X-ray PSA/Man complex (PDB ID 1RIN) are found for the interacting Man moiety: OH-3 interacts simultaneously with the NH group from Gly159 and the Asp111 carboxylate group, OH-6 interacts with the NH group from Glu248, and O5 establishes a hydrogen bond with the NH group from Ala247. Only in pose I, OH-4 interacts with the NH group from Gly129. No direct interactions are found for OH-2 from the interacting Man. This binding mode is also equivalent to that for Glc in the PSA/sucrose complex (PDB ID 1OFS). For the best predicted binding poses, additional interactions were found between the other terminal Man and the lectin, mainly with the Tyr154 CO group (Figure 2.6). Nevertheless these interactions were lost along the MD simulation, resulting in this terminal Man residue dangling into the solvent.

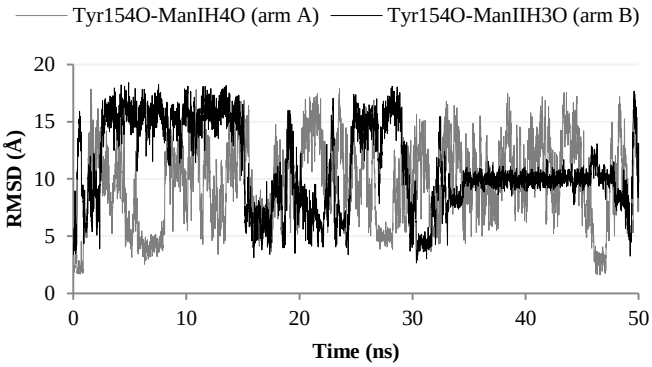


*Figure 2.6. MD simulation (50 ns) for the PSA/**2** complexes starting from the two alternative docked binding poses I (arm A, grey) and II (arm B, black). In both MD simulations, the interaction with Tyr154 is lost almost at the beginning.*

Other docked binding poses did not show this remote interaction and the alternative terminal mannoside (ManII for pose-I-like binding modes and ManI for pose-II-like binding modes) was always exposed towards the solvent with no interactions with the lectin. As for the branched ManIII, this sugar was placed close to the PSA only in the case of binding pose I (Figure 2.4).

Regarding the α1→3 and α1→6 difluorinated dimannosides **4** and **3**, the docking calculations predicted binding poses inside the Man recognition site with basically the same interactions involving the residues participating in the binding of the trimannoside: Gly129, Asn155, Asp111, Glu248, and Ala247. Interestingly, for the α1→6 derivative (compound **3** in Chart 2.1), only the binding pose with the ManII inside the binding site was obtained, while for the α1→3 derivative (compound **4** in Chart 2.1), both binding poses through ManI and through ManIII were predicted (Figure 2.7).

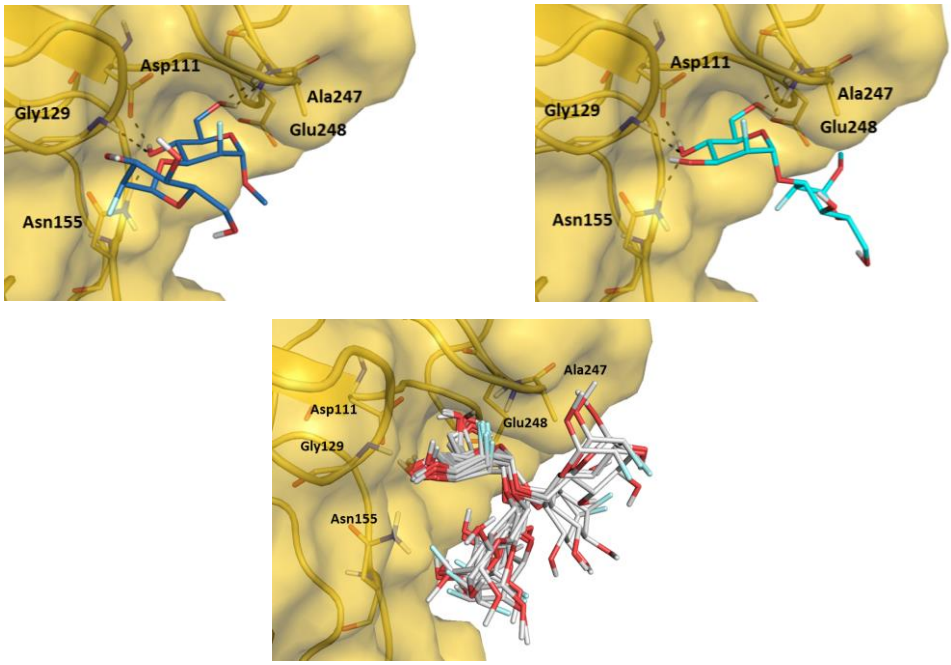


*Figure 2.7. Top: docked binding poses of the α1→3 difluorinated dimannoside (**4**) with PSA (yellow). Two main binding poses were observed: one with the ManIII in the binding site (blue-left) and one with the ManI (cyan-right) in the binding site. Bottom: superimposition of several binding poses (grey) for the α 1→6 derivative (**3**). Docking calculations exclude the possibility to have ManIII inside the binding site. The same ligand/lectin interactions present in the X-ray crystallographic structure of the PSA/Man complex (with residues Gly129, Asn155, Asp111, Glu248, and Ala247, PDB ID 1RIN) were observed in all the docked poses.*

In order to validate experimentally these models, 1D $^{1}H,^{19}F$-STDreF NMR experiments using the complexes of PSA with the α1→3 and α1→6 difluorinated disaccharide mimics **4** and **3** were first acquired. Fittingly, the analysis of the results demonstrated the participation of the reducing sugar unit, ManIII, of the glycan only when the terminal ManI, α1→3 linked is recognized (compound **4**). In contrast, for the alternative analogue **3**, the reducing end does not show any STD and thus, remains solvent exposed (Figure 2.8), in full agreement with the docking predictions.

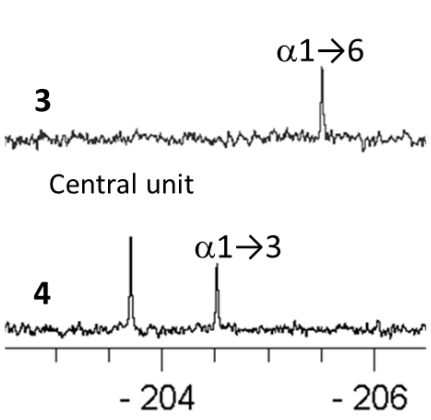


*Figure 2.8. 1D $^{1}H,^{19}F$-STDreF spectra. Top: STDreF spectrum of **3**. Only the α1→6 Man residue shows STD signal, being the reducing moiety solvent exposed. Bottom: STDreF of **4**. Both Man residues take part in the binding event.*

The stability of the two possible PSA/**2** docked complexes (through arm A, pose I, and arm B, pose II) was assessed by MD simulations (Figure 2.9, left), which also provided a detailed picture of the lectin-ligand interactions. Both complexes were maintained stable along the 50 ns of MD simulations, mainly through the interactions with either the 1→3 arm (starting from docked pose I, Figure 2.4, top) or the 1→6 arm (starting from docked pose II, Figure 2.4, bottom).

Thus, from a qualitative perspective, neither of these simulations would point to a preferred binding pose. This suggests difficulties to clearly resolve the conformation of the complete trimannoside by X-ray crystallography, reason why only clear electron density maps for the terminal Man was provided.[12]

In fact, the RMSD of the heavy atoms of the ligand in both complexes (binding pose I and binding pose II) shows a high instability during the MD simulation (Figure 2.9, right and Figure 2.10).

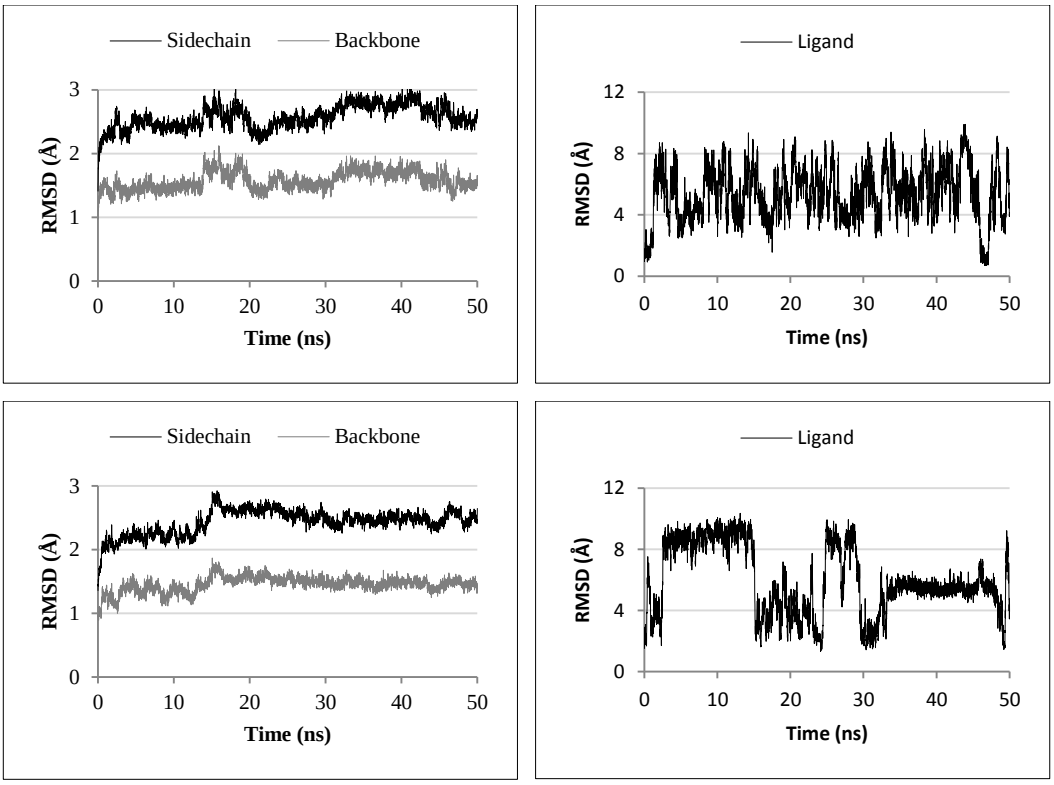


*Figure 2.9. Representative plots from MD simulation of the two complexes PSA (PDB ID 1RIN)/**2** from pose I (top) and pose II (bottom). Left: variation of the RMSD of the protein backbone and side chains found for the 50 ns MD simulation starting from pose I (top) and pose II (bottom) of the PSA/**2** complex. Right: variation of the RMSD of the ligand (heavy atoms) for the 50 ns MD simulation starting from pose I (top) and pose II (bottom) of the PSA/**2** complex.*

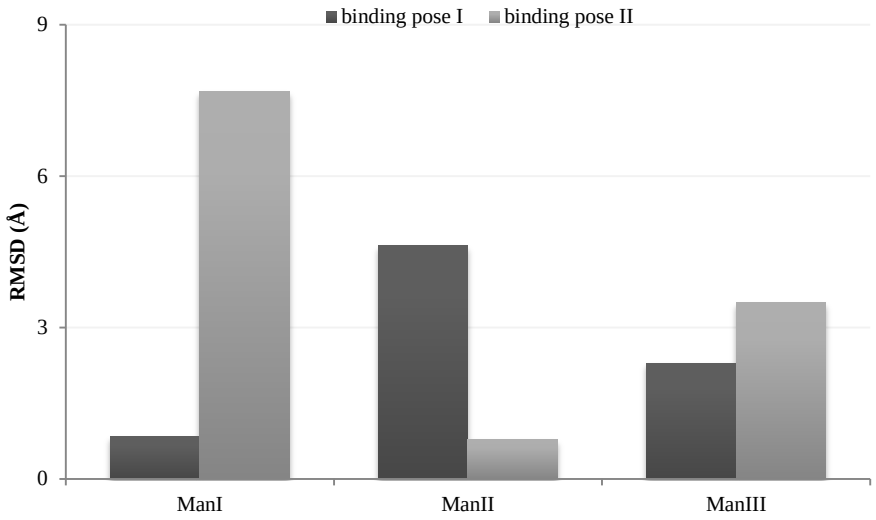


Figure 2.10. RMSD fluctuation of single monomers in both binding poses.

The interactions of the ManIII with the PSA were minor throughout both simulations, although in the docked binding pose I this ManIII was predicted to be closer to PSA. As outlined above in the docking calculations, in the complex through the arm A (pose I), the hydrogens H1, H2, and H4 of the reducing ManIII were close to Phe153 and Ala247 side chains (Figure 2.11).

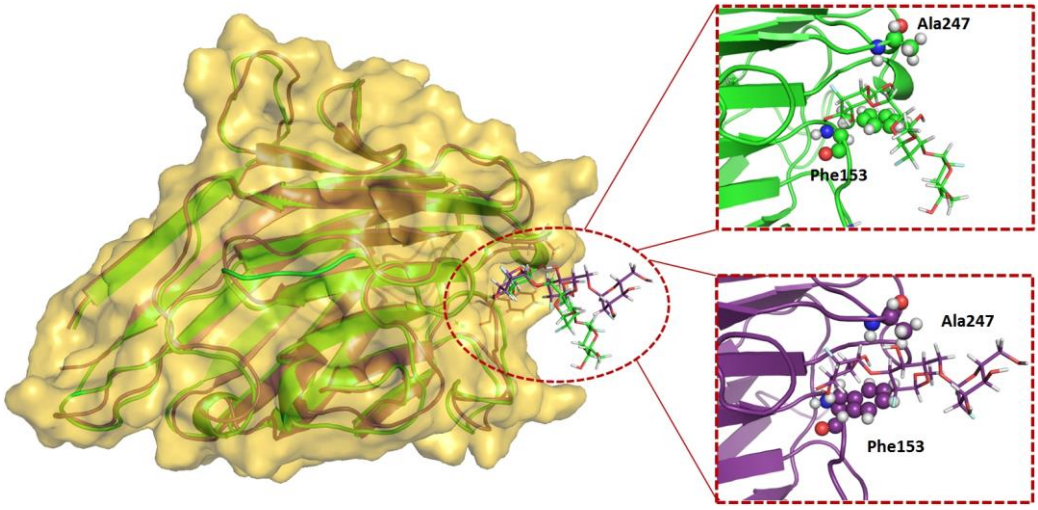


*Figure 2.11. Superimposition of two PSA/**2** complexes from the MD simulations. Top (green): compound **2** is bound through the α1→3 linked ManI (arm A), and hydrogens H1, H2, and H4 of the reducing ManIII are close to Phe153 and Ala247. Bottom (purple): compound **2** is bound through the α1→6 linked ManIII (arm B) and ManIII is more apart from Phe153 and Ala247.*

Interestingly, the analysis of the contributions to the protein/ligand binding energy by means of the MM-PBSA method suggests a slight preference binding pose (Figure 2.12 and Table 2.2). Indeed looking in details the van der Waals, the electrostatic and the polar solvation contribution of the key residues in Figure 2.12 and the energy contribution of each mannose to binding (Table 2.2), a minor preference for the binding pose II is observed. In particular, this preference is dictated by electrostatic interaction between Glu248 and ManII in the binding pose II (Figure 2.13). Since the energy differences are not significant, there results are not conclusive towards a preferred binding pose.

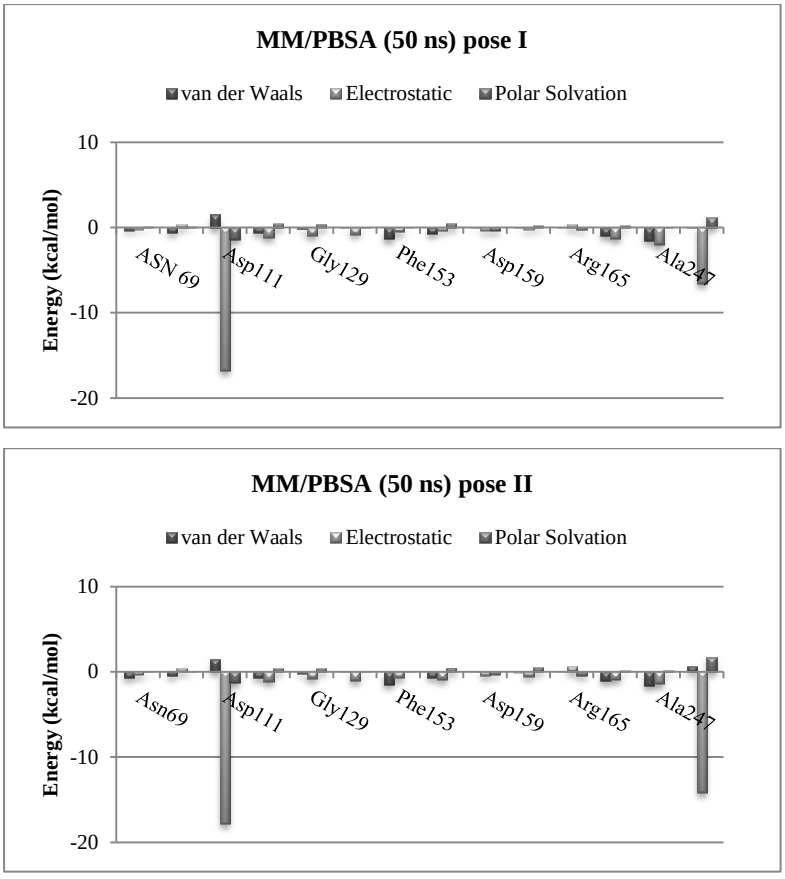


*Figure 2.12. MD simulation (50 ns) for the PSA/**2** complexes starting from the two alternative docked binding poses I (arm A) and II (arm B). The MM-PBSA suggests basically the same aminoacids are involved for the two alternative complexes.*

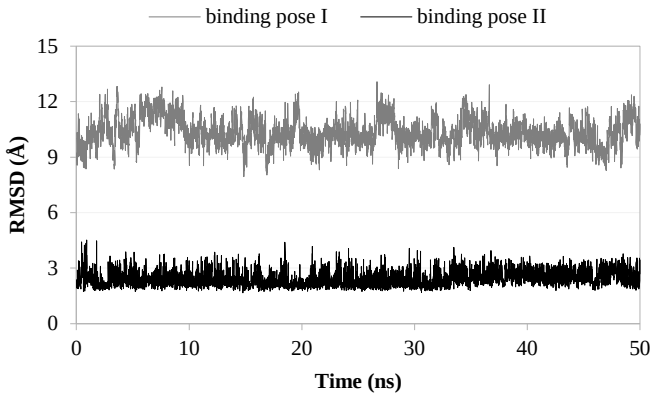


Figure 2.13. Plot showing the variation of the RMSD, during 50 ns of MD simulation, of the distance between the H of the NH group of the Glu248 from PSA and the O6 from ManI in the binding pose I (grey) and the O6 fromManII in binding pose II (black)

Table 2.2. Energetic contribution of each mannose to the binding in the both docked poses

	Contribution of each mannose to the binding energy in the pose I (kcal/mol)	**Standard Deviation**	**Contribution of each mannose to the binding energy in the pose II (kcal/mol)**	**Standard Deviation**
ManI	-10,316	0,0661	-2,947	0.0614
ManII	-0,100	0,0216	-11,932	0.0635
ManIII	-1,163	0,0056	-1,626	0.0094
Total	**-11,579**		**-16,505**	

In a second step, regular ^{1}H-STD-based competition binding experiments were also employed to determine that the fluorinated glycomimetics indeed compete with the natural mannosides for the same binding site[23] and to estimate their relative affinities. The addition of the fluorinated mimics to one sample containing α-methyl-Man in the presence of PSA led in all cases to the decrease in the STD signals of the reference ligand, and to the appearance of STD signals from the glycomimetics (Figure 2.14), thus confirming that a competitive binding process is taking place.

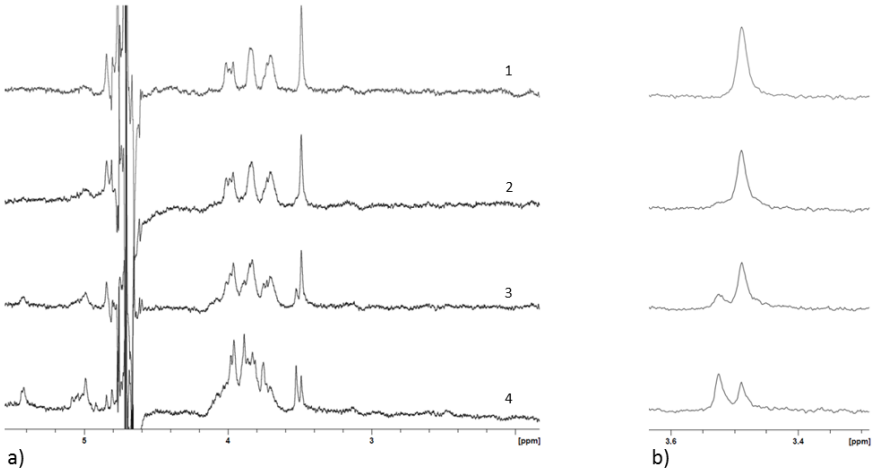


*Figure 2.14 a) ^{1}H-STD competition experiments: PSA (30 μM) in presence of a 0.7 mM concentration of α-methyl-Man and upon the gradual addition of a 0.1 mM (2), 0.4 mM (3) and 1.6mM (4) concentration of compound **2**, as a competitor. b) Expanded regions region are shown: the α-methyl-Man methyl group (3.49 ppm) STD signal intensity is progressively reduced, while the STD signal for the methyl group of compound **2** at 3.52 ppm was observed.*

The obtained binding affinities for the ^{19}F-containing saccharides **2-4** from the STD competition experiments are very similar and only slightly weaker than those reported for trisaccharide **1** (Table 2.3**)**. The binding affinity for **4** is in between those obtained for the trisaccharides **1** and **2** and that reported for the single monosaccharide. However, the estimated K_D for **3** is basically identical to that of the monosaccharide, as also suggested from the STDreF experiment. Only the non-reducing Man moiety contributes to the binding event. Therefore, the binding data are in agreement with the X-ray structure (PDB ID 1RIN) and the MD simulations, and point out that there is one residue providing the major ligand/lectin contacts (K_D for α-methyl-Man is 0.65±0.16 mM), while the additional Man moieties only provide an additional minor stabilization of the complexes (K_D for the disaccharides are ca. 0.41-0.65 mM, while K_D for the trisaccharides are ca. 0.19-0.23 mM).

Table 2.3. Affinity constants determined through ^{1}H-STD NMR experiments

Compounds	***Estimated K_D (mM)***
Methyl α-D-mannopyranoside	0.65 ±0.16[11]
1	0.19 ±0.07[11]
2	0.23 ±0.03
3	0.63 ±0.17
4	0.41 ±0.03

The overlapping of key ^{1}H NMR resonances at the sugar region precluded further analysis of the binding epitope at the residue level by using standard STD analysis. In order to address this problem, Prof. Jiménez-Barbero's group employed novel 2D ^{1}H, ^{19}F STD-TOCSYreF NMR experiments (data not shown). These experiments take advantage from the presence of the ^{19}F nuclei, thus additionally allowing for further specific resolution enhancement comparison with classical experiments such as STD-TOCSY. This information may also be deduced from the already described TOCSYreF experiment,[19, 24-25] although the version by Prof. Jiménez-Barbero's group is improved with some technical modifications regarding the TOCSY mixing sequence. Thus, the experiment allows for the integration of the STD signals at every particular residue and for every proton at one particular sugar ring. Interestingly, the results underline the major participation of the α1→3 linked ManI (arm A) in PSA binding, being the signals from this residue the most intense ones (globally stated as 100%, Table 2.4) in the STD-TOCSYreF spectrum. Nevertheless, the protons belonging to the α1→6 linked ManII (arm B) also show significant STD signals (60% of relative intensity, Table 2.4) demonstrating, beyond all doubts, the simultaneous existence of a second binding mode,[13] for which the mannose moiety α1→6 linked is at the recognition site of Pea lectin.

According to the STD data, the contribution of this binding mode II is smaller than that of the alternative pose I, although still significant. The *O*-methylated ManIII at the reducing end of the glycan also shows STD signals, although at lower extent (20% of relative intensity, Table 2.4).

Table 2.4. Quantitative and relative estimation of the STD signals for every fluoromannose unit. The individual (per proton) and global (per residue) STD intensities are given.

Relative STD Intensity			
	ManI	ManII	ManIII
H_1	37 %	17 %	n.d*
H_2	100 %	60 %	n.d*
H_3	37 %	31 %	25 %
H_4	35 %	48 %	20 %
H_5	43 %	-	-
H_{6R}	-	-	-
H_{6S}	-	-	-
Residue	*100 %*	*62 %*	*20 %*

**Due to the overlapping between H1 and H2 signal it is not possible to estimate quantitatively the individual STD contribution arising from those protons.*

This observation also agrees with that described above for the disaccharide **4**. The docking and MD simulations analysis from the PSA/trifluoromannoside **2** discarded the possibility of direct interaction of the reducing Man moiety at the binding site, but pointed at the proximity of the ManIII hydrogens H1, H2, and H4 with Phe153 and Ala247 side chains. Therefore, in this particular case, it is highly plausible that the interaction of ManI at the major binding site also allows for minor contacts of the reducing ManIII with the lectin, as pointed out in the MD simulations. Therefore, the combination of the modelling and NMR experimental data indicates the co-existence of two different binding modes for the trifluorinated trimannoside when interacting with the plant lectin from *Pisum sativum*. This example accounts for the complementarity between these techniques for the elucidation of complex molecular recognition problems where experimental techniques cannot provide a full response. Moreover, we have provided a novel example of the usefulness of the ^{19}F NMR applied to the recognition of fluorinated-glycans, technique with a many potential applications in the near future.

2.3. Material and Methods

The ring numbering system used in this work is standard for sugar systems, from C1 to C6. The arm A corresponds to the α1→3 linked Man also dubbed ManI. The arm B corresponds to the α1→6 linked Man also dubbed ManII. The reducing (branched) Man is dubbed ManIII. Glycosidic torsion angles are defined as: $\phi = H1_{(i)}\text{-}C1_{(i)}\text{-}O1_{(i)}\text{-}Cn$, $\psi = C1_{(i)}\text{-}O1_{(i)}\text{-}Cn\text{-}Hn$, $\omega = O5_{(i)}\text{-}C5_{(i)}\text{-}C6_{(i)}\text{-}O6_{(i)}$, where *i* indicates a given residue and *n* a ring position.

Ligand preparation. Ligands **1-4** were generated starting from no-fluorinated analogues obtained by the carbohydrate builder from GLYCAM.[21] For the compound **2-4**, the OH-2 groups of each Man residue were replaced by fluorine atoms. Optimized geometries and charge calculations of the ligands **1-4** were obtained by using the MM3* force field within Macromodel.[18] Also sucrose (αGlc1-2βFru) and α-mannose, used for the validation of the docking protocol (see below).

Conformational analysis, DFT calculations and MD simulations of the ligand at the free state. For the glycosidic linkages 1→3 and 1→6 a coordinate scan was performed with MM3* and adiabatic curves were obtained (e.g. Figure 2.3, top). The energy profile torsion (e.g. Figure 2.3, bottom) for the glycosidic linkages were obtain by coordinate scan with DFT calculation with Gaussian03.[26] B3LYP/3-21G was used as basis set. MD simulations were performed using AMBER12 suite of programs.[27] Charges for the fluorinated moieties were calculated using Gaussian03[26] at the B3LYP level of theory and the 6-31G++ basis set. RESP charges were derived with Antechamber, using the RESP fitting method. GLYCAM06 was used as force field. Each ligand was solvated within a cubic periodic water box of TIP3P waters at a minimum distance of 10 Å and a MD simulation of 20 ns was run. The behaviour of the ligand was analysed with the cpptraj module, included in the AmberTools13 package.[27] The torsion angles were monitored along the MD trajectory.

Protein preparation. The X-ray structures of PSA in complex with trimannoside (PDB ID 1RIN, resolution of 2.6 Å) and sucrose (PDB ID 1OFS, resolution of 1.80 Å) were used. Water molecules were removed. Hydrogens were added with Epik[28] at physiological pH.

The manganese and calcium ions were kept. Histidine protonation was established as HIE with the exception of His166 which was assigned as HIP because of the proximity to Glu150 carboxylate side chain and Asp165 CO group. The protein structure was minimized with OPLS 2005 with implicit solvent (water). For the docking calculations, Gasteiger charges were assigned by AutoDockTools.[29] Missing residues Leu213-Glu214-Glu215-Glu216-Asn217, far away from the binding site, were added with PyMOL.[30] Finally, MD simulations were performed during 20 ns using the AMBER12 package and the Amberff10 force field, and the stability of the protein was corroborated by RMSD monitorization.

Validation of the docking methodology. The docking protocols with AutoDock4 and VINA were followed for the prediction of the binding of sucrose and mannose towards the Pea lectin (from PDB IDs 1OFS and 1RIN, respectively). The X-ray crystallographic pose was predicted in both cases with both programs AutoDock4 and VINA.

Docking calculations. Docking calculations were performed with AutoDock4 and VINA.[31] With AutoDock4, the Lamarckian genetic algorithm was used to sample different conformations of the ligands, by randomly changing all the torsion angles and overall orientation of the molecule. A three-dimensional grid was defined centred on an equidistant point to the side chains of key binding residues: Phe153, Tyr154 and Gly216. The grid spacing was 0.375 Å (50 x 48 x 68 Å), and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock were also used. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

MD simulations of the complexes. The two best docked solutions for ligand **2** (binding pose I and binding pose II) were considered as starting geometries to perform the MD simulations. Five Na^+ counterions were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. GLYCAM06, gaff, and Amberff10 were used as force fields.

The MD simulations were carried out by using the sander module in the GPU accelerated version of AMBER12, using CUDA programming language and NVIDIA graphics cards. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 50 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 5000 structures for each complex, which were finally analysed. The residue contribution to the free binding energy along the simulations was calculated by means of the MM-PBSA method.

Bibliography

1. Solís, D.; Bovin, N. V.; Davis, A. P.; Jiménez-Barbero, J.; Romero, A.; Roy, R.; Smetana, K.; Gabius, H.-J., A guide into glycosciences: how chemistry, biochemistry and biology cooperate to crack the sugar code. *BBA. General subjects* **2015,** *1850* (1), 186-235.
2. Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **2011,** *36* (6), 298-313.
3. Gringhuis, S. I.; Kaptein, T. M.; Wevers, B. A.; Mesman, A. W.; Geijtenbeek, T. B., Fucose-specific DC-SIGN signalling directs T helper cell type-2 responses via IKKε-and CYLD-dependent Bcl3 activation. *Nat. Commun.* **2014,** *5*, 3898.
4. Gabius, H.-J.; Kayser, K., Introduction to glycopathology: the concept, the tools and the perspectives. *Diagn. Pathol.* **2014,** *9* (1), 4.
5. Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009,** *5* (10), 1087-1104.
6. Gupta, D.; Dam, T. K.; Oscarson, S.; Brewer, C. F., Thermodynamics of lectin-carbohydrate interactions. Binding of the core trimannoside of asparagine-linked carbohydrates and deoxy analogs to concanavalin A. *J. Biol. Chem.* **1997,** *272* (10), 6388-6392.
7. Solis, D.; Jiménez-Barbero, J.; Martín-Lomas, M.; Díaz-Mauriño, T., Probing hydrogen-bonding interactions of bovine heart galectin-1 and methyl β-lactoside by use of engineered ligands. *FEBS J.* **1994,** *223* (1), 107-114.
8. Matei, E.; André, S.; Glinschert, A.; Infantino, A. S.; Oscarson, S.; Gabius, H.-J.; Gronenborn, A. M., Fluorinated carbohydrates as lectin ligands: Dissecting glycan–cyanovirin interactions by using 19F NMR spectroscopy. *Chem. Eur. J.* **2013,** *19* (17), 5364-5374.
9. Vulpetti, A.; Hommel, U.; Landrum, G.; Lewis, R.; Dalvit, C., Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. *J. Am. Chem. Soc.* **2009,** *131* (36), 12949-12959.
10. Leclerc, E.; Pannecoucke, X.; Ethève-Quelquejeu, M.; Sollogoub, M., Fluoro-C-glycosides and fluoro-carbasugars, hydrolytically stable and synthetically challenging glycomimetics. *Chem. Soc. Rev.* **2013,** *42* (10), 4270-4283.
11. Ojima, I., Use of fluorine in the medicinal chemistry and chemical biology of bioactive compounds-A case study on fluorinated taxane anticancer agents. *ChemBioChem* **2004,** *5* (5), 628-635.
12. Rini, J. M.; Hardman, K.; Einspahr, H.; Suddath, F.; Carver, J. P., X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 A resolution. *J. Biol. Chem.* **1993,** *268* (14), 10126-10132.
13. Cheong, Y.; Shim, G.; Kang, D.; Kim, Y., Carbohydrate binding specificity of pea lectin studied by NMR spectroscopy and molecular dynamics simulations. *J. Mol. Struct.* **1999,** *475* (2), 219-232.
14. Stubbs, M. E.; Carver, J. P.; Dunn, R., Production of Pea lectin in Escherichia coli. *J. Biol. Chem.* **1986,** *261* (14), 6141-6144.
15. Diercks, T.; Ribeiro, J. P.; Cañada, F. J.; Andre, S.; Jimenez-Barbero, J.; Gabius, H.-J., Fluorinated carbohydrates as lectin ligands: Versatile sensors in

19F-Detected Saturation Transfer Difference NMR spectroscopy. *Chem. Eur. J.* **2009,** *15*, 5666-5668.
16. Mikkelsen, L. M.; Hernáiz, M. J.; Martín-Pastor, M.; Skrydstrup, T.; Jiménez-Barbero, J., Conformation of glycomimetics in the free and protein-bound state: Structural and binding features of the C-glycosyl analogue of the core trisaccharide α-d-Man-(1→ 3)-[α-d-Man-(1→ 6)]-d-Man. *J. Am. Chem. Soc.* **2002,** *124* (50), 14940-14951.
17. Allinger, N. L.; Yuh, Y. H.; Lii, J. H., Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989,** *111* (23), 8551-8566.
18. *MacroModel, version 10.7* Schrödinger Release 2015-1; Schrödinger, LLC: New York, NY, 2015.
19. Sayers, E. W.; Prestegard, J. H., Solution conformations of a trimannoside from nuclear magnetic resonance and molecular dynamics simulations. *Biophys. J.* **2000,** *79* (6), 3313-3329.
20. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
21. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
22. Naismith, J. H.; Field, R. A., Structural basis of trimannoside recognition by concanavalin A. *J. Biol. Chem.* **1996,** *271* (2), 972-976.
23. Ribeiro, J. P.; Diercks, T.; Jiménez-Barbero, J.; André, S.; Gabius, H.-J., Fluorinated carbohydrates as lectin ligands: 19F-based direct STD monitoring for detection of anomeric selectivity. *Biomolecules* **2015,** *5* (4), 3177-3192.
24. Canales, A.; Mallagaray, A.; Pérez-Castells, J.; Boos, I.; Unverzagt, C.; André, S.; Gabius, H.-J.; Cañada, F. J.; Jiménez-Barbero, J., Breaking pseudo-symmetry in multiantennary complex N-glycans using lanthanide-binding Tags and NMR pseudo-contact shifts. *Angew. Chem.* **2013,** *125* (51), 14034-14038.
25. Nishima, W.; Miyashita, N.; Yamaguchi, Y.; Sugita, Y.; Re, S., Effect of bisecting GlcNAc and core fucosylation on conformational properties of biantennary complex-type N-glycans in solution. *J. Phys. Chem. B* **2012,** *116* (29), 8504-8512.
26. *Gaussian03, Revision E.01*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, Gaussian, Inc.: Wallingford CT, 2004.

27. *AMBER 12*, D.A. Case, T.A. Darden, 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. Götz, 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.
28. *Schrödinger Release 2015-1: Epik, version 3.1, Schrödinger, LLC, New York, NY*, 2015.
29. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
30. Choi, J. K.; Lee, B. H.; Chae, C. H.; Shin, W., Computer modeling of the rhamnogalacturonase-"hairy" pectin complex. *Proteins* **2004,** *55* (1), 22-33.
31. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.

Annex II

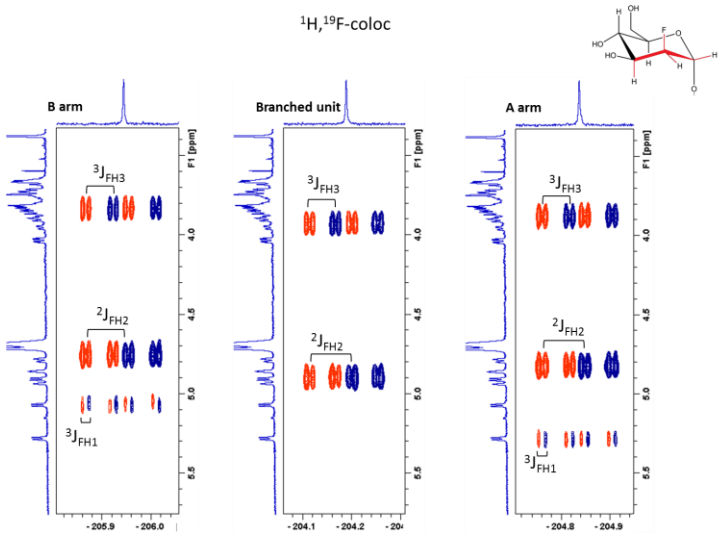


*Figure 2.A1. Strips of 2D ^{1}H,^{19}F-COLOC NMR spectrum of **2** at the corresponding frequency of the ^{19}F atom at every mannose unit. This experiment allows the analysis for the $^{x}J_{FH}$ coupling constants of the protons directly coupled to the ^{19}F atom at two or three bonds. In this way, the ring conformation could be easily assessed.*

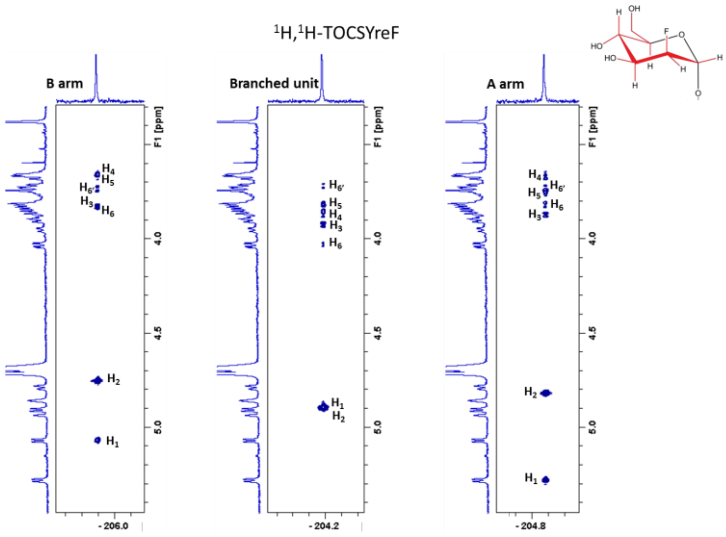


*Figure 2.A2. Strips of 2D ^{1}H,^{19}F TOCSYreF experiment (80 ms mixing time) of **2** at the corresponding frequency of the ^{19}F atom at every mannose unit. The spin systems for each one of the monosaccharidic units are now determined in a non-ambiguous manner.*

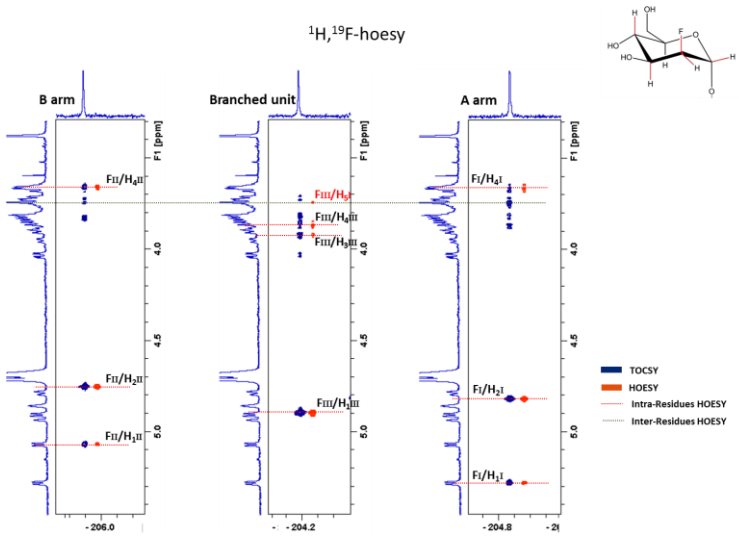


*Figure 2.A3. Strips of 2D ^{1}H,^{19}F HOESY (400 ms mixing time) of molecule **2** for each fluorine atom at the proper mannose unit (orange). Intra-residue ^{1}H-^{19}F nOe cross peaks are highlighted with a red line; while inter-residue nOe cross peak is labeled with a green line. No other inter-residue heteronuclear nOe are observed above the noise level. In blue, ^{1}H, ^{19}F-TOCSY spectrum (80 ms mixing time) as chemical shift reference.*

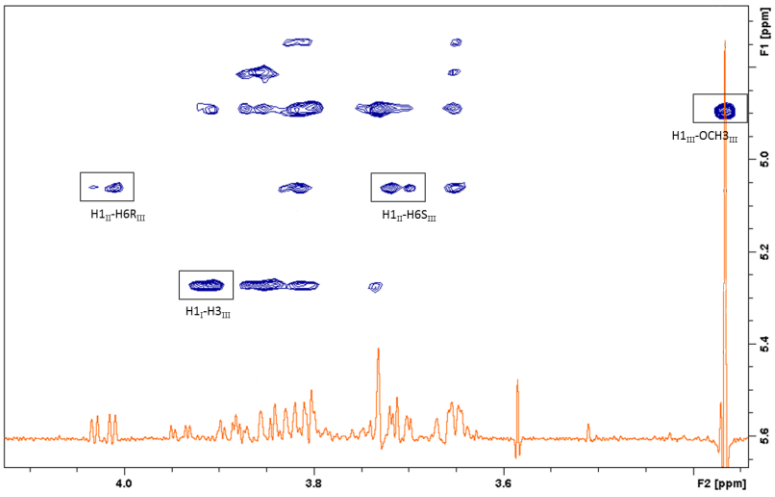


*Figure 2.A4. Strip of 2D NOESY spectra (600 ms mixing time) for **2** taken at H1 frequencies (axis y) of the three Man moieties. Key inter-residual NOEs are highlighted.*

*Table 2.A1. Relevant inter-atomic distances (Å) for the key conformers of mimic **2** from molecular mechanics calculations. Relative energies are reported. The descriptors in Arabic numeral refer to the particular mannose sugar ring. The comparison with the essential experimental NOE contacts with conformational information is also provided. The relevant intra and inter residue and ensemble average $<r^{-6}>^{-1/6}$ proton-proton and proton-fluorine distances (Å) were estimated from the integration of the observed NOEs cross peaks using the ISPA approximation.*

Compound **2**

Atom Pair		Calculated distance for each conformer (Å) and steric energy values [kcal·mol^{-1}]						Observed NOE Intensity	Experimental distance (Å) (±5%)	Average derived conformer*
		2a [+5.7]	**2b** [+0.2]	**2c** [+5.5]	**2d** [0.0]	**2e** [+18.8]	**2f** [+13.2]			
H1 III	OCH_3	2.4	2.4	2.4	2.4	2.4	2.4	Very strong	2.4	2.4
H3 III	H1 I	3.1	2.1	3.1	2.1	3.1	2.1	strong	2.6	2.6
H4 III	H1 I	2.7	4.3	2.7	4.3	2.7	4.3	medium	3.0	3.5
$H6_R$ III	H1 II	3.0	3.0	3.0	3.0	3.6	3.6	medium	3.0	3.1
$H6_S$ III	H1 II	2.4	2.4	2.4	2.4	3.0	3.0	strong	2.6	2.5
H2 II	H3 II	2.5	2.5	2.5	2.5	2.5	2.5	Very strong	2.5	2.5
H2 III	H5 I	3.3	2.6	3.3	2.6	3.3	2.6	Medium strong	2.8	2.9
F III	H4 III	2.7	2.7	2.7	2.7	2.7	2.7	strong	2.6	2.7
F III	H5 I	4.6	2.7	4.6	2.7	4.6	2.7	Medium weak	3.4	3.6

*Table 2.A2. Chemical shift assignments (ppm) and $^3J_{HH}$ values (Hz) of the trimannoside analogue **2**.*

	1→3 arm (ManI)		1→6 arm (ManII)		Reducing residue (ManIII)	
Position	^{1}H	^{19}F	^{1}H	^{19}F	^{1}H	^{19}F
1	5.27 (2.0)	(7.6)	5.06 (2.0)	(7.3)	4.89 (2.0)	(7.4)
2	4.81 (2.5)	-204.2 (49.0)	4.74 (2.3)	-206.0 (49.0)	4.88 (2.3)	-204.2 (49.0)
3	3.87 (9.3)	(32.0)	3.82 (9.0)	(32.0)	3.91 (9.3)	(30.3)
4	3.66 (10.0)		3.65 (‡)		3.86 (9.6)	
5	3.75 (‡;‡)		3.67 (3.5;‡)		3.81 (3.6; 2.0)	
6R	3.81 (11.3)		3.80 (12.7)		4.02 (11.4)	
6S	3.72		3.73		3.70	
OCH_3					3.36	

Bracketed values in the ^1H and ^{19}F columns are $^3J_{HH}$ and $^xJ_{FH}$ values, respectively. ‡ J coupling could not be quantified due to spectral overlap.

CHAPTER 3

STRUCTURAL INSIGHTS ON THE RECOGNITION OF THE ALPHA-TYPE HISTO-BLOOD ALPHA-GAL EPITOPE BY GALECTIN-3

3.1. Introduction

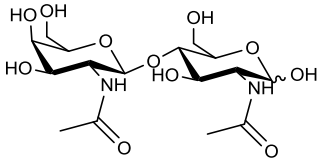
Molecular recognition events are essential for life processes. Glycan-protein interactions are the key for many diverse beneficial and pathological events.[1] Among glycan receptors, galectins are specific galactose binding lectins that have been reported to participate in a myriad of biological events of different nature, mainly through the specific recognition of galactose contained in glycans on cell surfaces, in a crosstalk process between glycan and lectin containing cells. Therefore, in the last years the interest in these proteins as therapeutic targets has increased since the development of selective galectins inhibitors could improve the prognosis of many diseases.

Recently, their role in innate immunity, through the recognition of glycans on pathogens, has been proposed.[2-4] The recent demonstration that galectins 4 and 8 are able to kill *E. coli* bacteria expressing blood group B antigen[5] has reinforced this idea, and raised the question whether immuno lectins would have a role in protection against pathogens expressing self-like antigens. The presence of blood group epitopes on erythrocytes is intriguing and it has been related with evolutionary processes arising from the pressure to react against pathogen producing blood group type molecules for attachments to cells and subsequent invasion.[6] Other glycans very similar to self antigens, have been proposed as molecular patterns for the recognition of parasites by the immune system, a process in which galectins would be involved.

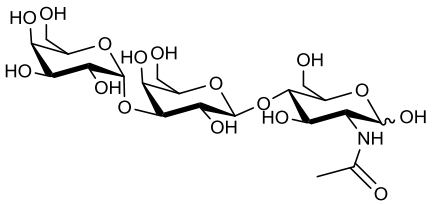
In particular LacdiNAc (compound **1** in Scheme 3.1, βGalNAc1-4GlcNAc), very similar to lactose (Lac, βGal1-4Glc), is present in vertebrates as constituents of *N*-glycans, is highly abundant in helminth parasites, and it has been proposed as a parasite pattern for galectin-3 (gal-3) mediated immune response.[7]

Even though gal-3 has been reported before to indeed recognize A-type histo-blood group type II tetrasaccharide[8] and LacdiNAc.[7] LacdiNAc-glycans constitute a parasite pattern for galectin-3-mediated immune recognition.[7] With higher affinity than that for lactose,[9] the structural basis of such interactions is not available yet. Herein we describe the atomic details of the interaction of gal-3 with a variety of glycans of different size, from the simple LacdiNAc disaccharide (**1**) to the blood group A type II tetrasaccharide (compound **4** in Scheme 3.1, GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAc), passing by the α-Gal trisaccharide (compound **2** in Scheme 3.1, Galα1-3Galβ1-4GlcNAc) and α-Gal tetrasaccharide (compound **3** in Scheme 3.1, Galα1-3Galβ1-4GlcNAcβ1-3Gal). Compounds **2** and **3** contain the α-Gal epitope (Galα1-3Galβ1-4GlcNAc), which is a major xenoantigen expressed on glycoconjugates of non-primate mammals. In turn, apes and humans produce a natural antibody (anti-Gal), which specifically binds α-Gal epitope. Anti-Gal is produced as the most abundant antibody (1% of immunoglobulins) in all individuals, and represents an important barrier for xenotransplantation.[10] However, this antibody is not the only responsible for the rejection of pig xenograft organs in humans, since lectins have also been shown to be involved in that process. Specifically, gal-3 has been identified as the receptor for monocyte-dependent recognition of porcine endothelium via binding to α-Gal epitope, playing an important role in delayed xenograft rejection.[11] The work has been performed through a combination of experimental NMR methods, both from the ligand and the galectin's perspective, and molecular modelling protocols.

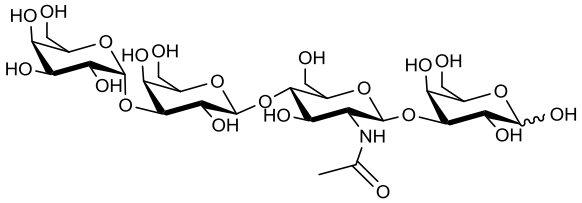
Using this combined approach, we here report the ligand-receptor interactions accounting for the gal-3 recognition of the important xenoantigen α-Gal epitope, showing the important role of gal-3 in the initial recognition and adhesion of human monocytes to porcine aortic endothelial cells which may contribute to delayed xenograft rejection.[10]



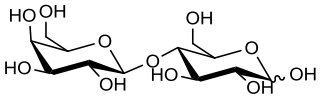
1. LacdiNAc
(GalNAcβ1-4GlcNAc)



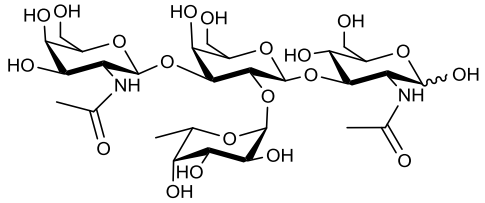
2. α-Gal trisaccharide
(Galα1-3Galβ1-4GlcNAc)



3. α-Gal tetrasaccharide
(Galα1-3Galβ1-4GlcNAcβ1-3Gal)



Lactose, Lac
(Galβ1-4Glc)



4. blood group A type II tetrasaccharide
(GalNAcα1-3(αFuc1-2)Galβ1-4GlcNAcβ)

Scheme 3.1. Chemical structure of the studied ligands.

3.2. Results

3.2.1. The recognition of LacdiNAc by gal-3

Docking calculations were undertaken to predict the binding pose of compounds **1**-**4** in galectin-3. Three different docking protocols (AutoDock4, VINA, and Glide) were used in order to check whether the results were in agreement, and avoid any bias due to the scoring function. Overall, the docking results from the three programs were in very close agreement, predicting low energy poses similar to that for lactose in the X-ray structure (PDB ID 3ZSJ). For MD simulations, we took into consideration the best docked solution energy from AutoDock4.

Docking calculations showed the best pose of compound **1** (LacdiNAc) bound to gal-3 to be very similar to that of lactose (Figure 3.1). Indeed, all the sugar-lectin intermolecular interactions are analogous to those described for lactose binding.[12] In particular, the CH-π interaction between the GalNAcβ and Trp181 is present, as well as the Gal-specific hydrogen bond between the OH4 of the GalNAcβ with His158. Other intermolecular hydrogen bonds are also present involving Arg162, Asn174, Glu184, and Arg186 of the lectin and several polar groups of the disaccharide. This docking-based structure was then submitted to 50 ns MD simulation to assess the stability of the complex.

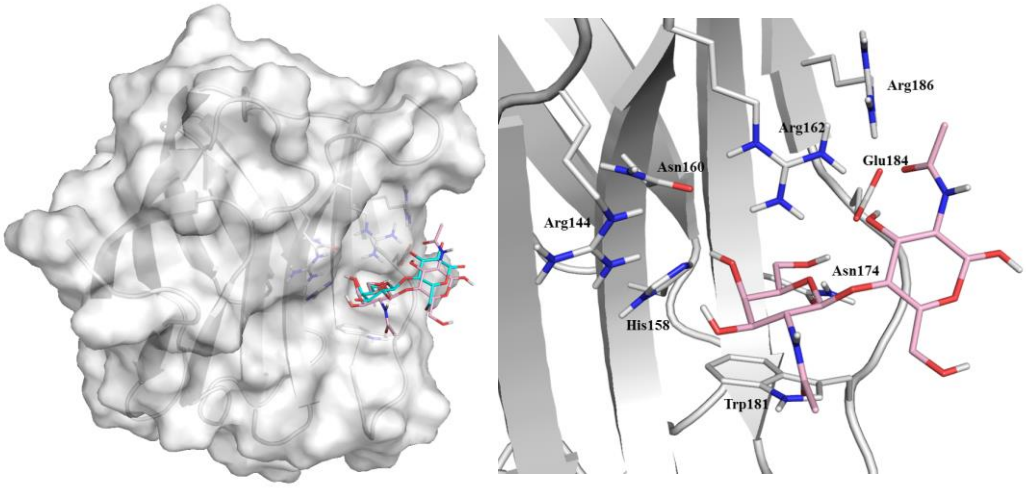


*Figure 3.1. Left: Docked binding pose of LacdiNAc (compound **1**, represented in pink sticks) in complex with gal-3 (X-ray crystal structure PDB ID 3ZSJ, grey surface) superimposed with crystallographic lactose binding pose (cyan sticks). Right: atomic details of the interactions established by LacdiNAc (in pink sticks) and the carbohydrate recognition domain (CRD). A net of hydrogen bonds is formed by the hydroxyl groups of the LacdiNAc and the side chains of Arg186, Arg162, His158, Arg144 and Asn160. Moreover the GalNAc moiety establishes a stack interaction with Trp181.*

The RMSD values for the heavy atoms of the ligand and the protein backbone and side chain were monitored along the 50 ns of MD simulation. Variations smaller than 2 Å were registered, showing that the complex was stable along the MD simulation time (Figure 3.2).

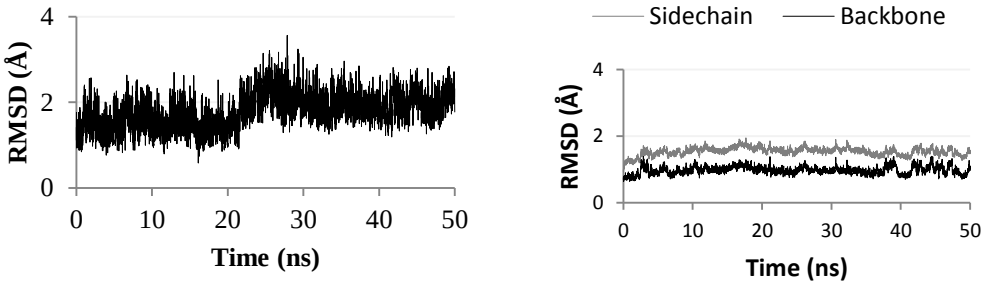


*Figure 3.2. Left: Computed RMSD for heavy atoms of the ligand **1** in the MD simulation of the complexes of gal-3 (PDB ID 3ZSJ) with LacdiNAc. Right: Computed RMSD for the lectin backbone (black) and the side chain (grey) in the MD simulation of the complexes of gal-3 with LacdiNAc (**1**).*

The predicted 3D structure was in complete agreement with the NMR experimental data. From the ligand point of view, the bound conformation is analogous to that in the free state, with protons H1GalNAc-H4GlcNAc at a distance of 2.5 Å, in agreement with the trNOESY experiment (Figure 3.A1 in Annex III). The STD data (Figure 3.3) show in fact that the GalNAc residue of **1** is in close contact with the protein, in agreement with the docked structure, with both acetyl groups exposed to the solvent. In particular H4 and H6 of GalNAc are involved in CH-π interaction with Trp181. Clear chemical shift perturbations are observed for the interacting residues in the HSQC (Figure 3.4) recorded for the gal-3/**1** complex, including the lateral side chain signals for Arg186 and Arg162, which are directly involved in intermolecular hydrogen bonds.

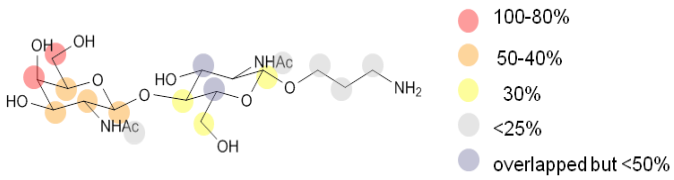


*Figure 3.3. STD for gal-3/**1** at a 1/90 molar ratio. Both acetyl groups received similar saturation transfer from the protein that is below 20% (STD relative). The protons of the linker showed no significant STD effect.*

Moreover a correlation between HSQC data and energetic analysis contribution was performed. In particular, the region of the arginine showed a clear perturbation of their signal (Figure 3.4) in agreement with the energy contribution of these residues in the MM-PBSA analysis (Figure 3.5, top). As in HSQC analysis, a comparison between energetic values of the lactose/gal-3 complex and LacdiNAc/gal-3 complex was analysed. Globally, slight differences in terms of energy were detected. In particular, LacdiNAc/gal-3 complex shows a higher electrostatic contribution of the Arg162 and Arg186 than in the lactose-gal-3 complex, in agreement with HSQC data (Figure 3.4-right). From the MM-PBSA analysis is possible to note as the switch from Gal moiety in the lactose to the GalNAc moieties in LacdiNAc seems to be more favourable for the stacking interaction with Trp181.

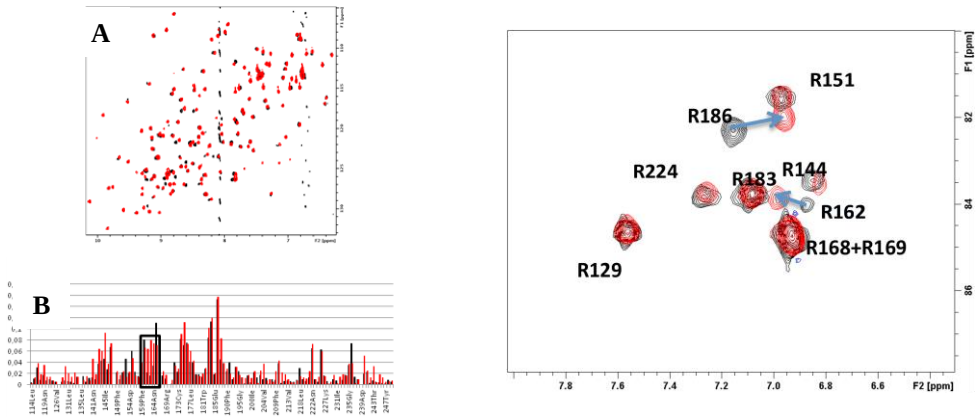


*Figure 3.4. The interaction of **1** with gal-3. Left: **A**. Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 CRD (black) and in the presence of 20 equivalents of **1** (red). **B.** Comparison between chemical shift differences for **1** (red) and lactose (black), highlighting a group of residues differentially perturbed by **1** (Arg162-Phe163-Asn164). Right: Plot of the arginine region in the ^{1}H-^{15}N HSQC spectra of gal-3 (in black) upon addition of **1** (red). Arg186 and Arg162 show significant chemical shift variations. Arg144 shifts very little. In contrast, Arg151, Arg129, Arg168, Arg169, and Arg183 do not show any perturbation upon ligand binding.*

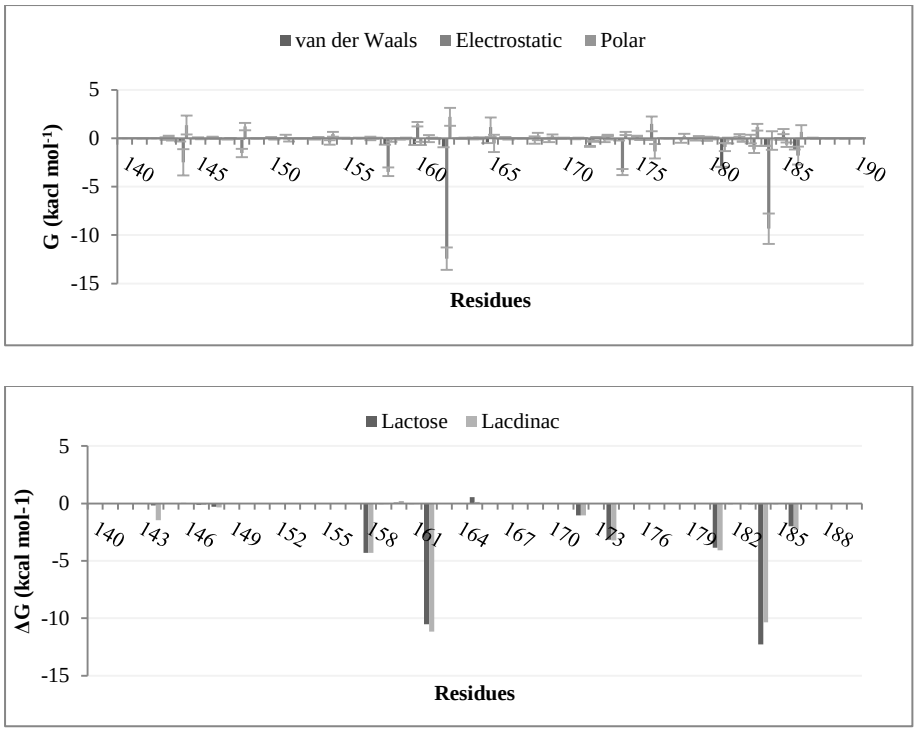


Figure 3.5. Pair wise energy decomposition analysis by MM-PBSA. Top: Energy contribution of the CRD aminoacids with relative standard deviations in gal-3-LacdiNAc complex. Bottom: Comparison of the energy contribution of the CRD aminoacids in the gal-3-lactose and gal-3-LacdiNAc complexes.

3.2.2. The recognition of the α-Gal epitope by gal-3

Complexes for gal-3 with the two α-Gal epitope glycans (**2** and **3**) were generated and submitted to docking calculations. The 3D perspective showed by the docked structures suggested that the presence of additional sugars in compounds **2** and **3** does not interfere with the recognition of the central lacNAc (Galβ1-4GlcNAcβ) moiety. The predicted binding mode of compound **2** (Figure 3.6) indicated that the interactions with the lacNAc moiety are kept as in the corresponding crystal structure, i.e. hydrogen bonds involving side chains of aminoacids His158, Arg162, Arg186 and Glu184, and the typical CH-π interaction between the Trp181 indole ring and the α-face of the β-linked GalNAc. Besides, the docking calculations predicted the positioning of the additional Galα1-3 onto the vicinal strand, providing extra hydrogen-bonds with the Arg144 and Asp148 side-chains. The docked binding pose was then submitted to 50 ns of MD simulation. The complex gal-3/compound **2** has been observed to be stable along all the simulation (Figure 3.7), as it possible to see from the RMSD of the backbone and sidechain of the protein. Nevertheless during the MD simulation, compound **2** establishes new interactions with the vicinal pocket through a slight relocation into the binding site (Figure 3.6-bottom).

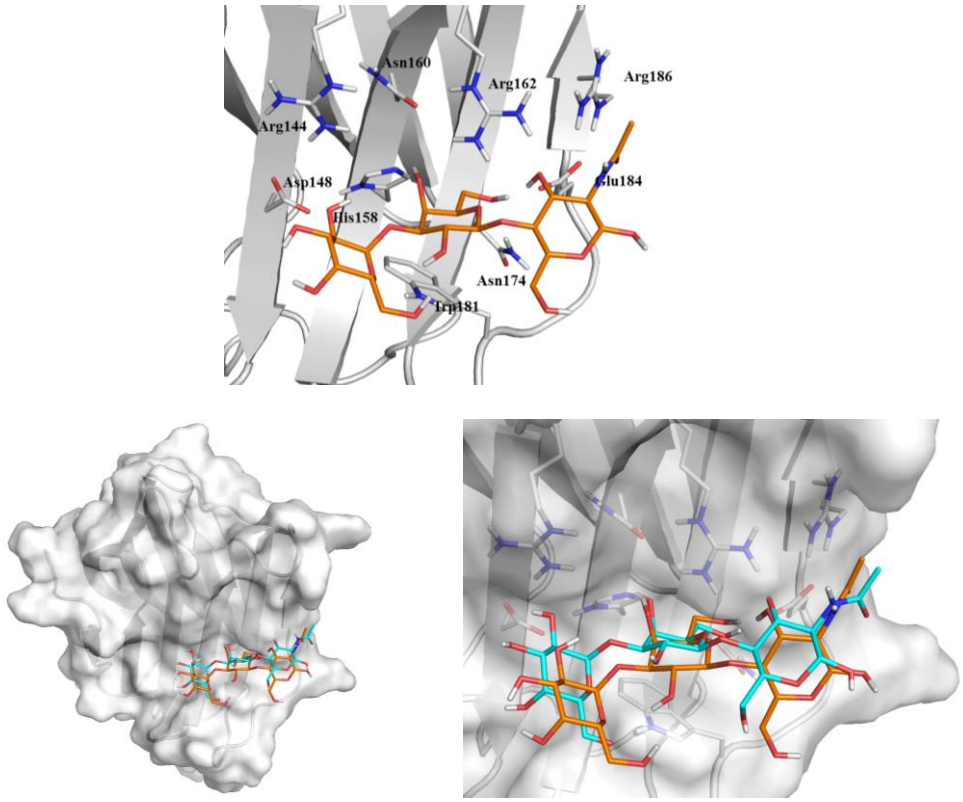


*Figure 3.6. Top: Structural docking model of the complex of the X-ray crystallographic structure of gal-3(grey surface, PDB ID 3ZSJ) with compound **2** (Galα1-3Galβ1-4GlcNAc orange). Bottom: Comparison between the docked binding pose of compound **2** (orange) and the average structure of MD simulation (cyan) at 23 ns are superimposed. New hydrogen bond interactions are predicted between the additional α(1-3)-linked galactose moiety and the Arg144 and Asp148 side-chains (shown in sticks).*

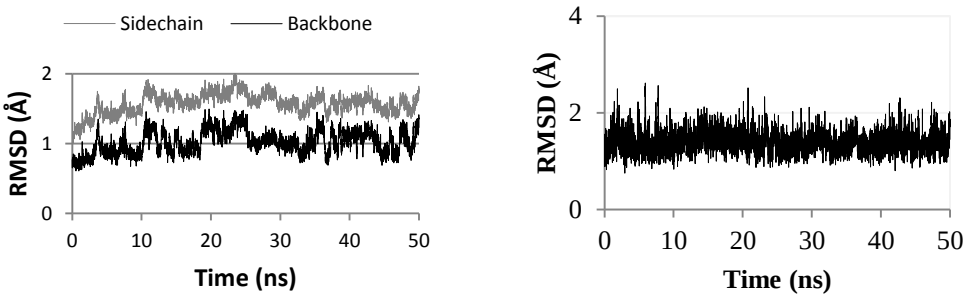


*Figure 3.7. Left: RMSD of the backbone and sidechain of gal-3 in complex with compound **2**. Right: RMSD of the heavy atoms in compound **2** along the 50 ns of MD simulation.*

For tetrasaccharide **3** two different binding modes were proposed. Additionally to the crystallographic-like pose (Figure 3.8, left), another binding pose was identified for which the reducing end Galβ unit makes contacts with the carbohydrate binding site and Galα 1-3-linked OH-4 establishes a hydrogen bond with Asp239 (Figure 3.8, right). These docking-based complexes are in good agreement with the selective chemical shift perturbations, occurring only in the presence of **2** and **3**, but not with lactose.

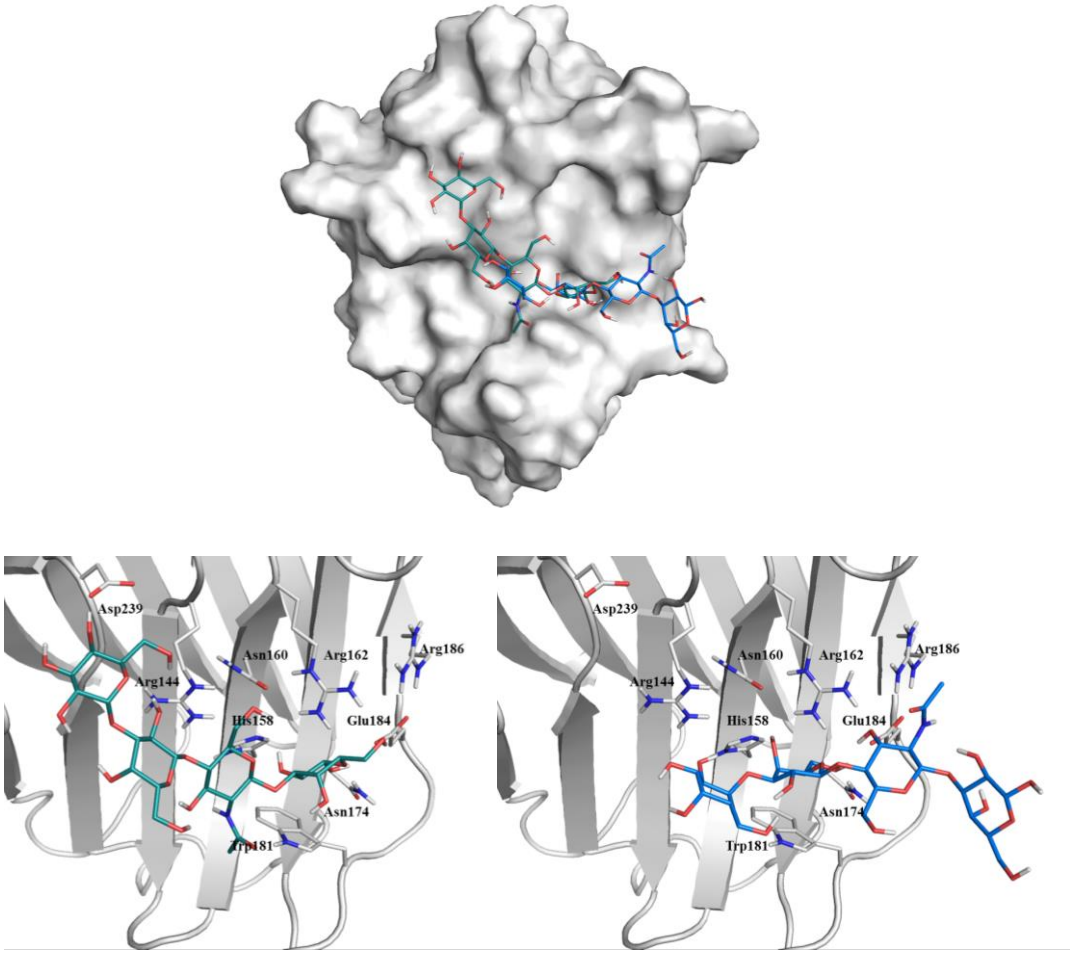


*Figure 3.8. Left: View of the major docked pose for compound **3** (deep teal sticks) bound to the X-ray crystallographic structure of gal-3 (PDB ID 3ZSJ, grey cartoons). The predicted binding pose for the LacdiNAc moiety closely resembles that of the lactose-like crystallographic structure. Main residues involved in the binding are depicted with grey sticks.*
*Right: Docked structure from the alternative cluster of the complex of the α-Gal tetrasaccharide **3** (blue sticks) with X-ray crystallographic structure of gal-3 (PDB ID 3ZSJ). The reducing Gal moiety makes contacts with the lectin, while the non-reducing residues are exposed to the solvent. This minor pose may explain the STD-data regarding the observations of certain degree of saturation at the reducing end gal unit.*

The stability of the different complexes was evaluated by MD simulations. The glycans-gal-3 complexes formed with both predicted poses for ligand **3** were highly stable, with very minor RMSD deviations in the Cα and sugar heavy atoms (below 2 Å, Figure 3.9).

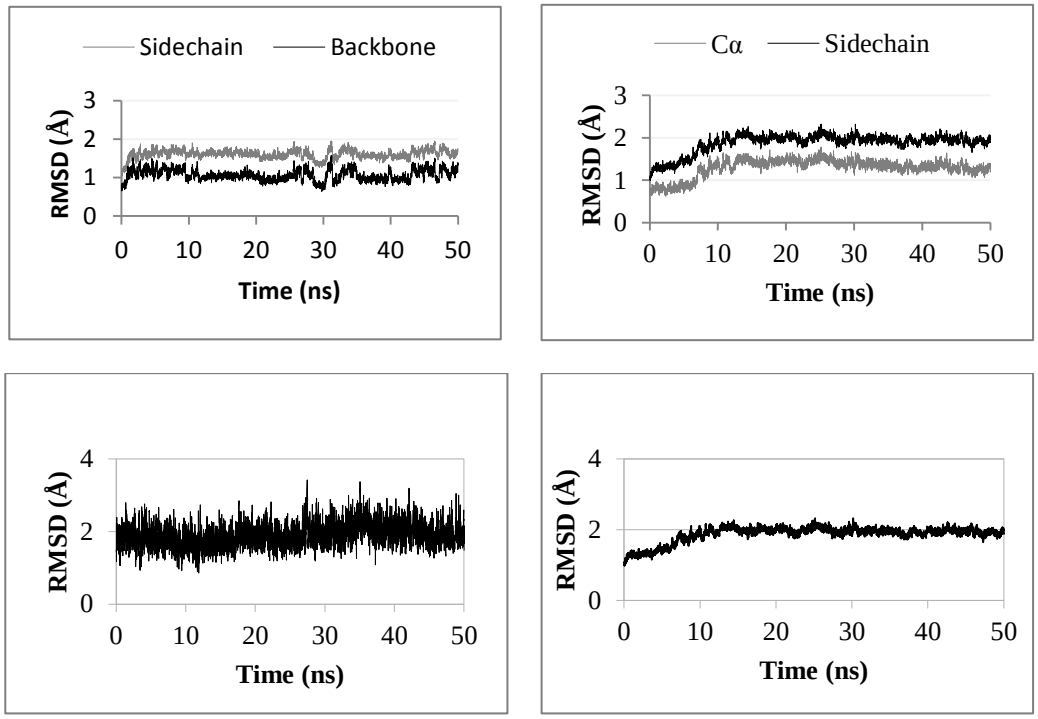


*Figure 3.9. Top: RMSD of the backbone and sidechain of gal-3 in complex with the major binding pose (left) and the minor one (right) of compound **3**. Bottom: RMSD of the heavy atoms the major binding pose (left) and the minor one (right) of compound **3** along the 50 ns of MD simulation.*

In any case, the HSQC data unequivocally indicate that the major complex corresponds to the most populated docked pose (structure shown in Figure 3.8-left), while the STD data suggest that a minor contribution of the alternative docked pose shown in Figure 3.8-right may also co-exist. The analysis of the data provided insights on the energy values of the new interactions for **2** and **3** relative to the disaccharides, as well as the associated movements.

From a MM-PBSA analysis, it was possible to assess subtle differences between the complexes. In particular, compound **2** establishes better interactions with Arg144 and Trp181 (Figure 3.10). This behaviour is also observed in the binding energy predicted from the docking studies.

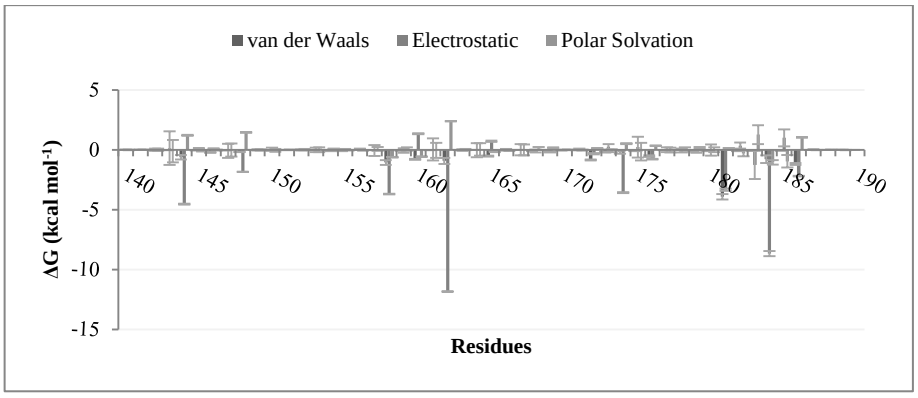


*Figure 3.10. Pair wise energy decomposition analysis by MM-PBSA. a) Energy contribution of the CRD aminoacids with their relative standard deviations in gal-3-compound **2** complex.*

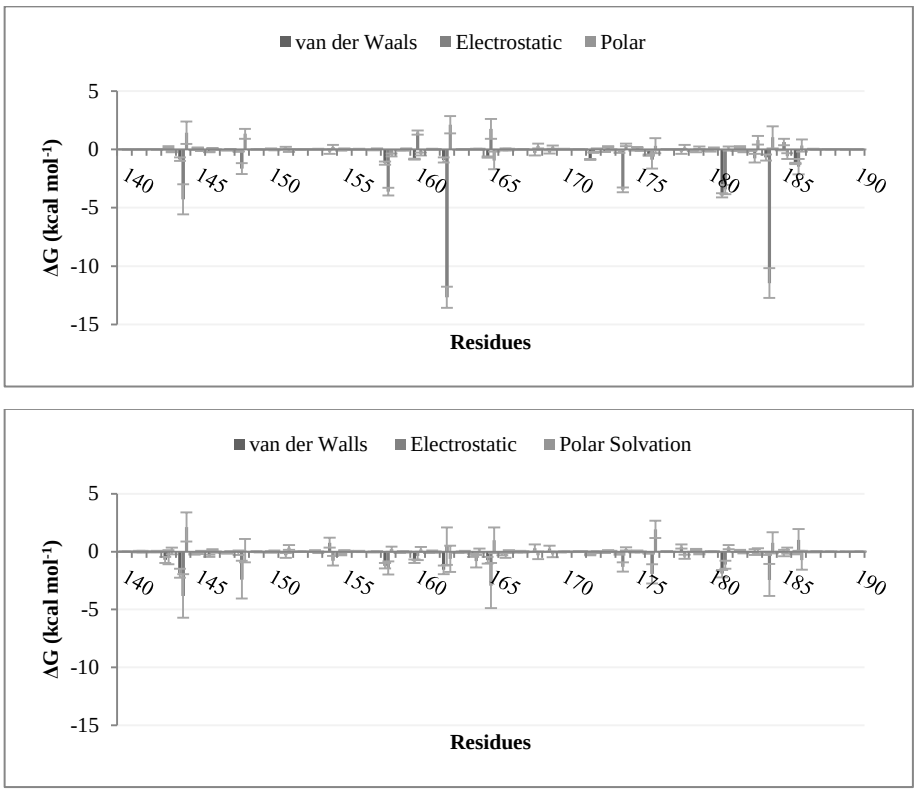


*Figure 3.11. Top: Energy contribution of the CRD aminoacids with their relative standard deviations of gal-3 in complex with the major binding pose of compound **3**.Bottom: Energy contribution of the CRD aminoacids with their relative standard deviations of gal-3 in complex with the minor binding pose of compound **3**.*

The energy contribution in the binding with ligand **2** (Figure 3.10) results to be very similar to the energy contribution of compound **3** (Figure 3.11). Appreciable difference is seen in the electrostatic contribution of Trp181 that in the case of the binding with compound **3** is higher due to the interaction between the NH of the indole ring and the O6 of the αGal moiety. As a general conclusion, the recognition of the LacdiNAc (compound **1**) and the α-Gal epitope-derived oligosaccharides **2** and **3**, all of which can be essentially understood as substituted lacNAc molecules, show analogous recognition features to those observed for that disaccharide and for lactose. In addition, selective local perturbations detected by NMR (Figure 3.12 and Figure 3.13), and interpreted by molecular modelling, permit to map the location of the additional Gal- decorating moieties at the adjacent pocket close to the Gal-binding site in *gal-3*.

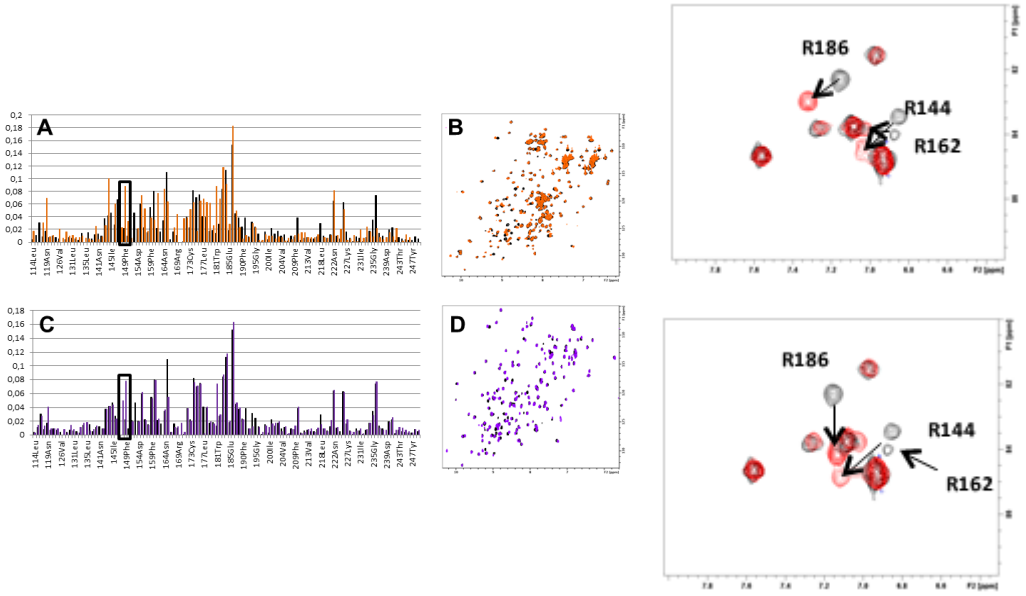


*Figure 3.12. Left: **A**: Comparison between the chemical shift perturbations observed for **2** (orange) and lactose (black). **B**: Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 (black) and with 10 eq. of **2** (orange). **C**: Comparison between the chemical shift perturbations for **3** (purple) and lactose (black). Key differences are highlighted. **D**: Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 CRD (black) and with 10 eq. of **3** (purple). Right: Top: Plot of the arginine region in the ^{1}H-^{15}N HSQC spectra of gal-3 (in black) upon addition of **2** (red). Arg186, Arg144, and Arg162 show clear chemical shift variations, especially Arg186 and Arg144. In contrast, Arg151, Arg129, Arg168, Arg169, and Arg183 do not show any perturbation upon ligand binding. Bottom: Plot of the arginine region in the ^{1}H-^{15}N HSQC spectra of gal-3 (in black) upon addition of **3** (red). Arg186, Arg144, and Arg162 show chemical shift variations, especially Arg186 and Arg144. In contrast, Arg151, Arg129, Arg168, Arg169, and Arg183 do not show any perturbation upon ligand binding. The changes for Arg144 and Arg162 are similar for both molecules. A different perturbation is observed for Arg186 in either complex. The variation of Arg144 is evident, while this was not the case for the complex with **1**.*

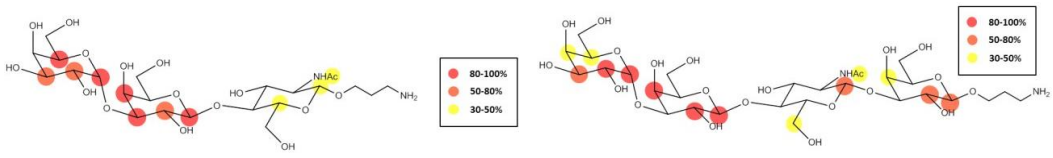


Figure 3.13. Left: Epitope mapping deduced from STD data, with colours showing relative STD intensities to the most saturated signal. Right: Epitope mapping deduced from STD data, with colours showing relative STD intensities to the most saturated signal.

3.2.3. The recognition of blood group A type II tetrasaccharide

A molecular model for the complex between gal-3 and **4** was then generated (Figure 3.14). From the STD (Figure 3.15) and the chemical shift perturbation analysis, it can be assumed that the lactose moiety of the tetrasaccharide is located at the lactose binding site and establishing the typical interactions: the α face of Gal establishes CH-π contacts with Trp181, and there are hydrogen bonds involving His158, Arg162, Asn174, Glu158 and Asn160 of the protein and OH4, OH6 and O5 of Gal and OH3 of GlcNAc, respectively. On the other side, the trNOESY analysis indicates that there is no conformational deviation from the free state upon binding. Thus, by using these two premises a starting complex was built, which was submitted to one initial 50 ns MD simulation. Different parameters were monitored in order to check the stability of the complex. The RMSD of the backbone of the protein was kept below 1.25 Å (Figure 3.14).

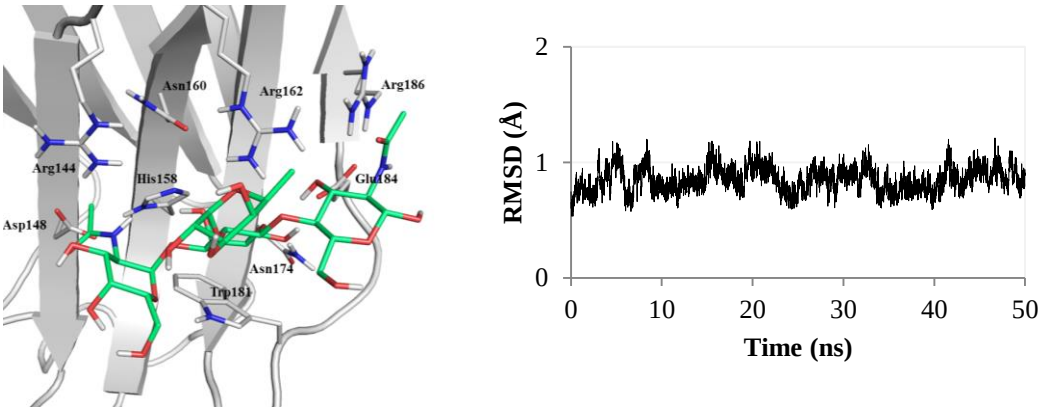


*Figure 3.14. Left: Structural docking model of the complex of X-ray crystallographic structure gal-3(PDB ID 3ZSJ) with compound **4** (green sticks). Aminoacids involved in the binding are shown in grey sticks. Right: RMSD in the Cα of gal-3 in complex with compound **4**.*

As mentioned above, the typical interactions for the recognition of lactose are kept. The only H-bond in which GlcNAc participates is that between OH6 and the indole NH of Trp181. Besides, a hydrophobic interaction between the methyl of the acetyl group with Ala146 could take place, although from the MM-PBSA analysis (Figure 3.16) the van der Waals contribution was small (-0.4 kcal mol^{-1}). The fitting of GlcNAc residues requires rearrangement of the side chain of Arg144 that folds back towards the protein backbone adopting a particular orientation. This fact is in agreement with the observation of a strong chemical shift perturbation of its backbone NH and side chain resonances upon ligand binding, effect not observed with other ligands. The fucose residue is far from the protein surface, and establishes no contact with the lectin.

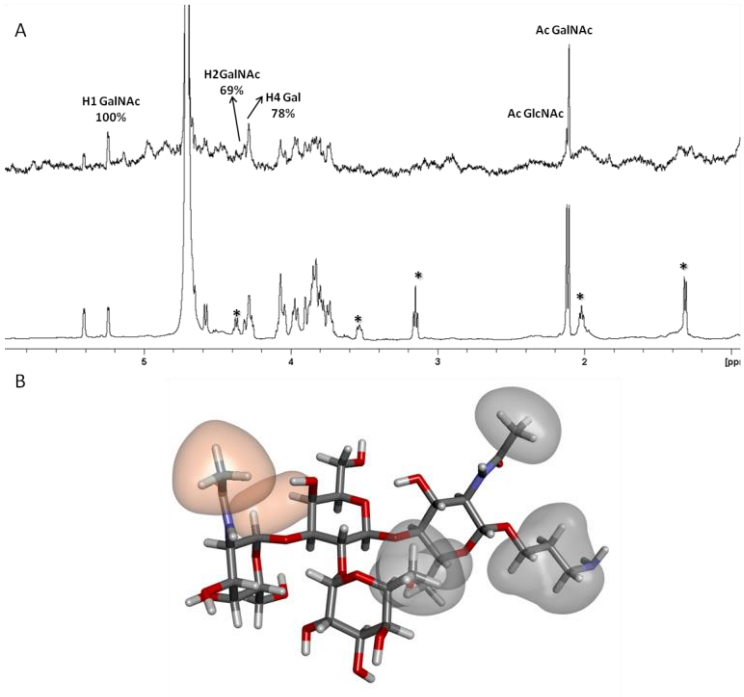


*Figure 3.15. A: On top STD and below corresponding reference spectra for a 43 uM/2mM gal-3/**4** sample. Signals with an asterisk indicate protons with no or very weak STD effect. The methyl groups of the Ac are assigned to their corresponding residues. Proton signals with strongest STD effect clearly identified are annotated with the corresponding STD percentage relative to the largest value (H1GalNAc).*

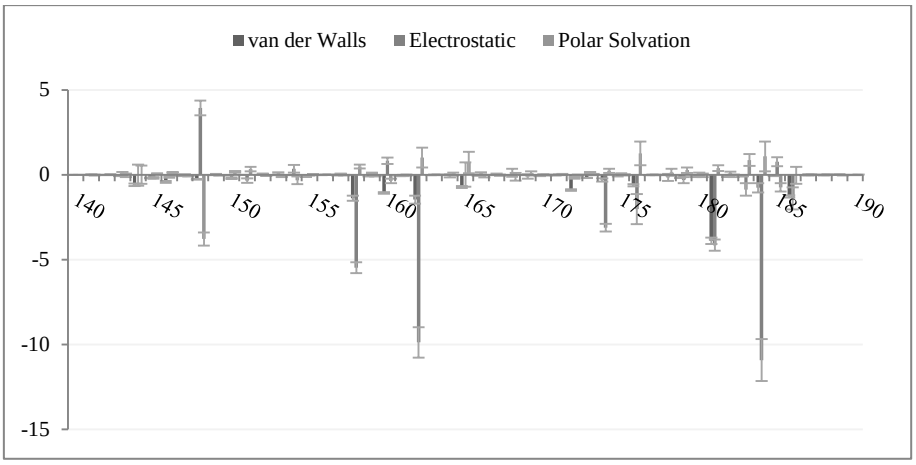


*Figure 3.16. Pair wise energy decomposition analysis by MM-PBSA. Energy contribution of the CRD aminoacids with their relative standard deviations in gal-3-**4** complex.*

Previous studies have shown that gal-3 recognizes the blood group antigens with higher affinities with respect to lactose.[9] In particular, **4** has been reported to bind gal-3 more than one order of magnitude better than lactose at 298K. Thermodynamic studies showed a difference of ca. 1.9 kcal mol^{-1}, with a strong enthalpy-entropy compensation phenomenon. The authors claimed that there is a positive enthalpic contribution, related to a water-mediated hydrogen bond between OH4 of the fucose and Glu165. However, our data suggest that the fucose moiety is not involved in any contact with the lectin and that, therefore, the reported increase of enthalpy could be related to a hydrogen bond between OH6 of GalNAc and NH of the Trp181 indol, as also suggested from MM-PBSA results.

3.3. Material and Methods

Protein Preparation: Around 42 crystallographic structures of galectin-3 are available at the Protein Data Bank,[13] 40 of those are from *Homosapiens*. For our studies we used PDB 3ZSJ[12] with a resolution of 0.86 Å, in complex with the natural ligand lactose. The protein were prepared using the Protein Preparation Wizard tool included in Maestro.[14] Water molecules, co-factors of crystallization and ligands were removed, missing atoms were added, side chains and loops were filled by Prime, and hydrogens were added with Epik at physiological pH. For the preparation of the protein, the protonation of His158 was considered as HID.[12] This final structure of the protein was minimized with OPLS 2005 force field as implemented in Maestro with implicit solvent (water). The final minimized structure was used for docking purposes.

Ligand Preparation The ligands (Scheme 3.1) used in this study were lactose (compound **1**, βGalNAc1-4GlcNAc), α-Gal trisaccharide (compound **2**, αGal1-3βGal1-4GlcNAc), α-Gal tetrasaccharide (compound **3**, αGal1-3βGal1-4βGlcNAc1-3Gal) and the blood group A type II tetrasaccharide (compound **4**, GalNAcα1-3(Fucα1-2)Galβ1-4GlcNAc). β configuration was used for all the terminal sugar. The ligands were generated using the carbohydrate builder from GLYCAM,[15] and minimized with semi-empirical method AM1. In order to get an optimal starting geometry for the docking calculations, a conformation analysis was performed with Macromodel[14] using MM3* force field. Most stable conformations for glycosidic linkages and O5C5C6O6 dihedral angle were chosen for each monosaccharide.

Validation of the docking methodology. The docking protocols with AutoDock4, VINA and Glide were followed for the prediction of the binding of lactose in the galectin-3 (from PDB ID 3ZSJ). The X-ray crystallographic pose was predicted in the three cases with the three programs AutoDock4, VINA and Glide.

Docking Calculations Docking calculations we performed by means of three different programs: AutoDock4,[16] Vina,[17] and Glide[18]. In order to validate the docking protocol, we docked lactose in gal-3, obtaining results in full agreement with the crystallographic binding pose with the three docking programs AutoDock4 (RMSD = 0.727 Å), Vina (RMSD = 0.906 Å) and Glide (RMSD = 0.686 Å). Lamarckian genetic algorithm in AutoDock was used to sample different conformations of the ligands, by randomly changing the torsion angles and overall orientation of the molecule. A three-dimensional grid was defined (68 x 60 x 40 Å) centered on an equidistant point to His158 and Try181. The grid spacing was 0.375 Å, and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock were also used. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å. The clusters were ranked by the lowest energy representative of each cluster. Calculations with Glide were only able to reproduce the crystallographic pose for lactose when two crystallographic water molecules were kept: the one close to Glu184 and Arg186, and the one close to Arg144 and Arg162. Van der Waals scaling factor and partial charge cutoff were kept as default, which are 1.0 and 0.25, respectively. The other parameters were kept as default as well. The key residues from the CRD (Trp181, His158 and Arg162) were used to generate the box center position, and box size was 20x26x20 grid point, being the spacing between points of 1 Å. The output file was generated writing 5 poses per ligand conformation and the RMSD for the clustering was set in less than 2 Å.

Molecular dynamics (MD) simulations of the ligands in the free state. All the studied ligands were submitted to 10 ns MD simulations with explicit water to monitor the preferred conformations in the free state. Every dihedral angle was monitored observing that O5C5C6O6 dihedral angle was maintained around -60 in the expected gauche-gauche conformation for glucose residues, and around +60 in the expected gauche-trans conformation for galactose residues according to the conformational analysis and the literature.[19]

Molecular dynamics (MD) simulations of the gal-3 complexes. The best docked solutions for each ligand were considered as starting geometries to perform the MD simulations. GLYCAM06, Gaff, and ff99EP were used as force fields. All MD simulations were carried out by using the sander module in AMBER12 (www.ambermd.org). Four Cl- counterions were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 50 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 5000 structures for each complex, which were finally analysed.

Bibliography

1. Solís, D.; Bovin, N. V.; Davis, A. P.; Jiménez-Barbero, J.; Romero, A.; Roy, R.; Smetana, K.; Gabius, H.-J., A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *BBA. General subjects* **2015,** *1850* (1), 186-235.
2. Vasta, G. R., Roles of galectins in infection. *Nat. Rev. Microbiol.* **2009,** *7* (6), 424-438.
3. Dam, T. K.; Brewer, C. F., Lectins as pattern recognition molecules: The effects of epitope density in innate immunity. *Glycobiology* **2010,** *20* (3), 270-279.
4. Sato, S.; Nieminen, J., Seeing strangers or announcing "danger": Galectin-3 in two models of innate immunity. *Glycoconjugate J.* **2002,** *19* (7-9), 583-591.
5. Stowell, S. R.; Arthur, C. M.; Dias-Baruffi, M.; Rodrigues, L. C.; Gourdine, J.-P.; Heimburg-Molinaro, J.; Ju, T.; Molinaro, R. J.; Rivera-Marrero, C.; Xia, B., Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* **2010,** *16* (3), 295-301.
6. Daniels, G.; Bromilow, I., *Essential guide to blood groups*. John Wiley & Sons: 2011.
7. Van Den Berg, T. K.; Honing, H.; Franke, N.; Van Remoortere, A.; Schiphorst, W. E.; Liu, F.-T.; Deelder, A. M.; Cummings, R. D.; Hokke, C. H.; Van Die, I., LacdiNAc-glycans constitute a parasite pattern for galectin-3-mediated immune recognition. *J. Immunol.* **2004,** *173* (3), 1902-1907.
8. Stowell, S. R.; Arthur, C. M.; Mehta, P.; Slanina, K. A.; Blixt, O.; Leffler, H.; Smith, D. F.; Cummings, R. D., Galectin-1,-2, and-3 exhibit differential recognition of sialylated glycans and blood group antigens. *J. Biol. Chem.* **2008,** *283* (15), 10109-10123.
9. Bachhawat-Sikder, K.; Thomas, C. J.; Surolia, A., Thermodynamic analysis of the binding of galactose and poly-N-acetyllactosamine derivatives to human galectin-3. *FEBS Lett.* **2001,** *500* (1-2), 75-79.
10. Greenwald, A. G.; Jin, R.; Waddell, T. K., Galectin-3-mediated xenoactivation of human monocytes. *Transplantation* **2009,** *87* (1), 44-51.
11. Jin, R.; Greenwald, A.; Peterson, M. D.; Waddell, T. K., Human monocytes recognize porcine endothelium via the interaction of galectin 3 and α-GAL. *J. Immunol.* **2006,** *177* (2), 1289-1295.
12. Saraboji, K.; Håkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M., The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: Ultra-high-resolution structures and water dynamics. *Biochemistry* **2011,** *51* (1), 296-306.
13. Protein Data Bank. http://www.rcsb.org (accessed March 29, 2017).
14. *Maestro version 9.3, Schrödinger, LLC, New York, NY*, 2012.
15. http//glycam.ccrc.uga.edu.
16. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.

17. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
18. *Glide version 6.1, Schrödinger, LLC, New York, NY*, 2012.
19. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (19), 10541-10545.

Annex III

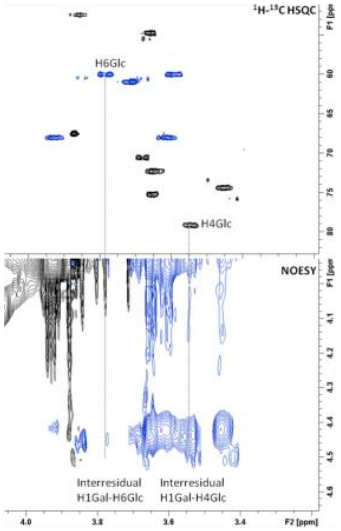


*Figure 3.A1. On top ^{1}H-^{13}C HSQC spectrum of **1**, and below NOESY spectrum. The dashed lines indicate the assignment along F2 of the GlcNAc protons showing interresidual NOE (positive) with H1GalNAc.*

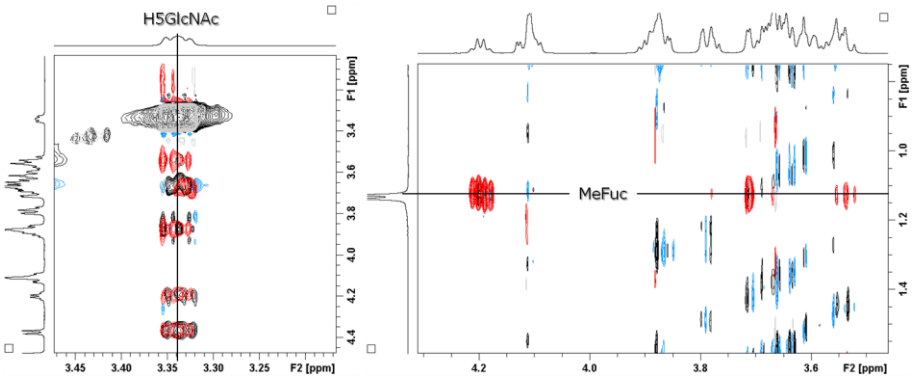


Figure 3.A2. Superimposition of the ROESY (grey and red contours) and NOESY (black and blue contours) of 4 in different regions. While the crosspeaks of H5GlcNAc are of similar intensity in the ROESY and the NOESY, those corresponding to MeFuc are lost in the NOESY spectrum.

*Table 3.A1. Interproton ROE intensities and derived distances for **4**. In bold interresidue ROEs.*

proton	proton	ROE crosspeak intensity	ROE derived distance (Å)	Average distance (Å) from MD (10n)
αGalNAc				
H1	H2	strong		
	H4 Gal	**strong**	**2.2**	**2.2**
	H3 Gal	**weak**	**3.3**	**2.9**
αFuc				
H1	H2	strong		2.4
	H3	weak *		
	H3 GalNAc	**strong** *	**2.3**	**2.8**
	H2 Gal	**strong**	**2.4**	**2.6**
	H3 Gal	very weak		
	H5 GalNAc	**weak (noise)**	**2.9 (noise)**	**2.8**
H5	H4	strong		
	H3	medium		2.5
	H5 GlcNAc	**medium**	**2.4**	**2.7**
H6 (Me)	H5	strong		
	H4	strong		
	H3 GlcNAc	**med/weak**	**2.8**	
	H5 GlcNAc	**weak**	**3.0**	
	H2 Gal	very weak		
βGal				
H1	H5 §	strong		
	H3 §	med (?)		2.7
	H4GlcNAc	**strong**	**2.2**	**2.4**
	H6 GlcNAc §	med (?)		

- *and § overlap*

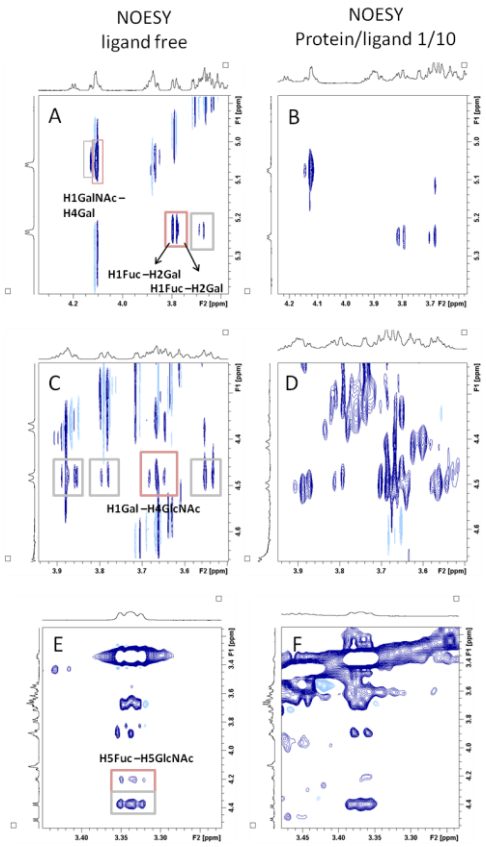


*Figure 3.A3. Comparison between the NOESY spectrum of **4** free in solution (A, C, E) and in the presence of gal-3 (B, D, F) in different regions. Inter-residue NOEs are annotated and marked with pink squares, intra-residue with grey squares.*

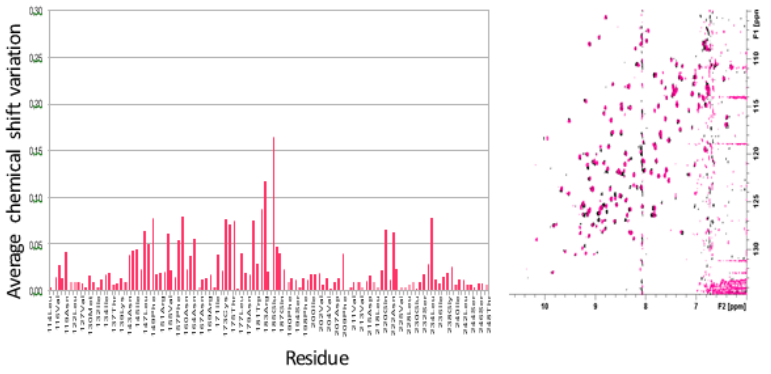


Figure 3.A4. ***A:*** *Average chemical shift differences for backbone amides* $\{[\otimes\delta(^{1}H)]^{2} + [(\otimes\delta(^{15}N))/2)]^{2}\}^{0.5}$ *of gal-3 upon addition of 20 equivalents of lactose.* ***B****: Superimposition of the* ^{1}H-^{15}N *HSQC spectra of apo gal-3 (black) and in the presence of 20 equivalents of lactose (pink).*

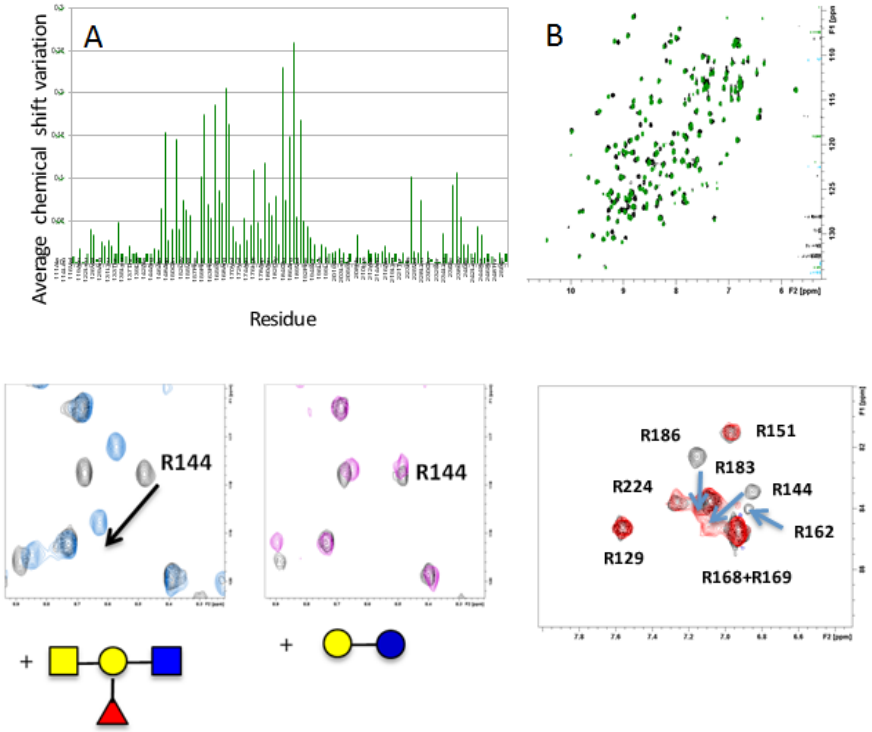


Figure 3.A5. ***A****: Average chemical shift differences for backbone amides* $\{[\otimes\delta(^{1}H)]^{2} + [(\otimes\delta(^{15}N))/2)]^{2}\}^{0.5}$ *of gal-3 upon addition of 20 equivalents of lactose.* ***B****: Superimposition of the* ^{1}H-^{15}N *HSQC spectra of apo gal-3 (black) and in the presence of 20 equivalents of lactose (pink).* ***C:*** *Left: Region of the backbone* ^{1}H-^{15}N *HSQC spectra of gal-3 (in black) upon addition of* ***4*** *(left, in blue) and lactose (right, in pink). Right panel. Plot of the arginine region in the* ^{1}H-^{15}N *HSQC spectra of gal-3 (in black) upon addition of* ***4*** *(red). Arg186, Arg144, and Arg162 show chemical shift variations variations, especially Arg186 and Arg144. In contrast, Arg151, Arg129, Arg168, Arg169, and Arg183 do not show any perturbation upon ligand binding.*

CHAPTER 4

COMPUTATIONALLY-DESIGNED GALECTIN BINDERS WITH HIGH-AFFINITY PROFILE

Synthesis of the compounds was performed at the laboratories of Prof. Stefan Oscarson (University College Dublin, Ireland) and Prof. José Carlos Menéndez (University Complutense of Madrid, Spain). NMR studies are being performed at the laboratory of Prof. Jesús Jiménez-Barbero (CIC bioGUNE, Derio, Spain). ITC studies are being performed at CIB-CSIC and IQFR-CSIC with the collaboration of Prof. Margarita Menéndez.

4.1 Introduction

Galectins are β-galactosides binding proteins involved in many physiological and pathological processes. In mammals, they exert a wide range of biological functions and play an important role in immune and inflammatory responses, and regulation of the cell cycle, among other processes. Also galectins have been identified as key players in cancer and immune pathologies pointing towards these proteins as important targets for development of drugs for cancer therapy and immunotherapy.[1-5] Current research confirms the involvement of galectins in tumour progression and indicates that galectins could serve as therapeutic targets, and have prognostic value.[1] In particular, galectins -1, -3, and -7 are adhesion/growth-regulatory galectins whose functions have been shown to be correlated among them in several tumoral processes.[6-7]

Galectin-3 participates in the regulation of inflammation,[8] and mediates important tumor-related functions (tumorigenesis, cell adhesion, proliferation, differentiation, angiogenesis, metastasis), being involved in the development and malignancy of several types of tumor.[9] Therefore, synthetic galectin-3 binders have been pursued for the understanding of the molecular regulation of gal-3 expression and for the development of new antitumoral therapeutic strategies.[10-12] Galectin-1 also acts in cellular adhesion, mobility and invasion, tumor-induced angiogenesis, and apoptosis, although the mechanisms of action in these processes are still not well understood. However, the overexpression of galectin-1 in cancer progression indicates that the role of galectin-1 is significant.[13-14] In fact, both galectins, -1 and -3, seem to be the major players in cancer biology and, therefore, have stimulated significant research interest. However, less is known about galectin-1 pathways compared to galectin-3.[1, 15]

Galectin-7 is one of the less well studied galectins. Information regarding altered galectin-7 expression between normal and tumor types is also mostly limited to single reports. In addition, some findings resemble the differences in galectin-3 expression with respect to cancer subtype and cellular localization. These observations further exemplify that galectin localization might be used to distinguish between histological subtypes.[1]

The possibility of targeting galectins differently has been pursued as an interesting aim in the glycomimetics field. Selective galectin binders could help in cellular essays to unravel the mechanisms regulated by galectins and their different (and sometimes opposite roles). Moreover, selective galectin ligands with improved affinity would also be interesting molecules with therapeutic applications. Although it is known that galectins bind carbohydrates through a carbohydrate recognition domain (CRD), a deeper knowledge of the interactions that take place and the search of new binding pockets to anchorage putative modulators, are key factors for the development of galectin modulators.[10]

Different strategies have been followed to obtain glycomimetics targeting galectins with improved affinity and selectivity. Given galectins specifically recognize β–galactosides, an evident starting point is the modification of the galactose moiety to produce different monosaccharides.

In particular, changes on the anomeric position and position 3 of the galactose ring have been performed. Nevertheless, the most used strategy has pursued the modification of disaccharide scaffolds, mainly the lactose. Chemical modifications include: the substitution of the glycosidic linkage in order to improve the glycolytic stability; the insertion of relatively long linkers on position 1 in order to have multivalent systems; the introduction of substituents in position 3 of the galactose moiety and/or positions 2 and 3 of the glucose moiety; and the use of other disaccharide scaffolds, such as the digalactose.[16-18] To date, a large number of galectin-1 and galectin-3 inhibitors have been synthesized, although it is scarce the information about selective galectin-7 inhibitors.

In rational drug design it is evident that structural information is of paramount importance.[19] For example, the knowledge of the conformational change of Arg144 in galectin-3, in presence of aromatic ring establishing a cation-π interaction, has permitted the synthesis of several galectin-3 inhibitors.[20] However, to the best of our knowledge, there are no computer-aided studies for the design of galectin ligands with improved selectivity and affinity.

In particular we are interested in the identifications of novel scaffolds able to lead to second generation of lactose derivatives with different selective binding properties towards human galectins 1, 3, and 7 that be helpful for the study of galectin mechanisms and putative activity as galectin modulators. We have undertaken fragment-based virtual screening studies to identify moieties which can be accommodated in the secondary pocket close to the carbohydrate recognition domain of the galectin. Our studies have led to the design of novel ligands with improved predicted binding affinity, and, putative selectivity towards gal-1, -3 and -7. These studies can assist in the further development of synthetic glycans with potential therapeutic applications.

4.2 Results

4.2.1 Design of selective galectin ligands.

From the study of galectins is very easy to underlines how CRD could be very similar among protein of the same family. They shared some common futures such as the presence of an arginine, a histidine and a tryptophan. So we start to explore the adjacent binding site of the CRD which is already known to have some conformational changes depending on the ligand interacting with them, such as the conformational change of the Arg144 in the galectin 3 in the presence of a ligand with an aromatic portion.[21] So, we focus the attention in particular on galectin-1, -3 and -7 which seem to be very interesting as therapeutical targets. The necessity to aim to a specific galectin is mandatory.

Given that β-galactose is the molecular pattern characteristic of the specific recognition of all galectins, this moiety is usually maintained in the reported synthetic glycomimetics aiming to target galectins. Moreover the natural ligand for galectins is a lactose or lactosamine. Therefore, we focused our attention in the adjacent pocket of the CRD in order to explore the possibility to identify possible specific moieties with good binding properties (Figure 4.1). We explore this region through binding site pocket finder software (SiteMap).

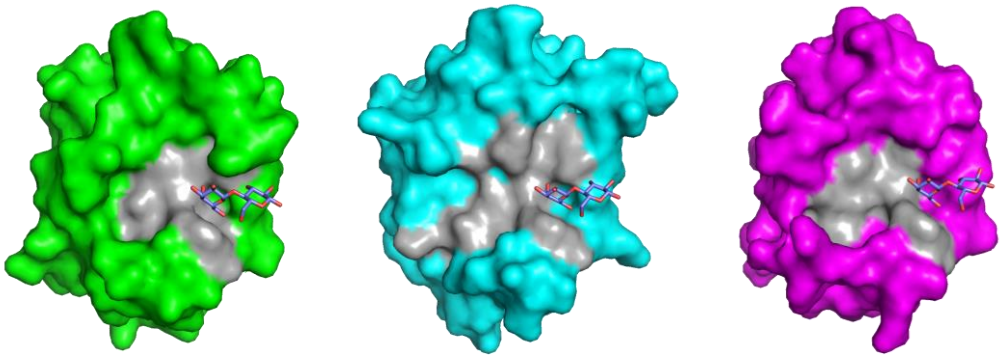


Figure 4.1. Surface representation of gal-1 (PDB ID 1GWZ, green), gal-3 (PDB ID 3ZKJ, cyan) and gal-7 (PDB ID 4GAL, magenta).

4.2.2 Virtual screening of a fragment library.

Keeping the lactose as in the crystal binding pose, we performed a fragment-based virtual screening in the adjacent region using two different protocols. In particular, we used the area delimited by Ser29-Lys63-Tyr119 for galectin-1, Arg144-Lys176-Lys233 in galectin -3 and Lys6-Arg31-Lys64 for galectin-7. The surface area of these pockets was approximately 1200 $Å^3$. The fragment library used in this work was the Maybridge data base, which contains more than one thousands drug-like fragments. FLAP and Glide were used as VS tools. The choice of the fragments was dictate from the best scoring marks.

From FLAP virtual screening, two amino acids per galectin were chosen to determine the centre of two different pockets that were merged together, in order to study the complete adjacent pocket. Pocket extension and thickness were 8 and 26 for all the cases, respectively, and high accuracy performance option was chosen. Best fragments, i.e. fragments with a GlobScore of more than 0,9 (26 for gal-1, 32 for gal-3, and 30 for gal-7), were considered to build the methyl-lactose (OMe-Lac) based ligands. In the case of Glide, VS protocol comprises 3 levels of performance: High-Throughput Virtual Screening (HTVS), Standard Precision (SP) docking and Extra Precision (XP) docking. In every step, only 30% of the fragments were kept. By applying this protocol, 26 fragments were asked to Glide to be selected for each galectin 1, 3 and 7, starting from the Maybridge database. An average of 25% of the fragments was repeated in more than one galectin by both FLAP and Glide VS protocols.

In total, taking into account both results from FLAP and Glide, and alternative binding poses and possible linkage connections with the lactose moiety, more than 300 possible fragment poses were analyzed in order to build the full OMe-Lac ligands based on these fragments. Only fragments with the appropriate orientation and position were considered, i.e. binding pose of the fragment should be near lactose binding site and the linkage position should be oriented towards position 3 of galactose residue. Finally, 106 fragments were selected to build OMe-Lac based ligands (Figure 4.2).

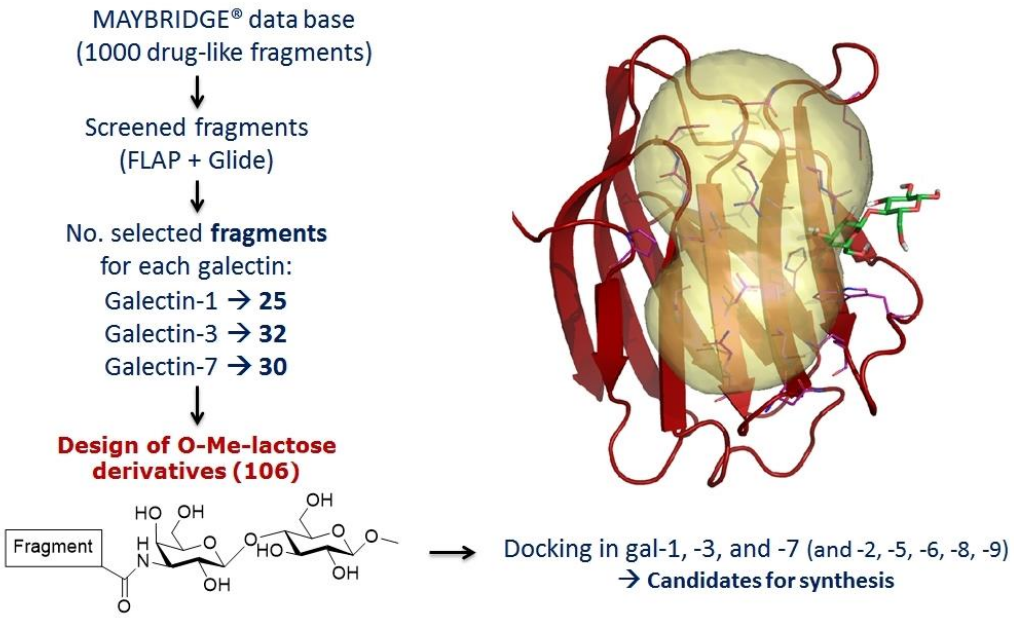


Figure 4.2. Left: Scheme of the protocol applied for the virtual screening on each galectin. Right: Cartoon representation of galectin 3, bound to lactose. Yellow sphere represents the merge of two pockets used to perform virtual screening with the Maybridge database.

4.2.3 Docking of the designed methyl-lactose derivatives.

For synthetic reason, we designed the ligands based on the OMe-Lac scaffold and more than one hundred (106) structures were built using the best binding fragments. The three-dimensional structures were then prepared (minimized and charged) to docking purpose. To validate the protocol and to estimate an improvement of affinity we firstly perform docking studies by means of AutoDock4 with OMe-Lac. Then the new 106 ligands were docked at first in galectin-1 (PDB ID 1GZW), galectin-3 (PDB ID 3ZSJ and 1KJR) and galectin-7 (PDB ID 4GAL). Moreover we apply the same protocol with the other human galectins (galectin-2 PDB ID 5DG2, galectin-4 PDB ID 4XZP, galectin-8 PDB ID 4BMB, galectin-9 PDB ID 3NV4 and galectin-10 PDB ID 1G86).

Focusing on galectin-1, -3 and -7 we compared the predicted docked poses for each designed compound with each galectin. For each compound the lowest energy lactose-like binding pose was taken into consideration. The choice of best candidates for synthesis was dictated by different criteria. Among lowest-energy binding poses we chose those with an improvement of energy in comparison with the predicted binding energy of the OMe-Lac for the same galectin and with a worsening of the binding energy for the other galectins, comparing with the OMe-Lac predicted binding energy. Among the best candidates, we chose compounds J013, J228 and J683 given the commercial availability or the synthetic accessibility of the corresponding moiety to be condensed to the amino-lactose (Figure 4.3).

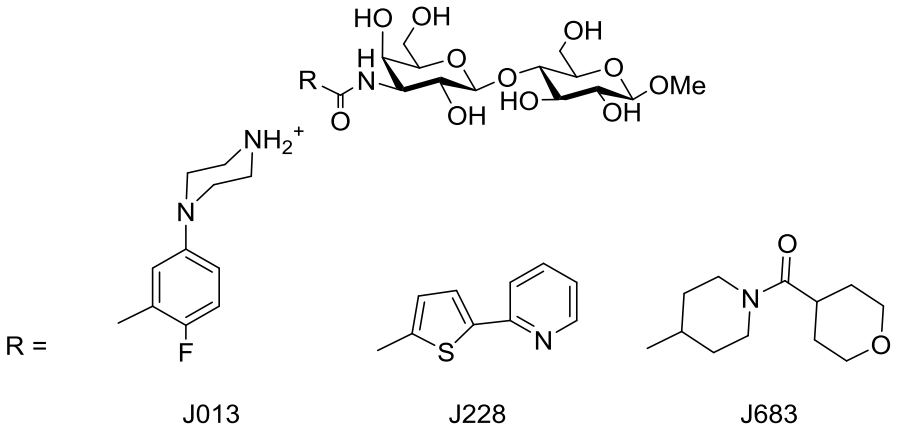


Figure 4.3. 2D representation of the best candidate compounds obtained by crossing the two VS protocols (FLAP and Glide).

Preliminary results from NMR and ITC studies confirm the high affinity gained by the designed compounds. At the present moment of writing this Doctoral Thesis, the NMR and ITC experiments are not completed and we do not have available accurate data to be included here. Deeper analysis is in progress, together with MD simulations of the corresponding galectin complexes in order to elucidate the atomic details of the recognition processes.

4.3 Material and methods

Protein Preparation: More than 150 X-ray crystallographic structures of human galectins are available at the Protein Data Bank[22]: 27 structures for gal-1, 4 for gal-2, 40 for gal-3, 12 for gal-4, 15 for gal-7, 28 for gal-8, 15 for gal-9, and 4 for gal-10. For our studies we used the X-ray structures with the following PDB IDs: 1GZW for gal-1, 5DG2 for gal-2, 3ZSJ and 1KJR for gal-3, 4XZP for gal-4, 4GAL for gal-7, 4BMB for gal-8, 3NV4 for gal-9, and 1G86 for gal-10. In the case of gal-3, two different structures were used in order to take into consideration the conformational change of the Arg144 (Figure 2.3.1). As shown in Figure 2.4, the presence of a ligand with an aryl group in position 3 of the galactose moiety induces a change in the conformation of the Arg144 side chain and the establishment of CH-π interaction, increasing the affinity of the ligand towards the protein.[28]

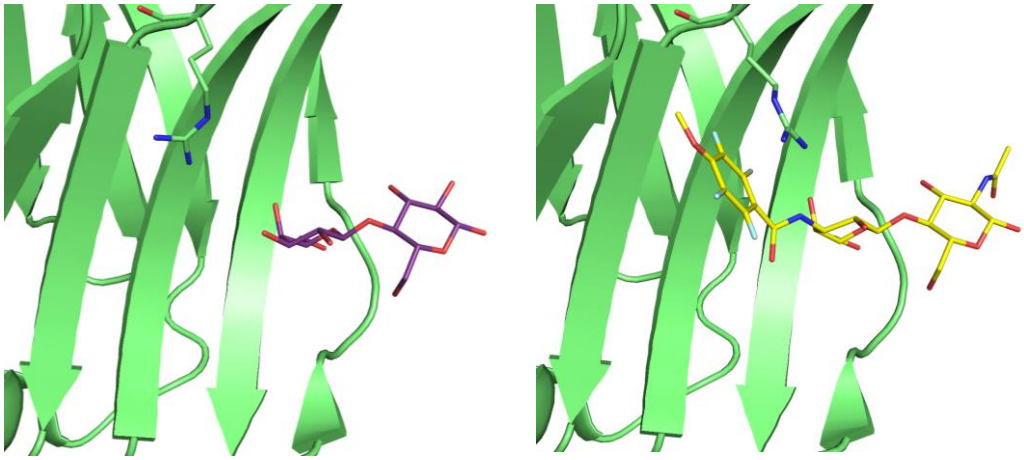


Figure 2.4. Crystal structures of gal-3 (in green) with different ligands. Left gal-3 complexed with lactose (represented in purple sticks, PDB ID 3ZSJ). Right: gal-3 complexed with L3 (represented in yellow sticks, PDB ID 1KJR). Arg144 in both crystal structures is shown.

The proteins were prepared using the Protein Preparation Wizard tool included in Maestro.[23] Water molecules, co-factors of crystallization and ligands were removed, missing atoms and cap termini were added, side chains and loops were filled by Prime, and hydrogens were added with Epik[24] at physiological pH. For the preparation of the protein, the protonation of the main histidine in the CRD (His158 in gal-3) considered as HID.[25] This final structure of the protein was minimized with OPLS 2005 force field as implemented in Maestro with implicit solvent (water). The final minimized structure was used for docking purposes.

Virtual Screening of Maybridge database. To perform the virtual screening in galectins -1, -3 and -7 Maybridge database was imported.[26] Maybrige database consists of 1002 different drug-like fragments of different chemical composition and properties. Additional entries were added for the L2 benzene fragment, L3 tetrafluorated benzene fragment and a naphthalene fragment as reference fragments with reported experimental data.[21] The protonation state of every fragment at physiological pH was calculated by Epik.[24] Two different virtual screening protocols were applied by means of two VS programs: FLAP[27] and Glide.[28]

VS *Protocol with* FLAP program. The pocket point radius was 2.0 Å and high accuracy performance was chosen. To explore only the adjacent accessory pocket of each galectin, we took in consideration the "lactose space" setting as macromolecule the complex of each galectin with lactose (**¡Error! No se encuentra el origen de la referencia.**). Table 4.1 shows the two aminoacids chosen as the centre of the pocket of each galectin. The pocket extension and thickness were 8 and 26 for all the cases, respectively.

Table 4.1. Chosen aminoacids for the pocket centre to perform virtual screening on galectin-3, -1 and -7.

Protein	***Galectin-3***	***Galectin-1***	***Galectin-7***
Centric aminoacids	*Arg144/Asp148*	*Val31/Asn33*	*His34/Arg32*

The VS protocol was validated using Chembl database.[29] For the validation protocol we use the galectin-3. In Chembl database we found 62 known as active galectin-3 binders and 48 as inactives. So, we performed a VS in the pocket centred in the Trp181 with extension and thickness of 15 and 5, respectively.

Building of the designed compounds. To build the designed compounds, OMe-Lac (βGal1-4βGlc-OMe) scaffold was used. The OMe-Lac moiety was obtained by means of GLYCAM[30] carbohydrate builder tool. The selected fragments from VS studies (32, Glob-Prob score up to 0.9 for FLAP + Glide) were attached through an amidic bond to the position 3 of the galactose. These 106 new ligands were then minimized with MMFFs force field with Macromodel[31] and Gasteiger charges were added with AutoDockTools.[27]

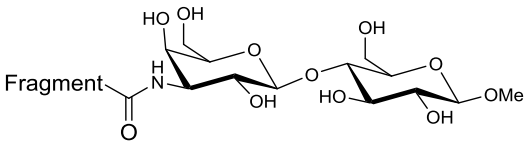


Figure 2.3.5. General structure of the novel compounds

Validation of the docking protocols. Different docking protocols were applied (described below): AutoDock4,[32] VINA[33] and Glide.[34] In order to validate the docking protocol, we docked lactose in hgal-3, obtaining results in full agreement with the X-ray crystallographic binding pose with the three docking programs AutoDock4 (RMSD = 0.727 Å), Vina (RMSD = 0.906 Å) and Glide (RMSD = 0.686 Å).

Docking with AutoDock4. The same grid box coordinates and size for all the ligands were used in the same protein. The box size chosen was big enough to ensure that all ligands could be docked. A three-dimensional grid was defined (68 x 60 x 40 Å) centered on an equidistant point to His158 and Try181. The centre of the box for each galectin was fixed as the centroid of the key residues from the CRD (Trp, His and Arg). In particular, Trp181, His158 and Arg162 were chosen for gal-3, Trp68, His44 and Arg48 for gal-1, and Trp69, His49 and Arg53 for gal-7.

The grid spacing was 0.375 Ǻ, and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock were also used. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Ǻ. The clusters were ranked by the lowest energy representative of each cluster.

Docking with VINA. The same gridboxes used with Autodock4 were employed also for VINA protocol. Twenty conformers for each ligand with each protein were generated.

Docking with Glide. The ligands were prepared by means of LigPrep tool in Maestro. Ligand minimization was performed with OPLS_2005 force field. The specified chiralities were retained as well as the ionization state. The same grid boxes in terms of size and centre were generated. Van der Waals scaling factor and partial charge cutoff were kept as default, which is 1.0 and 0.25, respectively. Unlike AutoDock4 and VINA, first a number of conformers were generated; the best poses were kept and posteriorly docked into the proteins. Standard precision Glide Docking was performed, maintaining all the default parameters except the partial charge cutoff of Van der Waals radii, which was set in 0.5, to enhance the contribution of the Trp interaction to the global binding energy. The output file was generated writing 5 poses per ligand conformation and the RMSD for the clustering was set in less than 2 Å. Calculations with Glide were only able to reproduce the crystallographic pose for lactose when two crystallographic water molecules were kept: the one close to Glu184 and Arg186, and the one close to Arg144 and Arg162. van der Waals scaling factor and partial charge cutoff were kept as default, which are 1.0 and 0.25, respectively. The other parameters were kept as default as well. The key residues from the CRD (Trp181, His158 and Arg162) were used to generate the box center position, and box size was 20x26x20 grid point, being the spacing between points of 1 Å. The output file was generated writing 5 poses per ligand conformation and the RMSD for the clustering was set in less than 2 Å.

MD simulation. The best docked solutions for each ligand were considered as starting geometries to perform the MD simulations. GLYCAM06, gaff, and ff14SB were used as force fields. All MD simulations were carried out by using the sander module in AMBER14 (www.ambermd.org). Counterions (Na+ and Cl-) were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 50 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 5000 structures for each complex, which were finally analysed.

Bibliography

1. Thijssen, V. L.; Heusschen, R.; Caers, J.; Griffioen, A. W., Galectin expression in cancer diagnosis and prognosis: a systematic review. *BBA. General subjects* **2015,** *1855* (2), 235-247.
2. Liu, F. T.; Rabinovich, G. A., Galectins as modulators of tumour progression. *Nature Reviews Cancer* **2005,** *5* (1), 29-41.
3. Hockl, P. F.; Wolosiuk, A.; Perez-Saez, J. M.; Bordoni, A. V.; Croci, D. O.; Toum-Terrones, Y.; Soler-Illia, G. J.; Rabinovich, G. A., Glyco-nano-oncology: novel therapeutic opportunities by combining small and sweet. *Pharmacol. Res.* **2016,** *109*, 45-54.
4. Farhadi, S. A.; Hudalla, G. A., Engineering galectin-glycan interactions for immunotherapy and immunomodulation. *Exp. Biol. Med.* **2016,** *241* (10), 1074-1083.
5. Arthur, C. M.; Baruffi, M. D.; Cummings, R. D.; Stowell, S. R., *Galectins: methods and protocols*. 2015; Vol. 1.
6. Saussez, S.; Cludts, S.; Capouillez, A.; Mortuaire, G.; Smetana Jr, K.; Kaltner, H.; André, S.; Leroy, X.; Gabius, H.-J.; Decaestecker, C., Identification of matrix metalloproteinase-9 as an independent prognostic marker in laryngeal and hypopharyngeal cancer with opposite correlations to adhesion/growth-regulatory galectins-1 and-7. *Int. J. Oncol.* **2009,** *34* (2), 433-439.
7. Grassadonia, A.; Tinari, N.; Iurisci, I.; Piccolo, E.; Cumashi, A.; Innominato, P.; D'Egidio, M.; Natoli, C.; Piantelli, M.; Iacobelli, S., 90K (Mac-2 BP) and galectins in tumor progression and metastasis. *Glycoconjugate J.* **2004,** *19* (7-9), 551-556.
8. Henderson, N. C.; Sethi, T., The regulation of inflammation by galectin-3. *Immunol. Rev.* **2009,** *230* (1), 160-171.
9. Newlaczyl, A. U.; Yu, L. G., Galectin-3 - a jack-of-all-trades in cancer. *Cancer lett.* **2011,** *313* (2), 123-128.
10. Campo, V. L.; Marchiori, M. F.; Rodrigues, L. C.; Dias-Baruffi, M., Synthetic glycoconjugates inhibitors of tumor-related galectin-3: an update. *Glycoconjugate J.* **2016,** *33* (6), 853-876.
11. Wang, L.; Guo, X. L., Molecular regulation of galectin-3 expression and therapeutic implication in cancer progression. *Biomed. Pharmacother.* **2016,** *78*, 165-171.
12. Ahmed, H.; AlSadek, D. M., Galectin-3 as a Potential Target to Prevent Cancer Metastasis. *Clin. Med. Insights. Oncol.* **2015,** *9*, 113-121.
13. Cousin, J. M.; Cloninger, M. J., The Role of Galectin-1 in Cancer Progression, and Synthetic Multivalent Systems for the Study of Galectin-1. *Int. J. Mol. Sci.* **2016,** *17* (9), E1566.
14. Ito, K.; Stannard, K.; Gabutero, E.; Clark, A. M.; Neo, S. Y.; Onturk, S.; Blanchard, H.; Ralph, S. J., Galectin-1 as a potent target for cancer therapy: role in the tumor microenvironment. *Cancer Metastasis Rev.* **2012,** *31* (3-4), 763-778.
15. Hughes, R. C., Galectins as modulators of cell adhesion. *Biochimie* **2001,** *83* (7), 667-676.

16. Blanchard, H.; Bum-Erdene, K.; Bohari, M. H.; Yu, X., Galectin-1 inhibitors and their potential therapeutic applications: a patent review. *Expert Opin. Ther. Pat.* **2016,** *26* (5), 537-554.
17. Blanchard, H.; Yu, X.; Collins, P. M.; Bum-Erdene, K., Galectin-3 inhibitors: a patent review (2008–present). *Expert Opin. Ther. Pat.* **2014,** *24* (10), 1053-1065.
18. St-Pierre, Y.; Biron-Pain, K.; Campion, C.; Lavoie, G.; Bouchard, F.; Couillard, J., Potential directions for drug development against galectin-7 in cancer. *Expert Opin. Drug Discov.* **2009,** *4* (6), 611-620.
19. Delaine, T.; Cumpstey, I.; Ingrassia, L.; Mercier, M. L.; Okechukwu, P.; Leffler, H.; Kiss, R.; Nilsson, U. J., Galectin-inhibitory thiodigalactoside ester derivatives have antimigratory effects in cultured lung and prostate cancer cells. *J. Med. Chem.* **2008,** *51* (24), 8109-8114.
20. van Hattum, H.; Branderhorst, H. M.; Moret, E. E.; Nilsson, U. J.; Leffler, H.; Pieters, R. J., Tuning the preference of thiodigalactoside- and lactosamine-based ligands to galectin-3 over galectin-1. *J. Med. Chem.* **2013,** *56* (3), 1350-1354.
21. Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J., Structural and thermodynamic studies on cation-Pi interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *J. Am. Chem. Soc.* **2005,** *127* (6), 1737-1743.
22. Bernardi, A.; Jiménez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K.-E.; Lahmann, M., Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev.* **2013,** *42* (11), 4709-4727.
23. *Schrödinger Release 2015-1: Maestro, version 10.1, Schrödinger, LLC, New York, NY,* 2015.
24. *Schrödinger Release 2015-1: Epik, version 3.1, Schrödinger, LLC, New York, NY,* 2015.
25. Saraboji, K.; Håkansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M., The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* **2011,** *51* (1), 296-306.
26. Choi, J. K.; Lee, B. H.; Chae, C. H.; Shin, W., Computer modeling of the rhamnogalacturonase-"hairy" pectin complex. *Proteins* **2004,** *55* (1), 22-33.
27. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S., A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** *47* (2), 279-294.
28. *Glide version 6.1, Schrödinger, LLC, New York, NY,* 2012.
29. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B., ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012,** *40* (D1), D1100-D1107.
30. Rouge, P.; Nerinckx, W.; Gough, C.; Bono, J. J.; Barre, A., Docking of chitin oligomers and Nod factors on lectin domains of the LysM-RLK receptors

in the Medicago-Rhizobium symbiosis. *Adv. Exp. Med. Biol.* **2011,** *705*, 511-521.
31. *Schrödinger Release 2015-1: MacroModel, version 10.7, Schrödinger, LLC, New York, NY*, 2015.
32. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
33. Trott, O.; Olson, A. J., AutoDock VINA: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
34. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K., Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-1749.

CHAPTER 5

TARGETING SIALIC ACID-MODIFIED RECEPTORS AS A POTENTIAL THERAPY FOR OSTEOARTHRITIS

The results here presented constitute a collaborative work with the group of Dr. María D. Mayán (CellCOM-SB Research Group) from the Instituto de Investigación Biomédica de La Coruña (INIBIC) in Spain. Dr. Mayán conceived the study and members of her group conducted the experiments and analyses (Paula Carpintero-Fernandez, Raquel Gago-Fuentes, Marta Varela-Eirin, Benigno Acea, and Eduardo Fonseca). Dr. Martín-Santamaría and Alessandra Lacetera designed, performed and analysed the computational studies. All the authors discussed the results. Dr. Goldberg, shareholder of Sentrimed Inc., acted as scientific advisor. Sentrimed Inc. supplied reagents including the lectin MASL. There is a patent to declare (International Application No.: PCT/US2014/045229). This does not interfere with the author's adherence to the journal policies on sharing data and materials.

5.1. Introduction

Osteoarthritis (OA) is the most prevalent form of arthritis and one of the most common diseases worldwide.[1-2] Nearly 10% of the adult population is affected by OA causing considerable socioeconomic cost. OA patients suffer from the degradation of articular cartilage at the synovial joints. Its prevalence and severity increase with age, but treatments only provide symptomatic relief. The only option to restore joint functionality is a total joint replacement with an artificial device. Articular cartilage functions as a cushion to cover and protect the joints between bones. The cartilage is formed by an extracellular matrix (ECM) of collagen 2 consisting of fibrils that are integrated within an abundant anionic network of proteoglycan aggregates.

These proteoglycans and collagen fibres, along with glycoproteins and water, form a macromolecular network to produce a protective matrix that facilitates articular movement and they are synthesized by chondrocytes. OA is characterized by the triggering of events that lead to the degradation of the cartilage ECM and loss of joint function.[3] The susceptibility of cartilage to arthritic degradation depends on specific posttranslational modifications of ECM proteins. Glycosylation is the most common posttranslational modification of cell surfaces and ECM proteins.[4] As a result, chondrocytes carry a dense coat of carbohydrates on their surfaces. These carbohydrate moieties mediate a wide variety of cell-cell and cell-matrix interactions that are critical for cartilage and bone development and function. Carbohydrate chains vary according to cell and tissue type and undergo modifications during a variety of processes including cellular differentiation, phenotypic changes, and oncogenesis.[5-8] The ability to regulate the events of signalling cascades to protect cartilage from the catabolic effects that induce ECM degradation will undoubtedly help to avoid or delay invasive therapeutic methods such as total joint replacement and to ameliorate the negative effects of arthritis, one the most common causes of disability that impacts over 350 million people worldwide.

In particular, sialic acids are 9-carbon carboxylated sugars that are usually found as terminal monosaccharides.[9] Sialyltransferases (SiaTs) transfer sialic acids in α-2,3-, α-2,6- or α-2,8- moieties to the *N*- or *O*-linked oligosaccharides of glycoproteins. The α-2,3- and the α-2,6-SiaTs show a mutually exclusive expression pattern and a remarkable tissue specificity. Specific sialylation motifs lead to differential effects of glycoproteins on fundamental aspects of cell behaviour including growth, migration, intercellular communication, inflammation, ECM production,[10-11] and chondrocyte function.[6, 12-13] Advances in clarifying the molecular mechanisms governing pathogenesis of OA have raised an emerging significance of glycan-protein (lectin) recognition in different aspects of pathophysiology.

Among the glycan-binding proteins, the *Maackia amurensis* seed lectin (MASL) can bind to sialylated glycoproteins,[14-16] in particular, to terminal α-2,3-sialylated oligosaccharides. MASL has been reported to be composed of two molecular species, leukoagglutinin (MAL) and hemagglutinin (MAH), although it is important to mention that a recent study has pointed to the presence of only one single species.[17]

Specific changes in lectin and glycan presentation underlie the involvement of molecular recognition aspects in the pathogenesis of several disorders. Mayán's group has undertaken an extensive analysis of the effects of the lectin MASL on primary human chondrocytes and cartilage integrity using samples from healthy donors and OA patients as well as animal models of arthritis. The results indicate that MASL preserves the structure and function of cartilage under diverse arthritic insults by interfering with the function of α-2,3-sialylated transmembrane receptors, such as the mucin-type transmembrane glycoprotein podoplanin (PDPN).[18] These findings suggest that MASL inhibits the activation of signal transduction pathways that lead to progressive cartilage destruction during the pathogenesis of arthritis by increasing reactive oxygen species (ROS), inflammatory cytokines, and metalloproteinases. We propose the direct interaction of MASL with PDPN as a plausible mechanism for this protective activity of the lectin, and our computational studies have provided a 3D molecular model for such interaction.

5.2. Results

5.2.1. Molecular binding characteristics of the interaction of PDPN with MASL

Dysregulation within the glycome of chondrocytes has been suggested as master regulators of clinically relevant inflammatory-response genes and cartilage degeneration.[6, 12, 19-21] Remarkably, previous reports indicate that osteoarthritic cartilage degeneration is promoted by factors that shift the expression of α-2,6-sialylated glycoproteins to α-2,3-sialylated glycoproteins in chondrocytes.[6, 12, 21] The results presented here indicate that MASL can target these α-2,3-sialylated glycoproteins in cartilage and protect articular chondrocytes from the detrimental effects of inflammatory and catabolic events. In particular, our data suggest that PDPN is a functionally relevant α-2,3-sialylated glycoprotein expressed by osteoarthritic chondrocytes.

PDPN is induced during oncogenesis and inflammatory processes.[18, 22-30] The PDPN receptor consists of an intracellular tail, a transmembrane domain, and an extracellular domain. While the intracellular tail can be modified by protein kinases,[24, 31] the majority of the protein consists of a highly glycosylated extracellular domain that impacts PDPN signalling and can be effectively targeted by antibodies and lectins.[32-34] In particular, the lectin MASL shows dynamic abilities to target PDPN and normalize the morphology and phenotype of tumour cells.[32-33] In addition, PDPN increases matrix metalloproteinases (MMP) activity in tumour cells,[35-37] and MASL, by targeting PDPN, blocks ECM degradation which is required for malignant cell invasion.[32-33] Results by Mayán's group indicate that MASL can protect and maintain the normal chondrocyte phenotype in the presence of damaging insults that would otherwise lead to cartilage degradation by shifting the sialylation patterns in chondrocyte glycoproteins and increasing the levels of ROS, MMPs and inflammatory cytokines (data not shown). Changes in the expression of sialyltransferases can reshape the arthritic cartilage glycophenotype reactivity contributing to activation of inflammatory pathways.

Additionally, previous studies have reported that TNF blockers inhibit PDPN expression[38] which is upregulated in synovial cells from RA patients.[25-28, 38-39] Results by Mayán's group for OA (data not shown) are consistent with these previous observations for tumour cells (regarding PDPN and MMP activity) and for RA (regarding PDPN and inflammation). Interestingly, MASL has been used in traditional medicine to treat inflammatory ailments including arthritis.[40] However, the active ingredients that underlie this purported activity have not yet been identified. In this study, an affinity chromatography-purified and commercially available lectin was used as a pure substance[32] to investigate the chondrocyte glycophenotype using various models and inflammatory conditions *in vitro* and *in vivo* (data not shown).

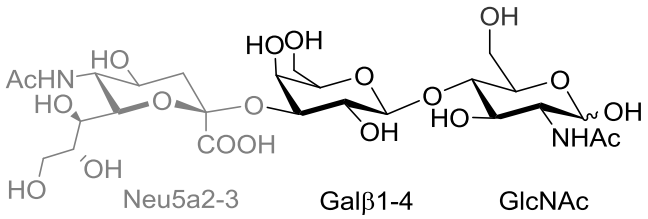
Moreover, we have provided a computational model describing the molecular recognition between MASL and the sialylated glycoprotein PDPN, identifying the relevant interactions at atomic detail. Terminal sialic acids are involved in many cellular functions, and changes in their biosynthesis or degradation are involved in degenerative disorders such as diabetes, inflammatory disorders or Alzheimer's disease by affecting ligands, masking antigenic sites, controlling signalling pathways or regulating immunological and inflammatory functions.[41-43] There is a growing interest in the targeting of catabolic and inflammatory signalling pathways for the prevention of cartilage and joint degeneration in OA and RA. Here, we focus on the lectin MASL that holds promise for drug discovery research for the treatment of arthritis. The increased levels of the α-2,3-SiaT isoforms and the corresponding increase in the levels of α-2,3-sialylated glycoproteins in osteoarthritic chondrocytes may shed mechanistic light on the pathophysiology of OA. The ability of MASL to target sialylated glycoproteins such as PDPN, which is also induced during OA and RA, and to protect chondrocytes from insults leading to cartilage degradation might offer further possibilities for therapeutic interventions and novel arthritis treatments that may include the regulation of sialylation during acute disease stages.[25-28]

5.2.2. α-2,3-sialylated glycoproteins are induced in arthritic chondrocytes

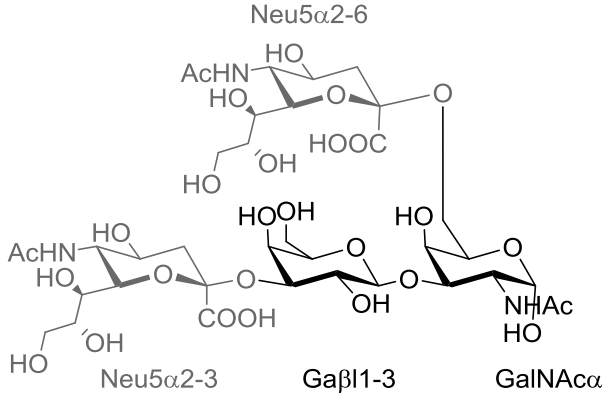
MASL was used to investigate the presence of sialic acid modifications in chondrocytes from normal and osteoarthritic articular cartilage. *In situ* analysis of surgical joint replacement samples (cartilage) from OA patients indicated that chondrocyte glycoproteins in osteoarthritic cartilage are highly α-2,3-sialylated compared with those in normal articular cartilage. Chondrocytes from healthy cartilage showed significantly reduced MASL binding, which was restricted primarily to the superficial zone. In contrast, the cartilage explants from OA patients showed strong MASL binding in the superficial and intermediate tissue zones. Expression of the α-2,3-sialylated glycoprotein PDPN receptor is induced during degenerative joint diseases including rheumatoid arthritis (RA).[25-28] In addition, PDPN plays important roles in tissue development, repair, and inflammation.[18, 23-24] For example, the binding of the *C*-type lectin-like receptor 2 (CLEC-2) to the sialylated extracellular domain of PDPN has been implicated in inflammatory reactions,[44-47] and the molecular binding characteristics for this interaction have been recently reported.[48] Because MASL has a strong affinity for these types of proteins based on their sialic acid motifs, we sought to determine if MASL recognized PDPN in arthritic chondrocytes. Higher levels of PDPN were detected in the superficial and deeper zones of osteoarthritic cartilage in comparison with cartilage from healthy donors. In addition, PDPN and MASL colocalized in human chondrocytes. Taken together, these results indicate that PDPN expression is induced in osteoarthritic chondrocytes and that MASL can target PDPN on chondrocytes.

5.2.3. PDPN and MASL: structural background for the study of sialyllactose interaction

We next sought to characterize the MASL/PDPN recognition at the atomic level using molecular modelling techniques. As previously mentioned, MASL is composed of two isolectins: MAL and MAH.[49] MAL has been reported to preferentially bind to sialyllactose or sialyllactosamine (Figure 5.1-top, Neu5Acα2-3Galβ1-4GlcNAc), typical structure of *N*-glycans, and to show a decreased affinity when the Galβ-1-4GlcNAc linkage is replaced by Galβ-1-3GlcNAc.[50-51] Conversely, MAH exhibits a higher affinity for a disialylated tetrasaccharide containing both α-2,3- and α-2,6-sialyl motifs (Figure 5.1-bottom, Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα), present in *O*-glycans. A model of the tetrasaccharide binding MAH has already been reported.[52]



Neu5Acα2-3Galβ1-4GlcNAc



Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα

Figure 5.1. 2D structure of the preferred Maackia amuresis ligands. Top: sialyllactosamine, preferred by MAL. Bottom: tetrasaccharide, preferred by MAH. Both sugars contain α-2,3 or/and α-2,6 sialyl linkages (in grey, Neu5Ac).

The 3D structure of MAL has been resolved by X-ray crystallography (PDB ID 1DBN) in complex with sialyllactose (Neu5Acα2-3Galβ1-4Glc) (Figure 5.2). MAL has 12-stranded β-sheets and contain two metal binding sites, one for the manganese and one for the calcium. The sialyllactose is located in a cavity which can accommodate the three sugar moieties. The sialic acid carboxylate establishes an ionic interaction with the ammonium group of Lys136. The glucose is perfectly accommodated between two tyrosine aromatic rings (Tyr165 and Tyr250).

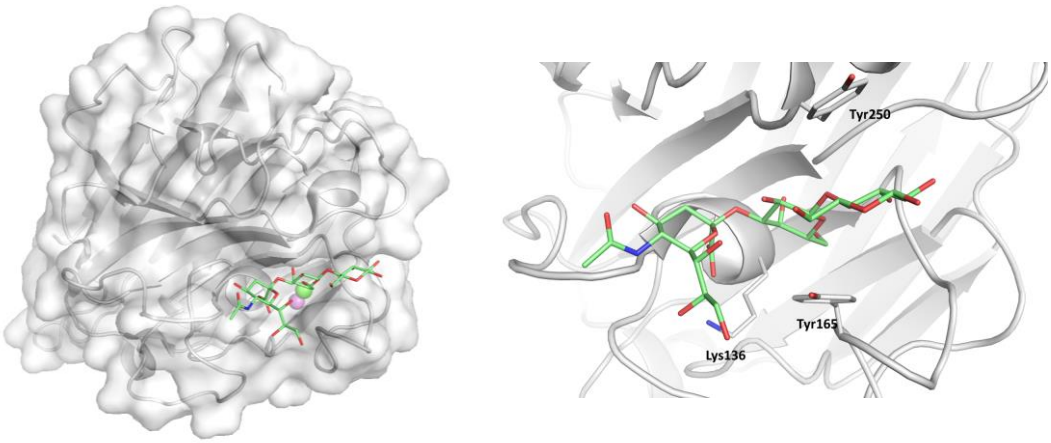


Figure 5.2. Representation of the X-ray crystallographic structure of MAL in complex with sialyllactose (PDB ID 1DBN). Sialyllactose is shown in green sticks. Left: a general view of the complex MAL-sialyllactose. MAL is represented in grey cartoon-surface and manganese and calcium are shown as violet and green spheres respectively. Right: atomic detail of the sialyllactose-MAL interaction. Two main interactions are established in this complex: a stacking interaction between the glucose of the sialyllactose and the Tyr165 and Tyr250. The carboxylate from the sialic acid establishes an ionic interaction with the Lys136 ammonium group.

Regarding the mucin-type sialoglycoprotein PDPN, it contains a highly conserved motif known as the platelet aggregation-stimulating (PLAG) domain,[53] in which a disialyl core is located at the Thr52 residue (Figure 5.3), attached by an *O*-glycosilation. Under normal conditions, this disialyl core consists of the tetrasaccharide Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα that contains both the α-2,3- and α-2,6-sialyl motifs (Figure 5.4). PDPN is a type-I transmembrane glycoprotein, with an extracellular domain rich in serines and threonines and a short intracellular domain, putative sites for cAMP phosporylation and protein C kinase. The 3D structure of this disialyl core in complex with the *C*-type lectin-like receptor 2 (CLEC-2) has been elucidated by X-ray crystallography (PDB ID 3WSR).[48]

29 EDDTETTG L EGGVAMPG A EDDVVTPG TSE 57

PLAG1 PLAG2 PLAG3

Figure 5.3. Sequence of three domains of the podoplanin (PDPN).In the third domain (PLAG 3) it is present the Thr52 (in green) which is usually glycosilated.

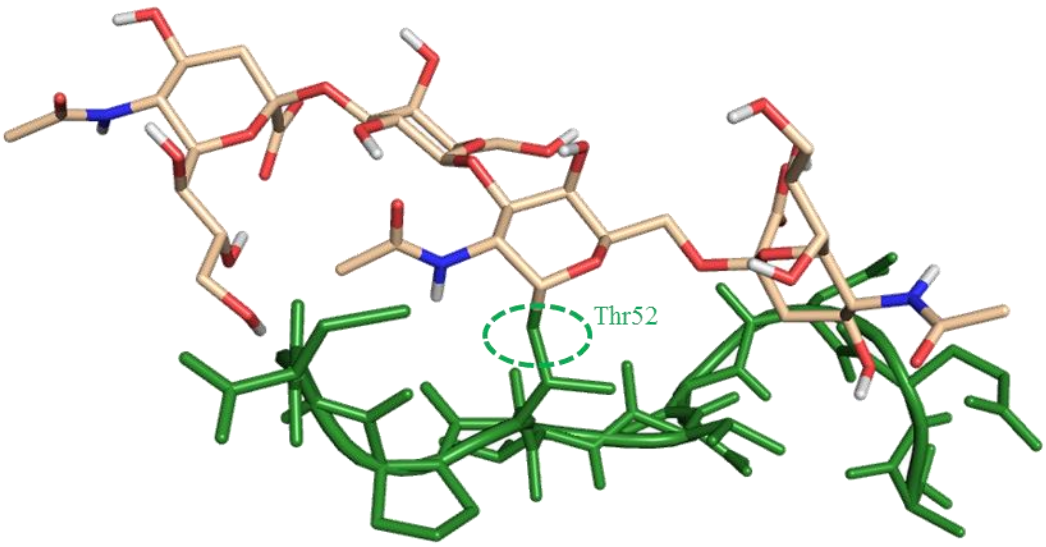


Figure 5.4. 3D structure of the extracellular glycosylated moiety of podoplanin (PDPN) adapted from PDB-ID 3WSR. The peptidic portion is shown in green. Thr52 is O-glycosylated with Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα (shown in beige).

5.2.4. Computational studies of the PDPN/MASL recognition

We were prompted to perform molecular docking of the PDPN tetrasaccharide core with the MASL structure to provide 3D clues regarding the MASL/PDPN recognition. For MAL, the X-ray crystal structure from PDB ID 1DBN was used. In the case of MAH, a homology model was obtained by using the Phyre2 web portal server[54] and MAL structure as template (see 1.3.7. section). MAH-sidechain was refined by performing MD simulations. The prediction of the binding poses of the PDPN tetrasaccharide in both MAL and MAH lectins was carried out by means of Glide[55] docking program.

All the docked poses in both proteins predicted the Neu5Acα2-3Galβ1-3GalNAcα portion to be placed in the site defined by two serines (Ser133 and Ser135 in MAL and Ser102 and Ser104 in MAH), three tyrosines (Tyr165, Tyr160 and Tyr250 in MAL and Tyr134, Tyr129 and Tyr45 in MAH), a lysine (Lys136 in MAL and Lys105 in MAH) and an aspartic acid (Asp166 in MAL and Asp135 in MAH) (Figure 5.5). Docking binding poses place the glycolic chain of the Neu5Ac2-6 moiety either exposed to the solvent or interacting with the close Tyr (Tyr165 in MAL and Tyr134 in MAH) through hydrogen bonds. In both docking studies, the carboxylic group of the Neu5Acα2-3 is involved in an hydrogen bond with the NH group of a Ser (Ser135 in MAL and Ser104 in MAH) and in a ionic interaction with the ammonium group of a Lys (Lys136 in MAL and Lys105 in MAH). The amide NH group of the Neu5Acα2-3 establishes a hydrogen bond with the OH group of Ser133 in MAL and Ser102 in MAH. Also the hydroxyl groups of the galactose moiety form several hydrogen bonds. In particular the OH groups in position 2 and 6 interact with the side chains of Tyr74 and Asp166 in MAL and Tyr45 and Asp135 in MAH respectively. In MAL, the GalNAc unit is not perfectly fitted between the Tyr250 and Tyr165, due to the steric hindrance of the Tyr250 that in MAH is an alanine (Ala219). As consequence, the binding with MAH results to be favourable. Nevertheless in MAL several hydrogen bonds can take place.

The anomeric hydroxyl group and the NH of GalNAc moiety establishes hydrogen bonds with the side chain of Tyr165 and the glycolic chain of the Neu5Acα2-6 with the side chain of Glu253. The Neu5Acα2-6 unit, instead, could point to different part of the protein in consequence to the flexibility of the 2-6 linkage. Although the predictions indicate that the binding to both lectins is possible, the theoretical binding energy was more favourable towards the MAH protein (docking score for MAL -8.431 kcal/mol anf for MAH -8.640 kcal/mol), in agreement with previously reported data.[50] Analysis of the ligand-receptor interactions showed that the recognition of the tetrasaccharide for the sialic acid α-2,3- moiety shares common characteristics for both isolectins, in particular, polar interactions involving Ser102, Ser104, and Lys105 of MAH and Ser133, Ser135, and Lys136 of MAL. These results are in agreement with mutagenesis studies with MAH in which a loss of ligand binding is shown when the Lys105 is mutated to Gly, or Asp135 to Asp.[56] Differences between MAH and MAL became apparent when analysing the anchorage of other ligand moieties aside from the sialic acid α-2,3- moiety. The Galβ-1-3 moiety is well accommodated into the MAH binding site, while the presence of Tyr250 in the MAL lectin makes this anchorage difficult. Conversely, the X-ray complex of sialyllactose and MAL (PDB ID 1DBN) shows that the alternative glycosidic linkage (the Galβ-1-4 moiety instead of the Galβ-1-3 moiety) permits binding without hindrance from the Tyr250 side chain, thus establishing CH-π interactions and a polar interaction with the Glu253 carboxylate.

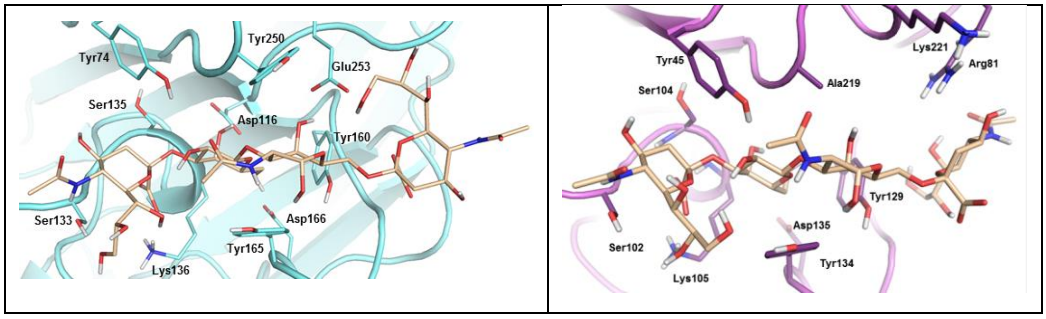


Figure 5.5. Docked tetrasaccharide (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα, shown in beige) in complex with X-ray crystallographic structure of MAL (cyan,left) from PDB ID 1DBN and the homology model of MAH (magenta,right). Labels of selected residues are depicted.

Of note, Tyr250 and Glu253 are not present in MAH, which instead holds Ala219 and Ala222. Moreover, the selectivity of the tetrasaccharide for MAH could also be explained by the presence of positively charged residues (Arg81 and Lys221) in the region harbouring the α-2,6 sialic acid interaction (Figure 5.6).

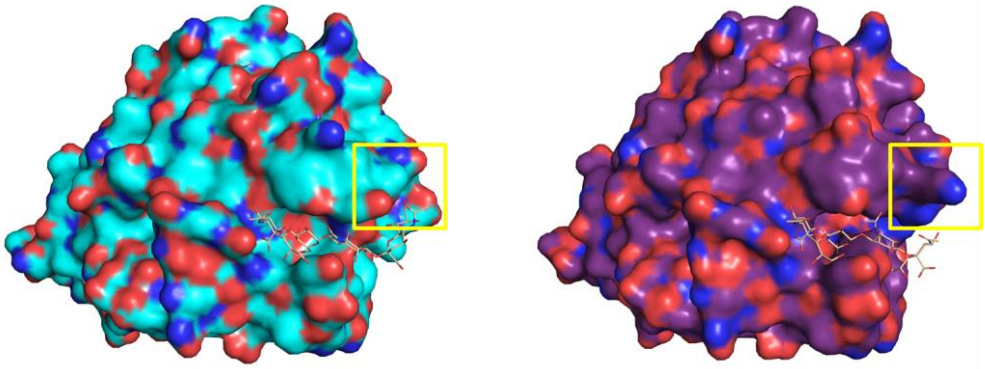


Figure 5.6. Docked tetrasaccharide (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα, shown in beige) in complex with MAL (left) and MAH (right). Proteins are represented with the electrostatic potential surface. Yellow boxes mark the most important differences in the binding site.

The side chains of these Arg81 and Lys221 in MAH residues establish hydrogen bonds interactions with the glycerol hydroxyl groups. However in MAL instead of the lysine and the arginine there are an asparagine (Asn110) and a threonine (Thr252). These two residues have side chains shorter than the residues of MAH, reason why Neu5Ac2-6 is not able to bind this region. The stability of both complexes was corroborated by 40 ns of MD simulation. Along all the simulation the complexes were stable according to the RMSD values (Figure 5.7).

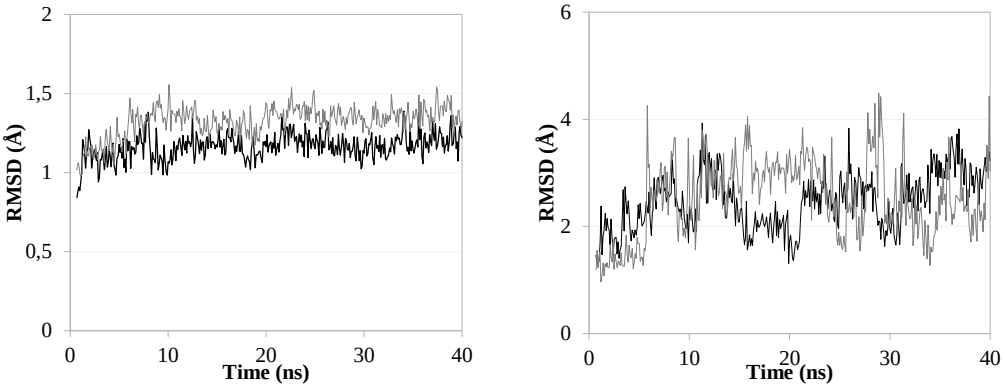


Figure 5.7. Left: RMSD of Cα along 40 ns of MD simulation. The RMSD of both, the tetrasaccharide in complex with MAL (black) and MAH (in grey) are stable along all the simulation. Right: RMSD of the heavy atom of the tetrasaccharide along the MD simulation. Also in this case the RMSD shows the stability of the ligand in complex with MAL (black) and MAH (grey). It has to be mentioned that the variation of the RMSD is due to the sugar moieties (Neu5Acα2-6) exposed to the solvent.

These results allow us to propose a molecular model for the interaction of the PDPN tetrasaccharide with the MASL isolectins MAH and MAL, in which the binding to both lectins is possible. Also, our models point towards a preferred binding towards MAH due to the establishment of stronger interactions than in the case of MAL. It is plausible that this molecular recognition mechanism involving MASL lectins could be shared by other sialylated glycoproteins such as the membrane-type metalloproteinases, integrins or CD44, providing an interesting example for the study of glycopeptide-lectin binding in cartilage.

5.3. Material and Methods

Computational modelling of the MASL proteins. MASL is a tetramer formed by a MAH protein[49] of 32 kD and a MAL protein of 37 kD. A crystal structure of MAL is available at the Protein Data Bank[57] (PDB ID 1DBN) at a resolution of 2.75 Å. Chain A was used, the co-factors of crystallization, the *N*-acetyl-D-glucosamine and the sialyllactose ligands were removed, and cap termini were inserted with the Protein Preparation Wizard within the Maestro suite 8.[58] Manganese and calcium ions were kept. All of the water molecules were removed, except those involved in the coordination of manganese and calcium. Protonation state was established by Epik, and the calculated one was maintained. The protein was minimized and charges were calculated with OPLS2005 using water as implicit solvent.[59]

The MAH homology model (Ref. Q7M1M0, UniProt) was generated with the Phyre2 web portal server[54] using MAL (Ref. P0DKL3, UniProt) as a template, with a query cover of 95% and a percentage of identity of more than 80% (Figure 5.8). The protein was prepared with the Protein Preparation Wizard, polar hydrogen ions were added, residues were maintained at a state of protonation calculated by Epik, and cap termini were added. To verify the stability of the model, we performed a molecular dynamics simulation of 5 ns using Amber12.[60] Molecular dynamics (MD) simulation of MAH was run using Amber 12. The protein was hydrated with a cubic box containing explicit TIP3P water molecules extending 10 Å away from any protein atom to simulate an aqueous environment with the help of AmberTools and counter ions were added to neutralize the system. Before the MD simulations, the system was equilibrated with the following protocol: an initial 8000 steps of steepest descent minimization, followed by heating of the system with position restraint (constant force of 20 kcal mol-1 Å-2) for all protein atoms during 10 ps of MD simulation to increase the temperature from 100 K to 300 K, plus an additional 15 ps at a constant temperature of 300 K. The position restraint was gradually decreased during 100 ps at a constant temperature of 300 K, until the entire system was under no restraints with a constant temperature (300 K) and pressure (1 atm).

After equilibration, MD simulation was run for 20ns at a constant temperature (300 K) and pressure (1 atm). Short and long-range forces were calculated every one and two time-steps, respectively (each time-step = 2.0 fs), constraining the covalent bonds involving hydrogen atoms to their equilibrium values. Long-range electrostatic interactions were accounted for by means of the particle mesh Ewald approach for applying periodic boundary conditions. The root mean square deviation (RMSD) as a function of time with respect to the starting structure for the α-C atoms was computed using cpptraj. For docking purposes, the final snapshot (after 20 ns) was taken in consideration.

tr|P0DKL3	MATSNSKPTQVLLATFLTFFFLLLNNVNSSDELSFTINNF	40
tr|Q7M1M0	SDELSFTINNF	11
tr|P0DKL3	VPNEADLLFQGEASVSSTGVLQLTRVENGQPQKYSVGRAL	80
tr|Q7M1M0	MPNQGDLLFQGVATVSPTGVLQLTSEENGQPLEYSVGRVL	51
tr|P0DKL3	YAAPVRIWDNTTGSVASFSTSFTFVVKAPNPDITSDGLAF	120
tr|Q7M1M0	YTAPVRIWDSTTGAVASFSTSFTFVVKAARG--ASDGLAF	89
tr|P0DKL3	YLAPPDSQIPSGSVSKYLGLFNNSNSDSSNQIVAVELDTY	160
tr|Q7M1M0	FLAPPDSQIPSGSVSKYLGLFNNSNSDSSNQIVAVEFDTY	129
tr|P0DKL3	FAHSYDPWDPNYRHIGIDVNGIESIKTVQWDWINGGVAFA	200
tr|Q7M1M0|Q7M1M	FGHSYDPWDPNYRHIGIDVNGIESIKTVQWDWINGGVAFA	169
tr|P0DKL3	TITYLAPNKTLIASLVYPSNQTTFSVAASVDLKEILPEWV	240
tr|Q7M1M0	TITYLAPNKTLIASLVYPSNQTSFIVAASVDLKEILPEWV	209
tr|P0DKL3	RVGFSAATGYPTEVETHDVLSWSFTSTLEANCDAATENNV	280
tr|Q7M1M0	RVGFSAATGAPKAVETHDVRSWSFTSTLEANSPADVDN	247
tr|P0DKL3	HIARYTA	287

tr|P0DKL3 = UniProt sequence for MAL
tr|Q7M1M0 = UniProt sequence for MAH

Figure 5.8. MAL and MAH sequences alignment. The alignment between the MAL and MAH sequences has been performed with the T-Coffee web server.[61] *The residue differences between the two sequences have been underlined in orange, indicating a high sequence similarity. Some key amino acids mentioned in the text are highlighted in blue (MAL: Asn110, Pro111, Tyr250, Thr252, Glu253; MAH: Arg81, Gly82, Ala219, Lys221, Ala222). Importantly, both isolectins have very similar carbohydrate binding sites. Cys272 in MAL has a critical role for homodimerization.*[32, 62]

Computational modelling of the ligand Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα. The tetrasaccharide Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα (Figure 5.1) was generated by the carbohydrate builder from GLYCAM.[63] In the X-ray crystallographic structure containing CLEC-2 (PDB ID 3WSR), this tetrasaccharide is present attached to the PDPN and the Neu5Acα2-3 unit is exposed to the solvent; for this reason, the corresponding electron densities were missing.[48] Eight conformations of the tetrasaccharide were generated by the carbohydrate builder, but only two structures fit with the X-ray crystallographic pose (PDB ID 3WSR).[48] The final structures were minimized with MacroModel[64] using the MM3* force field.

Docking calculations. A set of 122 possible conformations of the tetrasaccharide Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAcα was generated with MacroModel[64] by performing a conformational search with OPLS 2005 in the implicit solvent,[59] and the conformations were charged with the same force field. The docking was performed by means of the Glide program.[55] The prepared structures for MAL and MAH were employed as macromolecules. A cubic grid box of 25^3 $Å^3$ was generated defining the centre as the centre of mass between Tyr250, Tyr165, Tyr74, and Ser135 for MAL and between Ala219, Tyr134, Tyr45 and Ser104. A standard precision docking was applied.

MD simulation: The best docked solutions for MAH and MAL ligand were considered as starting geometries to perform the MD simulations. GLYCAM06, gaff, and ff14SB were used as force fields. All MD simulations were carried out by using the sander module in AMBER14 (www.ambermd.org). Na^+ counterions were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 40 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 5000 structures for each complex, which were finally analysed.

Bibliography

1. Carpintero Fernández, P.; Gago-Fuentes, R.; Lacetera, A.; Varela-Eirin, M.; Acea, B.; Fonseca, E.; Martin-Santamaria, S.; Mayán, M. D., FRI0031 Targeting sialic acid-modified receptors as a potential therapy for osteoarthritis. In *Ann. Rheum. Dis.*, 2016; Vol. 75, pp 437-437.
2. Chen, D.; Shen, J.; Zhao, W.; Wang, T.; Han, L.; Hamilton, J. L.; Im, H. J., Osteoarthritis: toward a comprehensive understanding of pathological mechanism. *Bone research* **2017,** *5*, 16044.
3. Loeser, R. F.; Goldring, S. R.; Scanzello, C. R.; Goldring, M. B., Osteoarthritis: a disease of the joint as an organ. *Arthritis Rheum.* **2012,** *64* (6), 1697-1707.
4. Moremen, K. W.; Tiemeyer, M.; Nairn, A. V., Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **2012,** *13* (7), 448-462.
5. Turner, G. A., N-glycosylation of serum proteins in disease and its investigation using lectins. *Clin. Chim. Acta* **1992,** *208* (3), 149-171.
6. Toegel, S.; Pabst, M.; Wu, S. Q.; Grass, J.; Goldring, M. B.; Chiari, C.; Kolb, A.; Altmann, F.; Viernstein, H.; Unger, F. M., Phenotype-related differential alpha-2,6- or alpha-2,3-sialylation of glycoprotein N-glycans in human chondrocytes. *Osteoarthr. Cartil.* **2010,** *18* (2), 240-248.
7. Schultz, M. J.; Swindall, A. F.; Bellis, S. L., Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metastasis Rev.* **2012,** *31* (3-4), 501-518.
8. Dube, D. H.; Bertozzi, C. R., Glycans in cancer and inflammation-potential for therapeutics and diagnostics. *Nat. Rev. Drug Discov.* **2005,** *4* (6), 477-488.
9. Schauer, R., Chemistry, metabolism, and biological functions of sialic acids. *Adv. Carbohydr. Chem. Biochem.* **1982,** *40*, 131-234.
10. Rao, F. V.; Rich, J. R.; Rakic, B.; Buddai, S.; Schwartz, M. F.; Johnson, K.; Bowe, C.; Wakarchuk, W. W.; Defrees, S.; Withers, S. G.; Strynadka, N. C., Structural insight into mammalian sialyltransferases. *Nat. Struct. Mol. Biol.* **2009,** *16* (11), 1186-1188.
11. Harduin-Lepers, A.; Vallejo-Ruiz, V.; Krzewinski-Recchi, M. A.; Samyn-Petit, B.; Julien, S.; Delannoy, P., The human sialyltransferase family. *Biochimie* **2001,** *83* (8), 727-737.
12. Toegel, S.; Bieder, D.; Andre, S.; Altmann, F.; Walzer, S. M.; Kaltner, H.; Hofstaetter, J. G.; Windhager, R.; Gabius, H.-J., Glycophenotyping of osteoarthritic cartilage and chondrocytes by RT-qPCR, mass spectrometry, histochemistry with plant/human lectins and lectin localization with a glycoprotein. *Arthritis Res. Ther.* **2013,** *15* (5), R147.
13. Toegel, S.; Bieder, D.; Andre, S.; Kayser, K.; Walzer, S. M.; Hobusch, G.; Windhager, R.; Gabius, H.-J., Human osteoarthritic knee cartilage: fingerprinting of adhesion/growth-regulatory galectins in vitro and in situ indicates differential upregulation in severe degeneration. *Histochem. Cell Biol.* **2014,** *142* (4), 373-388.

14. Geisler, C.; Jarvis, D. L., Effective glycoanalysis with Maackia amurensis lectins requires a clear understanding of their binding specificities. *Glycobiology* **2011,** *21* (8), 988-993.
15. Xiong, L.; Jianfang, L.; Dong, W.; Weining, W.; Zheng, C., Chromone and flavonoids from Maackia amurensis. *Asian J. Tradit. Med.* **2009,** *4* (3), 98-103.
16. Wang, W. C.; Cummings, R. D., The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha-2,3 to penultimate galactose residues. *J. Biol. Chem.* **1988,** *263* (10), 4576-4585.
17. Ochoa-Alvarez, J. A.; Krishnan, H.; Shen, Y.; Acharya, N. K.; Han, M.; McNulty, D. E.; Hasegawa, H.; Hyodo, T.; Senga, T.; Geng, J.-G., Plant lectin can target receptors containing sialic acid, exemplified by podoplanin, to inhibit transformed cell growth and migration. *PloS one* **2012,** *7* (7), e41845.
18. Astarita, J. L.; Acton, S. E.; Turley, S. J., Podoplanin: emerging functions in development, the immune system, and cancer. *Front. Immunol.* **2012,** *3*, 283.
19. Toegel, S.; Weinmann, D.; Andre, S.; Walzer, S. M.; Bilban, M.; Schmidt, S.; Chiari, C.; Windhager, R.; Krall, C.; Bennani-Baiti, I. M.; Gabius, H.-J., Galectin-1 couples glycobiology to inflammation in osteoarthritis through the activation of an NF-kappaB-regulated gene network. *J. Immunol.* **2016,** *196* (4), 1910-1921.
20. Osorio, J., Osteoarthritis: galectin-1 damages cartilage via inflammation. *Nat. Rev. Rheumatol.* **2016,** *12* (3), 132.
21. Pabst, M.; Wu, S. Q.; Grass, J.; Kolb, A.; Chiari, C.; Viernstein, H.; Unger, F. M.; Altmann, F.; Toegel, S., IL-1beta and TNF-alpha alter the glycophenotype of primary human chondrocytes in vitro. *Carbohydr. Res.* **2010,** *345* (10), 1389-1393.
22. Honma, M.; Minami-Hori, M.; Takahashi, H.; Iizuka, H., Podoplanin expression in wound and hyperproliferative psoriatic epidermis: regulation by TGF-beta and STAT-3 activating cytokines, IFN-gamma, IL-6, and IL-22. *J. Dermatol. Sci.* **2012,** *65* (2), 134-140.
23. Wicki, A.; Christofori, G., The potential role of podoplanin in tumour invasion. *Br. J. Cancer* **2007,** *96* (1), 1-5.
24. Krishnan, H.; Ochoa-Alvarez, J. A.; Shen, Y.; Nevel, E.; Lakshminarayanan, M.; Williams, M. C.; Ramirez, M. I.; Miller, W. T.; Goldberg, G. S., Serines in the intracellular tail of podoplanin (PDPN) regulate cell motility. *J. Biol. Chem.* **2013,** *288* (17), 12215-12221.
25. Del Rey, M. J.; Izquierdo, E.; Faré, R.; Miranda, V.; Criado, G.; Cañete, J. D.; Pablos, J. L., Podoplanin-Mediated Interaction Of Rheumatoid Arthritis Synovial Fibroblasts With Platelets Modulates IL-8 Expression. *Arthritis Rheum.* **2012,** *64* (Suppl 10), 1183.
26. Ekwall, A. K.; Eisler, T.; Anderberg, C.; Jin, C.; Karlsson, N.; Brisslert, M.; Bokarewa, M. I., The tumour-associated glycoprotein podoplanin is expressed in fibroblast-like synoviocytes of the hyperplastic synovial lining layer in rheumatoid arthritis. *Arthritis Res. Ther.* **2011,** *13* (2), R40.
27. Miyamoto, Y.; Uga, H.; Tanaka, S.; Kadowaki, M.; Ikeda, M.; Saegusa, J.; Morinobu, A.; Kumagai, S.; Kurata, H., Podoplanin is an inflammatory

protein upregulated in Th17 cells in SKG arthritic joints. *Molecular Immunology* **2013,** *54* (2), 199-207.
28. Faré, R.; Izquierdo, E.; Del Rey, M. J.; Usategui, A.; Celis, R.; Cañete, J. D.; Pablos, J. L., Podoplanin Expression In Rheumatoid Stroma Correlates With Lymphoid Neogenesis And Is Downregulated By Anti-TNF- Therapy. *Arthritis Rheum.* **2012,** *64* (Suppl 10), 1193.
29. Renart, J.; Carrasco-Ramirez, P.; Fernandez-Munoz, B.; Martin-Villar, E.; Montero, L.; Yurrita, M. M.; Quintanilla, M., New insights into the role of podoplanin in epithelial-mesenchymal transition. *Int. Rev. Cell Mol. Biol.* **2015,** *317*, 185-239.
30. Martin-Villar, E.; Borda-d'Agua, B.; Carrasco-Ramirez, P.; Renart, J.; Parsons, M.; Quintanilla, M.; Jones, G. E., Podoplanin mediates ECM degradation by squamous carcinoma cells through control of invadopodia stability. *Oncogene* **2015,** *34* (34), 4531-4544.
31. Krishnan, H.; Retzbach, E. P.; Ramirez, M. I.; Liu, T.; Li, H.; Miller, W. T.; Goldberg, G. S., PKA and CDK5 can phosphorylate specific serines on the intracellular domain of podoplanin (PDPN) to inhibit cell motility. *Exp. Cell Res.* **2015,** *335* (1), 115-122.
32. Ochoa-Alvarez, J. A.; Krishnan, H.; Shen, Y.; Acharya, N. K.; Han, M.; McNulty, D. E.; Hasegawa, H.; Hyodo, T.; Senga, T.; Geng, J. G.; Kosciuk, M.; Shin, S. S.; Goydos, J. S.; Temiakov, D.; Nagele, R. G.; Goldberg, G. S., Plant lectin can target receptors containing sialic acid, exemplified by podoplanin, to inhibit transformed cell growth and migration. *PLoS One* **2012,** *7* (7), e41845.
33. Ochoa-Alvarez, J. A.; Krishnan, H.; Pastorino, J. G.; Nevel, E.; Kephart, D.; Lee, J. J.; Retzbach, E. P.; Shen, Y.; Fatahzadeh, M.; Baredes, S.; Kalyoussef, E.; Honma, M.; Adelson, M. E.; Kaneko, M. K.; Kato, Y.; Young, M. A.; Deluca-Rapone, L.; Shienbaum, A. J.; Yin, K.; Jensen, L. D.; Goldberg, G. S., Antibody and lectin target podoplanin to inhibit oral squamous carcinoma cell migration and viability by distinct mechanisms. *Oncotarget* **2015,** *6* (11), 9045-9060.
34. Kato, Y.; Kunita, A.; Abe, S.; Ogasawara, S.; Fujii, Y.; Oki, H.; Fukayama, M.; Nishioka, Y.; Kaneko, M. K., The chimeric antibody chLpMab-7 targeting human podoplanin suppresses pulmonary metastasis via ADCC and CDC rather than via its neutralizing activity. *Oncotarget* **2015**.
35. Tsuneki, M.; Maruyama, S.; Yamazaki, M.; Cheng, J.; Saku, T., Podoplanin expression profiles characteristic of odontogenic tumor-specific tissue architectures. *Pathol. Res. Pract.* **2012,** *208* (3), 140-146.
36. Shindo, K.; Aishima, S.; Ohuchida, K.; Fujiwara, K.; Fujino, M.; Mizuuchi, Y.; Hattori, M.; Mizumoto, K.; Tanaka, M.; Oda, Y., Podoplanin expression in cancer-associated fibroblasts enhances tumor progression of invasive ductal carcinoma of the pancreas. *Mol. Cancer* **2013,** *12* (1), 168.
37. Inoue, H.; Miyazaki, Y.; Kikuchi, K.; Yoshida, N.; Ide, F.; Ohmori, Y.; Tomomura, A.; Sakashita, H.; Kusama, K., Podoplanin promotes cell migration via the EGF-Src-Cas pathway in oral squamous cell carcinoma cell lines. *J. Oral Sci.* **2012,** *54* (3), 241-250.
38. Del Rey, M. J.; Fare, R.; Izquierdo, E.; Usategui, A.; Rodriguez-Fernandez, J. L.; Suarez-Fueyo, A.; Canete, J. D.; Pablos, J. L., Clinicopathological correlations of podoplanin (gp38) expression in rheumatoid

synovium and its potential contribution to fibroblast platelet crosstalk. *PLoS One* **2014,** *9* (6), e99607.
39. Talmon, G.; Wake, L.; Muirhead, D., Podoplanin and clusterin are reliable markers of nonneoplastic synovium at various sites. *Int. J. Surg. Pathol.* **2013,** *21* (6), 587-590.
40. Liu, J.; Wang, L., A new medicinal plant resource-Maackia amurensis. *Chinese Wild Plant Resource* **1997,** *1*, 22-25.
41. Varki, A.; Freeze, H. H., Glycans in acquired human diseases. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
42. Freeze, H. H.; Schachter, H., Genetic disorders of glycosylation. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
43. Freeze, H. H., Genetic disorders of glycan degradation. In *Essentials of Glycobiology*, 2nd ed.; Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor (NY), 2009.
44. Bertozzi, C. C.; Schmaier, A. A.; Mericko, P.; Hess, P. R.; Zou, Z.; Chen, M.; Chen, C.-Y.; Xu, B.; Lu, M.-m.; Zhou, D.; Sebzda, E.; Santore, M. T.; Merianos, D. J.; Stadtfeld, M.; Flake, A. W.; Graf, T.; Skoda, R.; Maltzman, J. S.; Koretzky, G. A.; Kahn, M. L., Platelets regulate lymphatic vascular development through CLEC-2-SLP-76 signaling. *Blood* **2010,** *116* (4), 661-670.
45. Watson, S. P.; Herbert, J. M.; Pollitt, A. Y., GPVI and CLEC-2 in hemostasis and vascular integrity. *J. Thromb. Haemost.* **2010,** *8* (7), 1456-1467.
46. Mourao-Sa, D.; Robinson, M. J.; Zelenay, S.; Sancho, D.; Chakravarty, P.; Larsen, R.; Plantinga, M.; Van Rooijen, N.; Soares, M. P.; Lambrecht, B.; Reis e Sousa, C., CLEC-2 signaling via Syk in myeloid cells can regulate inflammatory responses. *Eur. J. Immunol.* **2011,** *41* (10), 3040-3053.
47. Suzuki-Inoue, K.; Inoue, O.; Ozaki, Y., Novel platelet activation receptor CLEC-2: from discovery to prospects. *J. Thromb. Haemost.* **2011,** *9 Suppl 1*, 44-55.
48. Nagae, M.; Morita-Matsumoto, K.; Kato, M.; Kaneko, M. K.; Kato, Y.; Yamaguchi, Y., A platform of C-type lectin-like receptor CLEC-2 for binding O-glycosylated podoplanin and nonglycosylated rhodocytin. *Structure* **2014,** *22* (12), 1711-1721.
49. Lehmann, K.; Janda, E.; Pierreux, C. E.; Ryt√∂maa, M.; Schulze, A.; McMahon, M.; Hill, C. S.; Beug, H.; Downward, J., Raf induces TGFbeta production while blocking its apoptotic but not invasive responses: a mechanism leading to increased malignancy in epithelial cells. *Genes Dev.* **2000,** *14* (20), 2610-2622.
50. Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T., Strong affinity of Maackia amurensis hemagglutinin (MAH) for sialic acid-containing Ser/Thr-linked carbohydrate chains of N-terminal octapeptides from human glycophorin A. *FEBS Lett.* **1994,** *342* (3), 334-338.
51. Knibbs, R. N.; Goldstein, I. J.; Ratcliffe, R. M.; Shibuya, N., Characterization of the carbohydrate binding specificity of the

leukoagglutinating lectin from Maackia amurensis. Comparison with other sialic acid-specific lectins. *J. Biol. Chem.* **1991,** *266* (1), 83-88.
52. Imberty, A.; Gautier, C.; Lescar, J.; Pérez, S.; Wyns, L.; Loris, R., An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **2000,** *275* (23), 17541-17548.
53. Kaneko, M. K.; Kato, Y.; Kameyama, A.; Ito, H.; Kuno, A.; Hirabayashi, J.; Kubota, T.; Amano, K.; Chiba, Y.; Hasegawa, Y.; Sasagawa, I.; Mishima, K.; Narimatsu, H., Functional glycosylation of human podoplanin: glycan structure of platelet aggregation-inducing factor. *FEBS Lett.* **2007,** *581* (2), 331-336.
54. Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J., The Phyre2 web portal for protein modeling, prediction and analysis. *Natural Protocols* **2015,** *10* (6), 845-858.
55. *Glide version 6.1, Schrödinger, LLC, New York, NY*, 2012.
56. Yamamoto, K.; Ishida, C.; Saito, M.; Konami, Y.; Osawa, T.; Irimura, T., Cloning and sequence analysis of the Maackia amurensis haemagglutinin cDNA. *Glycoconjugate J.* **1994,** *11* (6), 572-575.
57. Protein Data Bank. http://www.rcsb.org (accessed March 29, 2017).
58. *Maestro version 9.3, Schrödinger, LLC, New York, NY*, 2012.
59. Levkowitz, G.; Waterman, H.; Ettenberg, S. A.; Katz, M.; Tsygankov, A. Y.; Alroy, I.; Lavi, S.; Iwai, K.; Reiss, Y.; Ciechanover, A.; Lipkowitz, S.; Yarden, Y., Ubiquitin ligase activity and tyrosine phosphorylation underlie suppression of growth factor signaling by c-Cbl/Sli-1. *Mol. Cell.* **1999,** *4* (6), 1029-1040.
60. *AMBER 12*, D.A. Case, T.A. Darden, 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. Götz, 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.
61. Notredame, C.; Higgins, D. G.; Heringa, J., T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000,** *302* (1), 205-217.
62. Kaku, H.; Mori, Y.; Goldstein, I.; Shibuya, N., Monomeric, monovalent derivative of Maackia amurensis leukoagglutinin. Preparation and application to the study of cell surface glycoconjugates by flow cytometry. *J. Biol. Chem.* **1993,** *268* (18), 13237-13241.
63. Glycam Carbohydrate Builder. http//glycam.ccrc.uga.edu.
64. MacroModel version 10.2, Schrödinger, LLC, New York, NY. **2013**.

CHAPTER 6

GLYCOMIMETICS TARGETING GLYCOSYLTRANSFERASES: SYNTHETIC, COMPUTATIONAL AND STRUCTURAL STUDIES OF LESS-POLAR CONJUGATES

The results here presented constitute a collaborative work with the group of Prof. Pedro Merino (Universidad de Zaragoza) and Dr. Ramón Hurtado-Guerrero (BIFI, Universidad de Zaragoza) in Spain. The synthesis of the molecules was carried out at Merino´s laboratory and X-ray crystallographic studies were performed at Hurtado-Guerrero´s laboratory. Dr. Martín-Santamaría, Alessandra Lacetera and Mattia Ghirardello performed and analysed the computational studies at the CIB-CSIC.

6.1. Introduction

Glycosyltransferases (GTs) are key enzymes responsible of the incorporation of carbohydrates into a variety of acceptor biomolecules, including proteins, lipids, oligosaccharides and different metabolites.[1-3] The resulting glycoconjugates mediate a wide range of functions from structure and storage to signalling and, in consequence, they are related with important diseases. Therefore, chemical manipulation of GTs activity could lead to the development of useful therapeutic drugs.[4] Because of it, considerable synthetic efforts have been directed toward the preparation of efficient GT inhibitors.[5-11] Transfer of the sugar residue occurs from an anionic nucleotide sugar donor to the acceptor substrate; it can take place with retention or inversion of configuration at the anomeric centre of the sugar residue.[12] In this context, the elucidation of the mechanisms used by GTs has been pursued with much attention.[13-18] Of note, only nine sugar donors are known to be involved in protein glycosylation in mammal organisms,[19] which is the most abundant post-translational modification in nature. Six of these sugar donors contain the uridine moiety (Figure 6.1) that is in line with the existence of GTs employing UDP sugars as the most predominant in nature.[20]

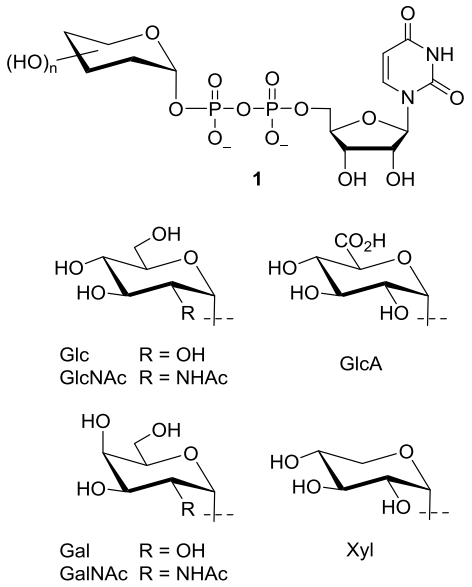


Figure 6.1. Uridine-based sugar donors.

Most of designed inhibitors mimic nucleotide phosphate sugars by incorporating anionic groups to emulate binding of the diphosphate bridge.[21-24] However, such compounds are not capable of permeating into cells due to their high polarity.[25] In order to overcome this problem, neutral inhibitors have also been prepared,[26-31] and derivatives in which the phosphate group was replaced by a different apolar group showed enhanced cell internalization[30-31] even though some reduction in binding affinity might be found. Vocadlo and co-workers demonstrated that UDP-5SGlcNAc acts as an inhibitor of OGT[32] but it should be generated inside the cell from precursor 5SGlcNAc by using the salvage pathway. Yet, most of the designed compounds resulted in poor biological activity because they do not incorporate the nucleoside moiety loosing recognition and selectivity by the target enzymes.

Design, preparation and binding studies of less polar uridyl-sugar analogues **6** have been performed as suitable GT binders in which the β-phosphate unit has been replaced by an alkyl chain (Figure 6.2).

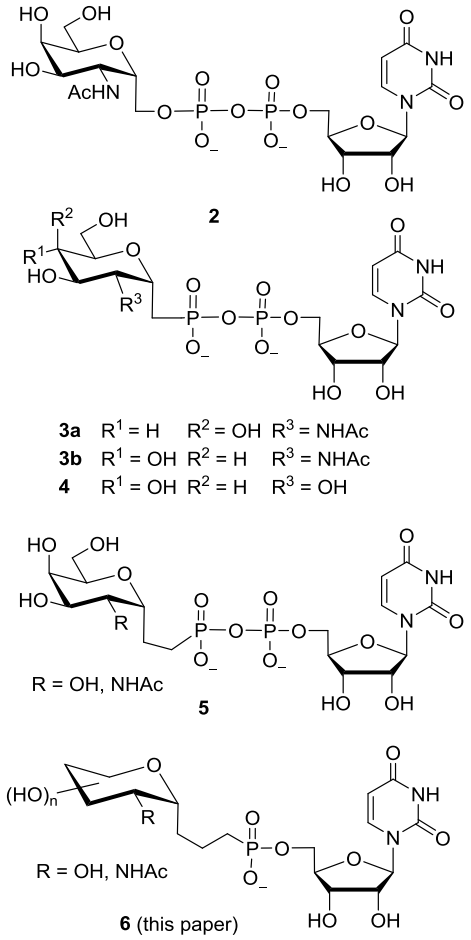


Figure 6.2. C-analogues of uridine dinucleotide sugars.

Insertion of methylene (**2**)[33] and replacement of the anomeric oxygen by a methylene (**3**)[34-38] or propylene (**4**)[39] groups has already been reported, but compounds **2-4** are still too polar due to the presence of two phosphate units (Figure 6.2). Compound **3a** has been evaluated against three bacterial UDP-galactopyranose mutases and showed only moderate inhibition although crystallographic studies showed that it is bound to the active site of the enzyme in a novel conformation not observed previously.[37] As model enzyme, we have selected GalNAc-T2, a member of important *N*-acetylgalactosaminyl transferases (GalNAcTs E.C. 2.4.1.41) involved in mucin biosynthesis.[40-41] In particular, a previous computational study with GalNAc-T2 reported by Masgrau and co-workers anticipated that UDP-2'-deoxyGal should not be a good donor substrate whereas UDP-Gal could be a valid sugar donor in some cases.[42]

Based on these precedents, analogues **6** have been designed, where one phosphate group has been replaced by an ethylene group, as suitable GT binders, exemplified on GalNAc-T2 as model enzyme. The synthetic strategy is based on coupling between uridine and phosphoalkyl sugars, which are prepared through radical hydrophosphonylation[43] of allyl sugars.

Through tryptophan fluorescence spectroscopy and X-ray studies have been demonstrated that two of these compounds bind moderately to the enzyme. In particular, trapping of a galactose moiety-containing compound bound to an inactive form of GalNAc-T2 provides hints for improving their affinity by the incorporation of the β-phosphate or alike negative groups. Furthermore, *in silico* studies support the aforementioned findings and as well as that the presence of only one phosphate group close to the sugar moiety could be enough for the uridyl-sugar analogue to interact with the protein. These combined studies also provide new evidences of the GalNAc-T2 catalytic mechanism cycle.

6.2. Results

6.2.1. Crystallographic and binding studies

In order to validate whether our compounds might have a biological effect, we selected the glycosyltransferase GalNAc-T2 as an exemplar. This enzyme is part of a large family of 20 members that transfers a GalNAc moiety from UDP-GalNAc onto Ser or Thr residues of proteins in the presence of manganese.[44] This enzyme predominantly uses UDP-GalNAc as the sugar donor though in some cases it might also use UDP-galactose.[42,44]

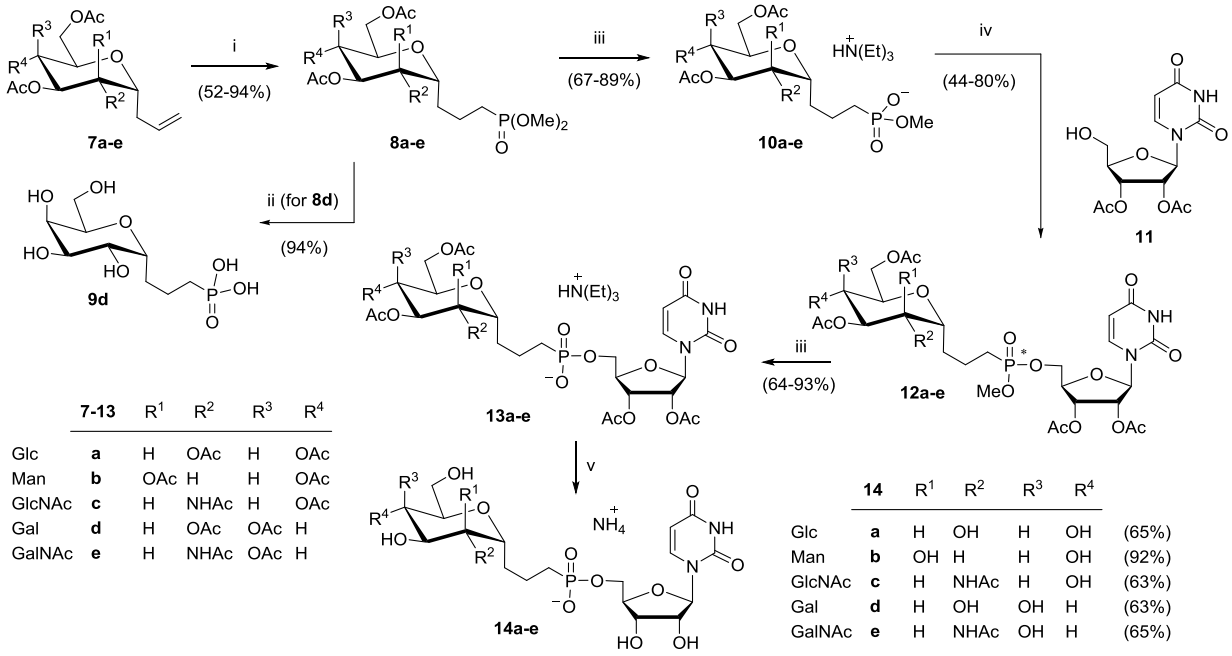


Figure 6.3. Scheme of the synthesis of partially free phosphonates.

Prompted by this, we determined whether the compounds **14d** and **14e** (Figure 6.3), that contains a Gal and GalNAc moiety, respectively, might bind to this enzyme. Tryptophan fluorescence spectroscopy studies for compound **14d** rendered K_d's of 269 ± 40 and 254 ± 20 μM, and for compound **14e** rendered K_d's of 800 ± 10 and 915 ± 180 μM, in the absence and presence of 2 mM Mn^{+2}, respectively, suggesting that the interaction with Mn^{+2} for these ligands might not be crucial for their binding (Figure 6.4). Contrary to GalNAc-T2 preferences for UDP-GalNAc as the favourite donor substrate for catalysis, the results also suggest that the compound with the galactose moiety (**14d**) was ca. 3.5-fold better binder than the one with the GalNAc moiety (**14e**).

Compound **14d** traps the enzyme in an inactive state. To understand the binding mode of these compounds to GalNAc-T2, we solved the crystal structure of GalNAc-T2 in complex with compound **14d**. Despite numerous attempts were carried out with both compounds (**14d** and **14e**) by co-crystallisation and soaking experiments, we only managed to get the complex with **14d** by soaking experiments in orthorhombic crystals previously grown with UDP/Mn^{+2}. The structure was solved at 2.07 Å, allowing us to solve and interpret the density maps (see Table 6.A1 in Annex VI).

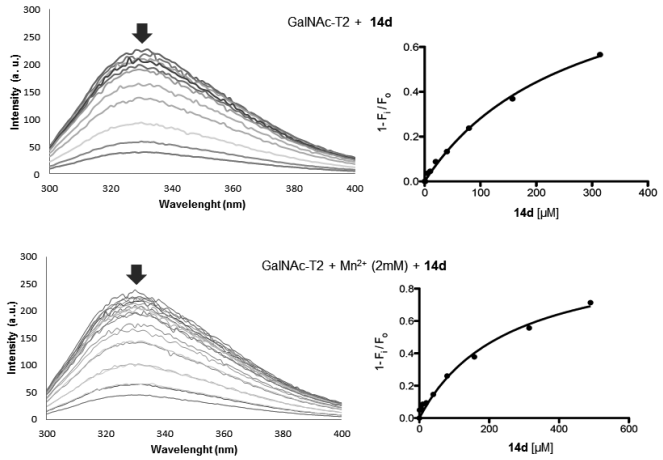


*Figure 6.4.Quenching of intrinsic GalNAc-T2 tryptophan fluorescence measured at increasing concentrations of the compound **14d** in the absence and in the presence of Mn^{+2}. All data points represent the means ± S.D. for three measurements. The K_d for compound **14d** was determined by fitting fluorescence intensity data against compound **14d** concentration.*

The asymmetric unit as was reported earlier displays 6 molecules of GalNAc-T2.[44] Three out of six GalNAc-T2 molecules present in the asymmetric unit showed clearly density maps for the presence of both UDP/Mn^{+2} and compound **14d** in each active site, while the other three remaining GalNAc-T2 molecules contained UDP (Figure 6.5). The crystal structure shows that the typical GT-A fold (catalytic domain) is located in the *N*-terminal region and the lectin domain is located in the *C*-terminal region (Figure 6.6).

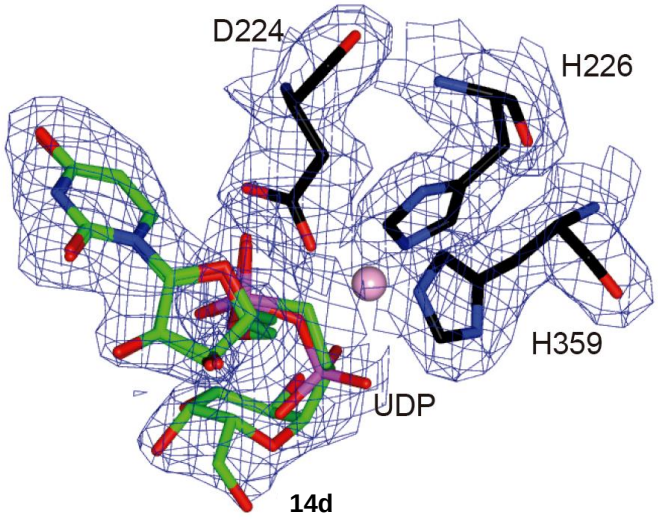


*Figure 6.5. Compound **14d** electron density map. Electron density maps are F_O–F_C syntheses (blue) contoured at 2.2 σ for UDP/**14d**, and 2F_O–F_C syntheses (blue) contoured at 1.0 σ for the residues Asp224, His226 and His359. The manganese atom is shown as a pink sphere.*

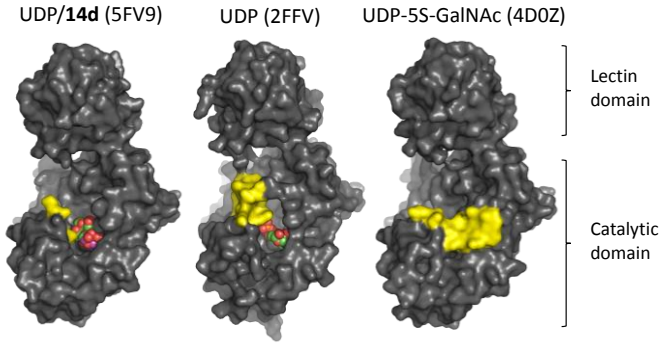


*Figure 6.6. Surface representation of GalNAc-T2 in complex with compound **14d**/UDP, UDP and UDP-5S-GalNAc. Protein, flexible loop, nucleotides/compound **14d** are coloured in grey, yellow and green, respectively. The active and inactive states correspond to closed (4D0Z) and open (5FV9 and 2FFV) conformations, respectively.*

Strikingly, GalNAc-T2 in complex with compound **14d** adopts an inactive state conformation, which has been reported previously for other binary complexes containing either glycopeptides or UDP (Figure 6.6).[40, 44-45] Both, active and inactive states have been found to be determinant during the catalytic cycle of this enzyme and are associated to the motion of a flexible loop that can oscillate between closed and open conformations, respectively.

In particular, UDP and compound **14d** are exposed to the solvent due to the open conformation of the flexible loop, whereas UDP-5S-GalNAc is covered by the flexible loop, which adopts a closed conformation, rendering the enzyme in an active state (Figure 6.6, PDB ID 4D0Z). Compound **14d** induces an inverted-uridine conformation and an unusual 4-coordinate Mn^{+2} complex. A closer inspection of the active site reveals that the uridine moiety of both compound **14d** and UDP adopts an unusual inverted conformation, which was also found earlier in a structure of this enzyme in complex with UDP (Figure 6.7 and PDB ID 2FFV).[45] This atypical conformation was proposed to represent a final step during the catalytic cycle of GalNAc-T2 in which UDP was ready to exit the enzyme.[44-45] One major difference stands out between both complexes: the metal is unusually 4-coordinate in our crystal structure, and 5-coordinate in the GalNAc-T2-UDP complex (PDB ID 2FFV; Figure 6.7). The metal in the former structure is coordinated by the α-phosphate, Asp224, His226 and His359 (Figure 6.7, A3), whereas in the latter complex, the metal is additionally coordinated by the UDP β-phosphate (Figure 6.7, B1 and B2). However, 6-coordinate Mn^{+2}, which requires the previous residues, pyrophosphate and an extra water molecule to form an octahedral complex, is not only the most common coordination in crystal structures of this enzyme but also the most abundant in proteins in general.[40, 44-46] These results suggest that these UDP-phosphonate scaffolds, with only the alpha-phosphate, are weaker ligands in terms of binding to GalNAc-T2 but they are still suitable binders for this enzyme. Previously, different crystal structures obtained from the same type of orthorhombic crystals suggested that UDP might be present in both active and inactive states of the enzyme. On the contrary to our structure, UDP in these crystals always adopted the typical conformation found for UDP-GalNAc, which was required for an efficient turnover.[44] Thus, it is tempting to speculate that compound **14d** has allowed us to trap a conformation at a very final step of the catalytic cycle of this enzyme in which UDP would be ready for an immediate departure from the enzyme.

The crystal structure also supports the aforementioned K_d's, confirming that compound **14d** does not rely on Mn^{+2} for binding to GalNAc-T2. Regarding direct interactions between the enzyme and the nucleotides/compound **14d** in the different complexes, it is important to mention that compound **14d** shows less-interactions than UDP (PDB ID 2FFV) and UDP-5S-GalNAc (PDB ID 4D0Z) (Figure 6.7).

It is important to note that Trp331, a critical residue in the catalytic cycle of these enzymes,[45] is adopting an "out" (out of the active site) conformation in the complex with **14d/UDP** (Figure 6.7, A1) which correlates well with the inactive state of GalNAc-T2. On the contrary, Trp331 adopts an "in" (inside the active site) conformation in the complex with UDP-GalNAc that sets up the enzyme in an active state (Figure 6.7, C1).

Most of the direct interactions between the uridine moiety of compound **14d** and the enzyme are kept as the UDP from the reported PDB ID 2FFV, whereas a larger number of direct interactions are established between UDP-GalNAc and the enzyme, presumably because the active site is covered by the flexible loop (Figure 6.6-6). Finally, the phosphonate of compound **14d** does not interact directly with any residue and the galactose moiety is recognised exclusively by Glu334 (Figure 6.7, A2), likely explaining the poor binding of this compound to the enzyme. In fact, in the docking studies, we found that the Gal moiety adopts several alternative binding poses within the sugar site.

6.2.2. Docking Studies

In order to have a 3D picture of the putative binding ability of this family of compounds, docking calculations were performed for analogues **14** in both conformations of GalNAc-T2: closed (active, PDB ID 4D0Z) and open (inactive, PDB ID 2FFV). The calculations were able to predict reasonable docked binding poses for all the ligands, although with some slight differences. UDP, UDP-GalNAc, and UDP-5S-GalNAc were also docked for comparison purposes in order to validate the computational protocol. Interestingly, all the compounds were predicted to bind in both conformations (open and closed) inside the active site of the enzyme with binding poses similar to those found in the crystallographic structures. For the subsequent studies, the best predicted binding poses in terms of binding energies from Glide were taken into consideration.

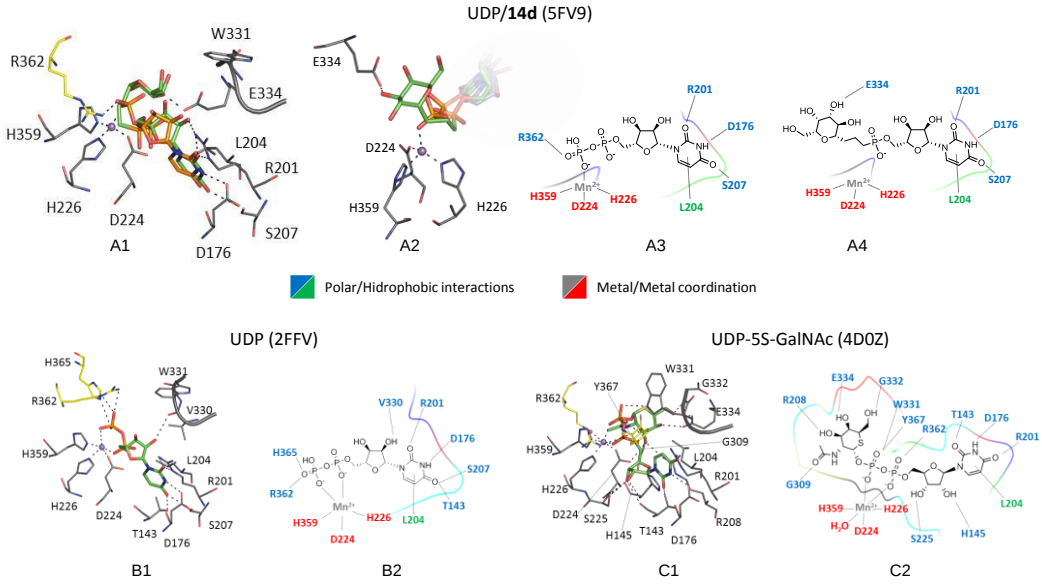


*Figure 6.7. Structural features of the sugar nucleotide binding site for the GalNAc-T2 in complex with UDP/**14d**. Panel A1 shows a close-up view of the active site in complex with UDP-Mn^{+2}/**14d**. Panel A2 shows a detail on metal interactions. Panels A3 and A4 depict a 2D scheme of the interactions between **14d** and UDP, respectively, and the residues of GalNAc-T2. Panels B1 and C1 show a close-up view of the active site in complex with UDP/Mn^{+2} and UDP-5S-GalNAc-Mn^{+2}, respectively. Panels B2 and C2 depict a 2D scheme of the interactions between the ligands and the residues of GalNAc-T2. The nucleotides are depicted as green carbon atoms with the exception of UDP in the crystal structure with compound **14d**, which is depicted with orange carbon atoms. Protein residues of the active site and the flexible loop are illustrated as grey and yellow carbon atoms, respectively. The manganese atom and a water molecule are shown as lilac and red spheres, respectively. Hydrogen bonds and protein/UDP-Mn^{+2} coordination are depicted as black dashed lines.*

Predicted binding energies were higher in comparison to the reference diphosphate compounds, due to the consequently decreased ionic phosphate-manganese interaction caused by the lack of the β-phosphate group (Table 6.1).

Table 6.1. Values of predicted binding energy from Glide docking

Predicted binding energy with 2FFV (in kcal/mol)				**Predicted binding energy with 4D0Z (in kcal/mol)**			
UDP	**UDP-GalNAc**	**14d**	**14e**	**UDP**	**UDP-GalNAc**	**14d**	**14e**
-6,642	-6,217	-6,57	-4,075	-8,651	-9,41	-6,73	-8,511

Even though, docked binding energies were favourable, thus validating the phosphonate scaffold of analogues **14** as a possible platform for the optimization of GTs binders.

The best binding energies were always predicted for the active (closed) conformation, in agreement with a binding pose with more GalNAc-T2/ligand interactions, although the docking results pointed to the ability to bind also to the open conformation, which in fact, is preferred in the crystal structure. Probably, both possibilities might co-exist in the equilibrium of the conformational ensembles. Docking calculations are able to estimate both binding poses while the limitations of X-ray, in this case, have not allowed us to obtain the closed 3D structure. No main differences in the predicted binding energy were found among the phosphonate analogues **14**, pointing to a lack of preferred binding. For compounds **14,** the main contributions to the binding are provided by the uridine-enzyme H-bonds as in the 4D0Z complex, and, to a lesser extent, from the ionic phosphonate-Mn^{+2} interaction through one or two of the phosphonate oxygens.

In the case of the active conformation of the GT, an interesting difference was found regarding the binding of the sugar moiety of analogues **14**, which also showed alternative binding poses where the sugar moiety reaches Ala307 CO group and Asn146 side chain, according to the highly flexible aliphatic chain. For the inactive conformation of GalNAc-T2, the docking calculations placed the synthetic analogues in the equivalent region than in the case of the active conformation (Figure 6.8).

The uridine moiety is observed to rotate in a similar way to the crystallographic pose, and the overall conformation of the ligand was found to adopt a more bended conformation, with the sugar moiety placed in front of the uridine plane. For comparison purposes we performed docking calculations for **14d** to our crystallographic structure (PDB 5FV9, chain "B") finding similar results between the crystallographic pose for **14d** and the docked pose for UDP-GalNAc.

We conclude, from the overall docking study, that compounds **14** are putative GalNAc-T2 binders where the uridine moiety plays the major role in the binding in agreement with the crystallographic binding poses, while the phosphonate group and the sugar moiety provide additional interactions, not finding differences among the type of sugar.

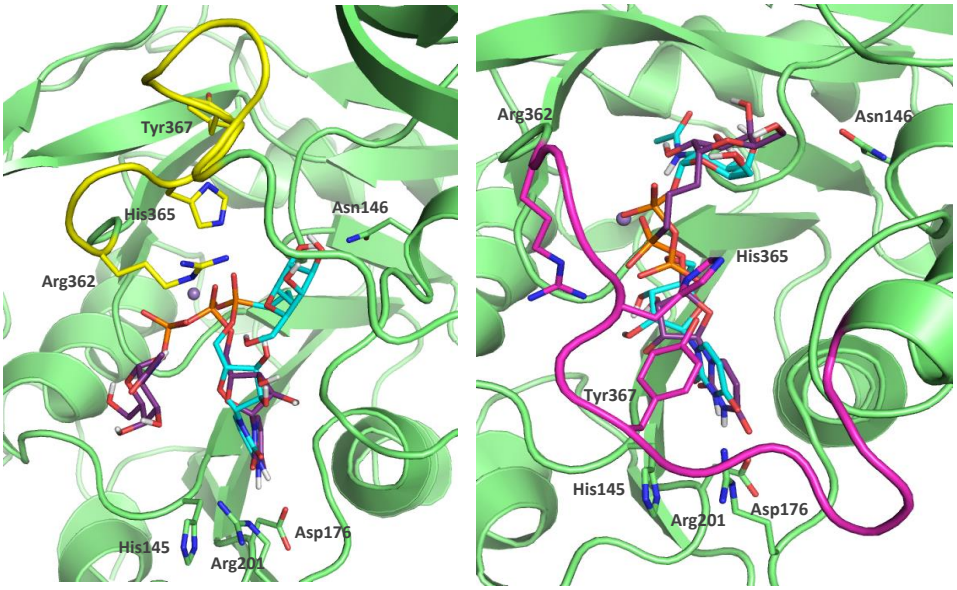


*Figure 6.8. Left panel: Superimposition of docked poses in GalNAc-T2 (PDB-ID 2FFV) of compound **14d** (cyan) and UDP-GalNAc (purple). Open loop is shown in yellow. Right panel: Superimposition of docked poses in GalNAc-T2 (PDB-ID 4D0Z) of compound **14d** (cyan) and UDP-GalNAc (purple). Closed loop is shown in magenta.*

6.2.3. tMD simulation

In order to shed light into the conformational changes that the protein needs to suffer from the inactive to the active conformation (opening/closing process), and to analyze the required interactions to host the substrate, targeted molecular dynamics (tMD) simulation was undertaken for the GalNAc-T2/UDP-GalNAc complex. The simulation was started from the complex in the inactive open conformation (from the docked binding pose) towards the complex in the active closed conformation (from the docked binding pose of UDP-GalNAc, very similar to the crystallographic complex with UDP-5S-GalNAc) (Figure 6.7). The analysis of the root mean square fluctuation (RMSF) (Figure 6.9) during the tMD simulation time identified a main fluctuation corresponding to the flexible loop Arg362-Ser373 with a maximum value for Pro370. This main change corresponds to the folding of the loop over the substrate, sealing it in the active site, and leading to the closed conformation. Additionally to Asp224, His226 and His359 side chains, Mn^{+2} ion was found to coordinate with two phosphate oxygens (Figure 6.10). During the progression of the tMD simulation, one water molecule entered into the coordination sphere, according to a 6-coordinate Mn^{+2} model (Figure 6.10).

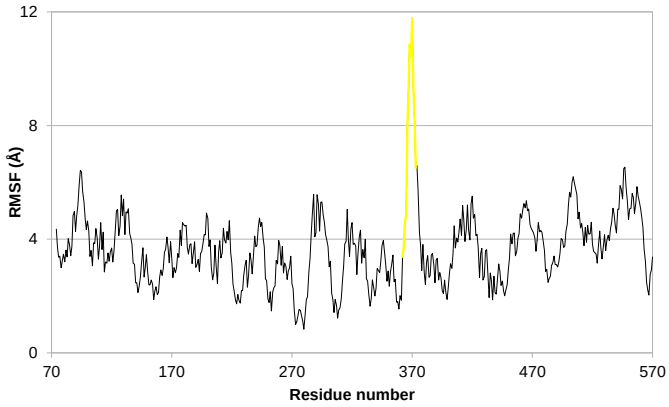


Figure 6.9. Root mean square fluctuation (RMSF) for the Cα atoms for GalNAc-T2 in complex with UDP-GalNAc during the tMD simulation. Values corresponding to the flexible loop Arg362-Ser373 are marked in yellow.

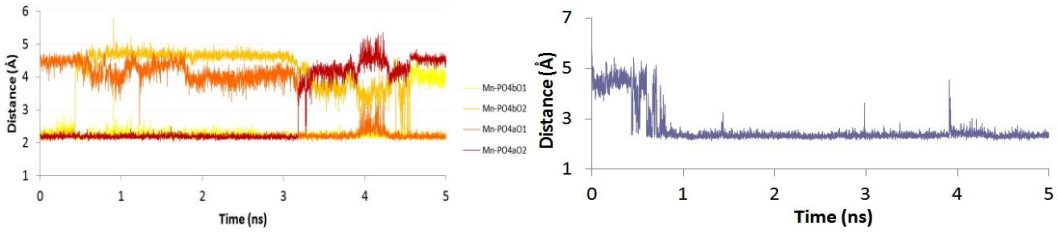


Figure 6.10. Left: Distances between the Mn^{+2} ion and the phosphate oxygens (alpha -a- and beta -b-) along the tMD simulation time. Right: Distances between the Mn^{+2} ion and one water molecule along the tMD simulation time entering into the coordination sphere.

Several functional groups from the GalNAc-T2 are establishing interactions, especially those involving the uridine moiety. In particular, the NH in position 3 of the uracyl ring establishes a stable hydrogen bond along the simulation with the Asp176 side chain (Figure 6.11-right and Figure 6.12). The CO groups at positions 2 and 4 of the uracyl ring are establishing alternating hydrogen bonds with Arg201 and Thr143 side chains (Figure 6.11-right and Figure 6.13-13) during the rotation of the uridine. This rotation is allowed for the torsional change undergone by the diphosphate linker. Also stacking interactions between the His145 side chain and the uracyl moiety are identified. Most of these interactions are always found in the crystallographic structures in agreement with a specific anchorage for the uracil moiety. On the contrary, the flexibility of the diphosphate linker allows the GalNAc to rotate and move in the closing process, changing the interactions along the simulation, adapting to the new surrounding environment.

For example, in the inactive conformation, the GalNAc is involved in two hydrogen bonds: one between the endocyclic GalNAc oxygen and the Arg362 backbone NH, and another one between the *N*-acetylamido CO group and the Arg362 guanidium group. After the closing, these interactions with the Arg362 are lost, and new hydrogen bonds are established: one between the GalNAc OH-3 and the backbone NH of Gly309; and another one between the OH-4 from the GalNAc moiety and the Ala307 CO group. Moreover new CH-π interactions are formed between the ribose and the Tyr367 indole ring.

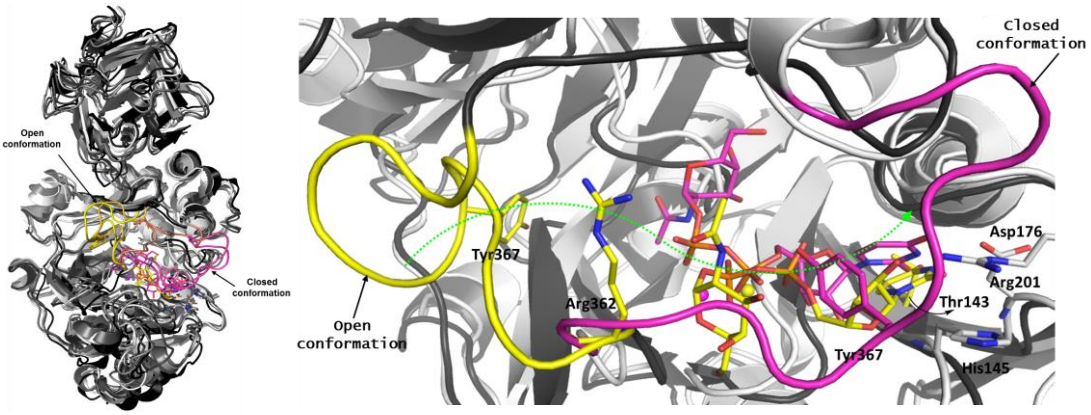


Figure 6.11. Superimposition of five structures of the GalNAc-T2/UDP-GalNAc complex along the tMD simulation of the transition from the open (black, loop in yellow) to the closed (white, loop in magenta) conformation. Intermediate protein structures are shown grading from yellow (starting geometry), orange (intermediate geometries) to magenta (final target geometry). Ligand UDP-GalNAc is shown in yellow for the open complex, and in magenta for the closed complex. Left: Full perspective. Right: Detail showing some representative residues.

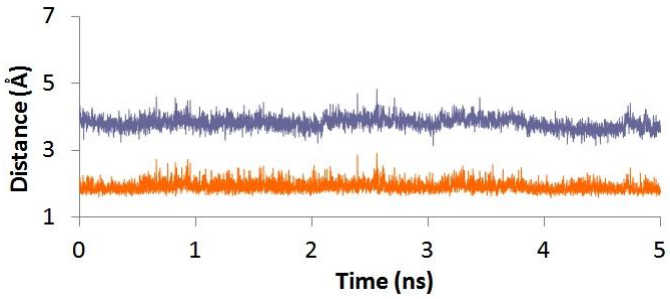


Figure 6.12. Distance between the two carboxylate oxygens (each distance is represented in orange and blue colours) from Asp176 side chain and the NH in position 3 from the uracyl along the tMD simulation time.

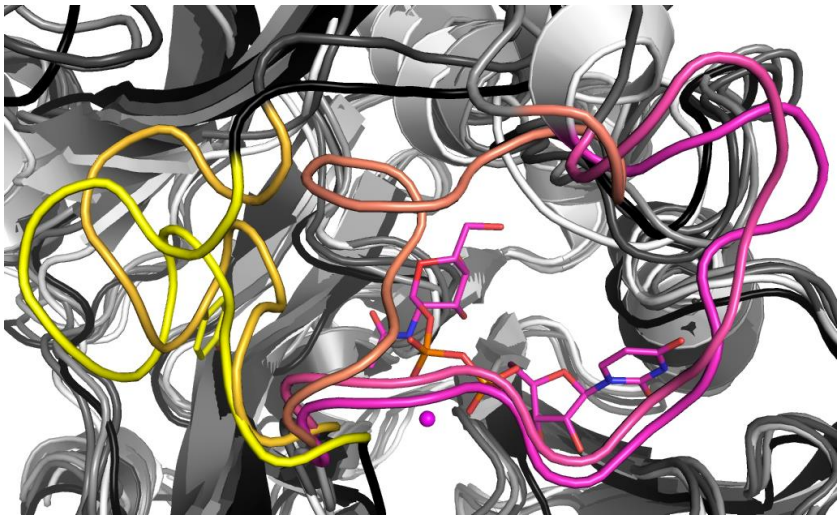


Figure 6.13. Superimposition of five structures of the GalNAc-T2/UDP-GalNAc complex along the tMD simulation of the transition from the open (black, loop in yellow) to the closed (white, loop in magenta) conformation. Intermediate structures are shown grading from yellow (starting geometry), orange (intermediate geometries) to magenta (final target geometry). Ligand UDP-GalNAc and Mn^{+2} ion are shown in magenta for the closed complex.

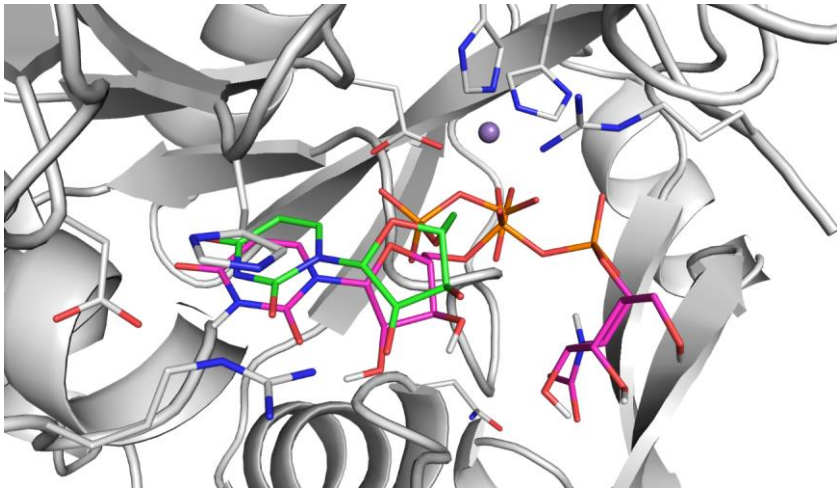


Figure 6.14. Superimposition of docked pose in GalNAc-T2 (PDB-ID 5FV9) of UDP-GalNAc (magenta). Crystallographic pose of UDP is shown in green as reference.

Summarizing, the tMD may give us an approximation to the binding pose changes that the UDP-GalNAc suffers in the closing/opening process when bound to the GT. The specific interactions for the GalNAc in the closing/opening process may be useful to understand the catalytic mechanism cycle of this enzyme, and to help for the designing of selective GT binders. In fact, we provide a plausible explanation for the role of the *N*-acetylamido group in the binding, which supports its higher affinity to the enzyme in comparison to UDP-Gal.

6.3. Material and Methods

Macromolecule preparation. Two X-ray crystal structure of GalNAc-T2 were selected: one in the closed active conformation (PDB ID 4D0Z, in complex with UDP-5S-GalNAc, EA2 peptide and Mn^{+2}), and the second one in the open inactive conformation (PDB ID 2FFV, in complex with UDP and Mn^{+2}). Chain "A" was selected in both PDB structures. All water molecules, ethylene glycol and ligands were removed, while Mn^{+2} ions were kept. By using the Maestro package,[47] missing residues were added and modeled, and neutral terminal N and C groups and hydrogen atoms were added. Charges were assigned to all atoms with OPLS_2005 force field, and the final protein structure was finally minimized with the same force field.

Ligand preparation. 3D structure for UDP was extracted from PDB 2FFV. 3D structure for UDP-GalNAc was built from the 3D structure of UDP-5S-GalNAc from PDB 4D0Z. UDP-5S-GalNAc was also prepared for docking as a reference compound. The 3D structure of compounds 16 were built from their SMILES codes. All ligands were subjected to geometry optimization by using the MMFFs force filed implemented in Maestro. This method was chosen comparing the docking results for reference compounds by applying different force fields.

Docking calculations. Docking was performed by means of Glide program[41] and AutoDock4 program.[48] In Glide, a cubic grid box of 25^3 $Å^3$ was generated for each protein, defining as centre the centre of mass among Asp176, His145, Arg362, and His226 residues. For all the ligands, the torsional angles were allowed to rotate with no restrictions. A standard precision docking was performed for both structures. In the case of AutoDock, different conformers of the starting geometries of the ligands were docked by using the Lamarckian genetic algorithm, by randomly changing the torsion angles and overall orientation of the molecule. A volume for exploration was defined with a grid spacing of 0.375 Å, centered on the centre of mass among Asp176, His145, Arg362, and His226 residues, and a 3D grid of 21 × 18 × 19 $Å^3$ for 4D0Z, and 21 × 21 × 18 $Å^3$ for 2FFV. After docking, the 200 solutions were clustered in groups with root mean-square deviations less than 2.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

Docking results from both Glide and AutoDock methods were analysed and taken together for the discussion.

Targeted molecular dynamics (tMD) simulation. The tMD simulation was performed with the AMBER12 software[49] by using the ff10, gaff and Glycam force fields. The Cα atoms of the starting structure (GalNAc-T2 in complex with UDP-GalNAc in the open conformation) were restrained to the corresponding positions of the target structure (GalNAc-T2 in complex with UDP-GalNAc in the closed conformation). Explicit water solvent (TIP3BOX model) was considered by adding the same number of water molecules to both complexes. Also, the same number of counterions (4 chlorine ions) was added. Energy minimization of all the initial structures (closed and open states) was carried out using the steepest descent method in AMBER12 for 500 steps and the non-bonded cut off was 9 Å. For the tMD simulation, the SHAKE algorithm was used, the time step was 0.002 ps, and the non-bonded cut off was 15 Å. The temperature coupling method was used to keep the temperature constant at 300 K and Berendsen temperature coupling scheme was used. No positional restrains were applied. The total simulation time was 5 ns.

Bibliography

1. Voglmeir, J.; Flitsch, S. L., Glycosyltransferases. In *Biocatalysis in Organic Synthesis 1*, 2015 ed.; Faber, K.; Fessner, W. D.; Turner, N. J., Eds. Georg Thieme Verlag: Stuttgart, 2015; Vol. 1, pp 507-542.
2. Gloster, T. M., Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **2014,** *28*, 131-141.
3. Roychoudhury, R.; Pohl, N. L. B., New structures, chemical functions, and inhibitors for glycosyltransferases. *Curr. Opin. Chem. Biol.* **2010,** *14* (2), 168-173.
4. Chen, X., Fermenting Next Generation Glycosylated Therapeutics. *ACS Chem. Biol.* **2011,** *6* (1), 14-17.
5. Wang, S.; Vidal, S., Recent design of glycosyltransferase inhibitors. *Carbohydr. Chem.* **2013,** *39*, 78-101.
6. Merino, P.; Tejero, T.; Delso, I.; Hurtado-Guerrero, R.; Gomez-SanJuan, A.; Sadaba, D., Recent progress on fucosyltransferase inhibitors. *Mini-Rev. Med. Chem.* **2012,** *12*, 1455-1464.
7. Kajimoto, T.; Node, M., Synthesis of glycosyltransferase inhibitors. *Synthesis* **2009,** (19), 3179-3210.
8. Whalen, L. J.; Greenberg, W. A.; Mitchell, M. L.; Wong, C.-H., Iminosugar-based glycosyltransferase inhibitors. *Iminosugars* **2007**, 153-176.
9. Jung, K.-H.; Schmidt, R. R., Glycosyltransferase inhibitors. *Carbohydrate-Based Drug Discovery* **2003,** *2*, 609-659.
10. Qian, X.; Palcic, M. M., Glycosyltransferase inhibitors. In *Carbohydrates in Chemistry and Biology*, 2000; Vol. 3, pp 293-312.
11. Tedaldi, L.; Wagner, G. K., Beyond substrate analogues: New inhibitor chemotypes for glycosyltransferases. *MedChemComm* **2014,** *5* (8), 1106-1125.
12. Liang, D.-M.; Liu, J.-H.; Wu, H.; Wang, B.-B.; Zhu, H.-J.; Qiao, J.-J., Glycosyltransferases: mechanisms and applications in natural product development. *Chem. Soc. Rev.* **2015,** *44* (22), 8350-8374.
13. Albesa-Jové, D.; Mendoza, M. F.; Rodrigo-Unzueta, A.; Gomollón-Bel, F.; Cifuente, J.; Urresti, S.; Comino, N.; Gómez, H.; Romero-García, J.; Lluch, J. M.; Sancho-Vaello, E.; Biarnés, X.; Planas, A.; Merino, P.; Masgrau, L.; Guerin, M. E., A natural ternary complex trapped in crystal reveals the catalytic mechanism of a retaining glycosyltransferase. *Angew. Chem. Int. Ed.* **2015,** *54*, 9898-9902.
14. Ardevol, A.; Rovira, C., Reaction mechanisms in carbohydrate-active enzymes: Glycosyl hydrolases and glycosyltransferases. Insights from ab initio quantum mechanics/molecular mechanics dynamic simulations. *J. Am. Chem. Soc.* **2015,** *137* (24), 7528-7547.
15. Gomez, H.; Polyak, I.; Thiel, W.; Lluch, J. M.; Masgrau, L., Retaining glycosyltransferase mechanism studied by QM/MM methods: Lipopolysaccharyl-α-1,4-galactosyltransferase C transfers α-galactose via an oxocarbenium ion-like transition state. *J. Am. Chem. Soc.* **2012,** *134* (10), 4743-4752.
16. Hurtado-Guerrero, R.; Davies, G. J., Recent structural and mechanistic insights into post-translational enzymatic glycosylation. *Curr. Opin. Chem. Biol.* **2012,** *16* (5-6), 479-487.
17. Tvaroska, I., QM/MM insight on enzymatic reactions of glycosyltransferases. *Mini-Rev. Med. Chem.* **2011,** *8* (3), 263-269.
18. Jakeman, D. L., Mechanisms of Glycosyltransferases: The in and the out. *ChemBioChem* **2011,** *12* (17), 2540-2542.

19. Rini, J.; Esko, J.; Varki, A., Glycosyltransferases and Glycan-processing Enzymes. In *Essentials of Glicobiology. 2nd edition*, Varki, A.; Cummings, R. D.; Esko, J. D.; Freeze, H. H.; Stanley, P.; Bertozzi, C. R.; Hart, G. W.; Etzler, M. E., Eds. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2009.
20. Malik, V.; Black Gary, W., Structural, functional, and mutagenesis studies of UDP-glycosyltransferases. *Advances in protein chemistry and structural biology* **2012,** *87*, 87-115.
21. Descroix, K.; Wagner, G. K., The first C-glycosidic analogue of a novel galactosyltransferase inhibitor. *Org. Biomol. Chem.* **2011,** *9* (6), 1855-1863.
22. Timmer, M. S. M.; Chumillas, M. V.; Donker-Koopman, W. E.; Alerts, J.; van der Marel, G. A.; Overkleeft, H. S.; van Boom, J. H., Selective cross-metathesis of C-allyl-glycosides. *J. Carbohydr. Chem.* **2005,** *24* (4-6), 335-351.
23. Dorfmueller, H. C.; Borodkin, V. S.; Blair, D. E.; Pathak, S.; Navratilova, I.; van Aalten, D. M. F., Substrate and product analogues as human O-GlcNAc transferase inhibitors. *Amino Acids* **2011,** *40* (3), 781-792.
24. Stolz, F.; Reiner, M.; Blume, A.; Reutter, W.; Schmidt, R. R., Novel UDP-glycal derivatives as transition state analogue inhibitors of UDP-GlcNAc 2-epimerase. *J. Org. Chem.* **2004,** *69* (3), 665-679.
25. Jiang, J.; Lazarus, M. B.; Pasquina, L.; Sliz, P.; Walker, S., A neutral diphosphate mimic crosslinks the active site of human O-GlcNAc transferase. *Nat. Chem. Biol.* **2012,** *8* (1), 72-77.
26. Jorgensen, R.; Grimm, L. L.; Sindhuwinata, N.; Peters, T.; Palcic, M. M., A Glycosyltransferase Inhibitor from a Molecular Fragment Library Simultaneously Interferes with Metal Ion and Substrate Binding. *Angew. Chem. Int. Ed.* **2012,** *51* (17), 4171-4175.
27. Schaefer, K.; Albers, J.; Sindhuwinata, N.; Peters, T.; Meyer, B., A new concept for glycosyltransferase inhibitors: Nonionic mimics of the nucleotide donor of the human blood group B galactosyltransferase. *Chembiochem* **2012,** *13* (3), 443-450.
28. Errey, J. C.; Lee, S. S.; Gibson, R. P.; Fleites, C. M.; Barry, C. S.; Jung, P. M. J.; O'Sullivan, A. C.; Davis, B. G.; Davies, G. J., Mechanistic Insight into Enzymatic Glycosyl Transfer with Retention of Configuration through Analysis of Glycomimetic Inhibitors. *Angew. Chem. Int. Ed.* **2010,** *49* (7), 1234-1237.
29. Wang, S.; Cuesta-Seijo, J. A.; Lafont, D.; Palcic, M. M.; Vidal, S., Design of Glycosyltransferase Inhibitors: Pyridine as a Pyrophosphate Surrogate. *Chem. Eur. J.* **2013,** *19* (45), 15346-15357.
30. Wang, S.; Shen, D. L.; Lafont, D.; Vercoutter-Edouart, A.-S.; Mortuaire, M.; Shi, Y.; Maniti, O.; Girard-Egrot, A.; Lefebyre, T.; Pinto, B. M.; Vocadlo, D.; Vidal, S., Design of glycosyltransferase inhibitors targeting human O-GlcNAc transferase (OGT). *MedChemComm* **2014,** *5* (8), 1172-1178.
31. Jiang, J.; Kanabar, V.; Padilla, B.; Man, F.; Pitchford, S. C.; Page, C. P.; Wagner, G. K., Uncharged nucleoside inhibitors of beta-1,4-galactosyltransferase with activity in cells. *Chemical Communications (Cambridge)* **2016,** *52* (20), 3955-8.
32. Gloster, T. M.; Zandberg, W. F.; Heinonen, J. E.; Shen, D. L.; Deng, L.-H.; Vocadlo, D. J., Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. *Nat. Chem. Biol.* **2011,** *7* (3), 174-181.
33. Schafer, A.; Thiem, J., Synthesis of novel donor mimetics of UDP-Gal, UDP-GlcNAc, and UDP-GalNAc as potential transferase inhibitors. *J. Org. Chem.* **2000,** *65*, 24-29.

34. Chang, R.; Vo, T.-T.; Finney, N. S., Synthesis of the C1-phosphonate analog of UDP-GlcNAc. *Carbohydr. Res.* **2006,** *341* (12), 1998-2004.
35. Hajduch, J.; Nam, G.; Kim, E. J.; Froehlich, R.; Hanover, J. A.; Kirk, K. L., A convenient synthesis of the C-1-phosphonate analog of UDP-GlcNAc and its evaluation as an inhibitor of O-linked GlcNAc transferase (OGT). *Carbohydr. Res.* **2008,** *343* (2), 189-195.
36. Clarke, A. J.; Hurtado-Guerrero, R.; Pathak, S.; Schuttelkopf, A. W.; Borodkin, V.; Shepherd, S. M.; Ibrahim, A. F. M.; van Aalten, D. M. F., Structural insights into mechanism and specificity of O-GlcNAc transferase. *EMBO J.* **2008,** *27* (20), 2780-2788.
37. Partha, S. K.; Sadeghi-Khomami, A.; Slowski, K.; Kotake, T.; Thomas, N. R.; Jakeman, D. L.; Sanders, D. A. R., Chemoenzymatic synthesis, inhibition studies, and X-ray crystallographic analysis of the phosphono analog of UDP-galp as an inhibitor and mechanistic probe for UDP-galactopyranose mutase. *J. Mol. Biol.* **2010,** *403* (4), 578-590.
38. Forget, S. M.; Jee, A.; Smithen, D. A.; Jagdhane, R.; Anjum, S.; Beaton, S. A.; Palmer, D. R. J.; Syvitskia, R. T.; Jakeman, D. L., Kinetic evaluation of glucose 1-phosphate analogues with a thymidylyltransferase using a continuous coupled enzyme assay. *Org. Biomol. Chem.* **2015,** *13*, 866-875.
39. Vidal, S.; Bruyere, I.; Malleron, A.; Auge, C.; Praly, J.-P., Non-isosteric C-glycosyl analogues of natural nucleotide diphosphate sugars as glycosyltransferase inhibitors. *Biorg. Med. Chem.* **2006,** *14* (21), 7293-7301.
40. Lira-Navarrete, E.; de las Rivas, M.; Companon, I.; Pallares, M. C.; Kong, Y.; Iglesias-Fernandez, J.; Bernardes, G. J. L.; Peregrina, J. M.; Rovira, C.; Bernado, P.; Bruscolini, P.; Clausen, H.; Lostao, A.; Corzana, F.; Hurtado-Guerrero, R., Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates proteinO-glycosylation. *Nat. Commun.* **2015,** *6*, 6937.
41. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., *Nucleic Acids Res.* **2009,** *37*, D233-D238.
42. Gomez, H.; Rojas, R.; Patel, D.; Tabak, L. A.; Lluch, J. M.; Masgrau, L., A computational and experimental study of O-glycosylation. Catalysis by human UDP-GalNAc polypeptide: GalNAc transferase-T2. *Org. Biomol. Chem.* **2014,** *12* (17), 2645-2655.
43. Dondoni, A.; Marra, A., Validating the alkene and alkyne hydrophosphonylation as an entry to organophosphonates. *Org. Biomol. Chem. 13*, 2212-2215.
44. Lira-Navarrete, E.; Iglesias-Fernandez, J.; Zandberg, W. F.; Companon, I.; Kong, Y.; Corzana, F.; Pinto, B. M.; Clausen, H.; Peregrina, J. M.; Vocadlo, D. J.; Rovira, C.; Hurtado-Guerrero, R., Substrate-guided front-face reaction revealed by combined structural snapshots and metadynamics for the polypeptide N-acetylgalactosaminyltransferase 2. *Angew. Chem. Int. Ed. Engl.* **2014,** *53* (31), 8206-8210.
45. Fritz, T. A.; Raman, J.; Tabak, L. A., Dynamic association between the catalytic and lectin domains of human UDP-GalNAc: Polypeptide alpha-N-acetylgalactosaminyltransferase-2. *J. Biol. Chem.* **2006,** *281* (13), 8613-8619.
46. Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M., Metal ions in biological catalysis: From enzyme databases to general principles. *J. Biol. Inorg. Chem.* **2008,** *13* (8), 1205-1218.
47. *Schrödinger Release 2015-1: MacroModel, version 10.7, Schrödinger, LLC, New York, NY*, 2015.

48. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
49. *AMBER 12*, D.A. Case, T.A. Darden, 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. Götz, 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.

Annex VI

Table 6.A1. Data collection and refinement statistics. Values in parentheses refer to the highest resolution shell. Ramachandran plot statistics were determined with PROCHECK.

	*GalNAc-T2-**14d***
Space group	$P2_12_12_1$
Wavelength (Å)	0.97
Resolution (Å)	20-2.07 (2.124-2.07)
Cell dimensions (Å)	*a* =116.49 *b* = 121.75 *c* = 250.15
Unique reflections	215817
Completeness	99.8 (99.84)
R_{sym}	0.061 (0.632)
$I/\sigma(I)$	15.6 (2.6)
Redundancy	5.7 (5.7)
R_{work} / R_{free}	0.189/0.246
RMSD from ideal geometry, bonds (Å)	0.017
RMSD from ideal geometry, angles (º)	1.996
<*B*> protein (Å^2)	48.81
X1<*B*> Ethlenglycol (Å^2)	64.24
<*B*> Glycerol (Å^2)	71.18
<*B*> DRG(Å^2)	64.33
<*B*> UDP (Å^2)	47.52
<*B*> Mn^{+2} (Å^2)	32.78
<*B*> solvent (Å^2)	49.87
Ramachandran plot: Most favoured (%) Additionally allowed (%) Disallowed (%)	 93.37 5.09 1.53
PDB ID	5FV9

CONCLUSIONS

Molecular modelling techniques have been extensively applied to the elucidation of glycan-protein interactions at the atomic level of several systems with biological interest, and have provided new insights for the understanding of the molecular recognition events underlying the biological functions of these systems.

A pluridisciplinary approach, in a collaborative work with experimental groups, has been carried out for the study of carbohydrate-protein interaction. Molecular modelling studies have assisted to NMR experimental methods to unravel the molecular recognition features of a biologically relevant trisaccharide (trimannoside) with *Pisum sativum* lectin as a model lectin. This combined study has required the use of synthetic fluorinated complex glycan mimics to provide detailed structural and binding information for every sugar ring, allowing the elucidation of the interacting ligand epitope and the bound conformations. Computational techniques have been proved to assist the ^{19}F NMR experiments and to provide the necessary 3D perspective for the elucidation of this example of molecular recognition of high complexity.

The interaction of a series of glycans with human galectin-3 has been studied, pointing out to the subtle differences in the recognition event. By means of combined molecular modelling and NMR techniques, we have identified the ligand-receptor interactions accounting for the gal-3 recognition of three glycans containing the lacdiNAc motif: blood group A type II tetrasaccharide, and two compounds derived from the important xenoantigen α-Gal epitope. Our findings have unravelled important details at atomic level of the gal-3 recognition.

Computational studies of human galectin-1, -3 and -7, focused on the binding pocket adjacent to the CRD, have provided structural basis for a fragment-based design of novel galectin binders with predicted selectivity and high affinity. In collaboration with other groups, selected compounds have been synthesized and preliminary NMR and ITC results are very promising. These selective ligands may contribute to the understanding of the highly relevant biological functions of galectins, and their role in cancer.

A computational model for the interaction of the PDPN epitope (Neu5α2-3Galβ1-3(Neu5α2-6)GalNAc) and *Maackia amurensis* seed lectins isoforms has accounted for the protective action of these lectins against cartilage degradation in osteoarthritis (OA). The docking of the tetrasaccharide shows the preference of this ligand to bind MAH isoform, due to the presence of disialyl-core and the specific Galβ-1-3GalNAc glycosidic linkage. The increased levels of the α-2,3-SiaT isoforms and the corresponding increase in the levels of α-2,3-sialylated glycoproteins in osteoarthritic chondrocytes may shed mechanistic light on the pathophysiology of OA. The ability of MASL to target sialylated glycoproteins such as PDPN, which is also induced during OA and rheumatoid arthritis and to protect chondrocytes from insults leading to cartilage degradation might offer further possibilities for therapeutic interventions and novel arthritis treatments that may include the regulation of sialylation during acute disease stages.

GalNAc-T2 has been used as a prototype GT for studying the binding mode of less-polar glycomimetics. In particular, from the docking studies it is inferred that the β-phosphate is required for binding to the Mn^{+2}, and thus the replacing of the α-phosphate while keeping the phosphate or alike groups in the β position should maintain or improve the binding of these compounds. Finally, the tMD simulation of the complex of GalNAc-T2 with UDP-GalNAc has provided new insights on the substrate recognition events involved in the preliminary steps of the catalytic cycle of GalNAc-T2.

CONCLUSIONES

Se han empleado técnicas de modelado molecular para la elucidación a nivel atómico de las interacciones carbohidrato-proteína de varios sistemas con interés biológico y terapéutico. En concreto, se han estudiado los siguientes sistemas: galectinas humanas -1, -3 y -7, lectina de *Pisum sativum*, lectina de semillas de *Maackia amurensis*, y la glicosiltransferasa GalNAc-T2. Nuestros estudios han permitido aportar nuevas perspectivas para la comprensión de los procesos de reconocimiento molecular implicados en las funciones biológicas de estos sistemas.

Se ha llevado a cabo un estudio pluridisciplinar, en colaboración con grupos experimentales, abordándose el estudio de los modos de unión de una trimanosida trifluorada, con enlaces α1 → 3 y α1 → 6, y sus derivados disacáridos fluorados, con una lectina modelo de *Pisum sativum*, combinando los estudios computacionales con datos experimentales de ^{19}F-RMN. Para llevar a cabo este estudio, se han empleado glicómiméticos fluorados sintéticos para la elucidación de detalles estructurales y de la interacción de cada manosa fluorada con la proteína, identificando el epítopo y las conformaciones de unión. Este tipo de estudio no hubiera sido posible sin la presencia del flúor. Las técnicas computacionales han proporcionado la perspectiva 3D necesaria para la elucidación de este ejemplo de reconocimiento molecular de gran complejidad, sineod así un soporte importante de los experimentos de ^{19}F-RMN. La combinación de los datos experimentales de modelado y los datos de RMN indica la coexistencia de dos modos de unión diferentes para la trimanosida trifluorada cuando se une a la lectina de *Pisum sativum*. Este ejemplo ilustra la complementariedad entre estas dos técnicas para la elucidación de problemas complejos de reconocimiento molecular donde las técnicas experimentales no pueden proporcionar una respuesta completa.

Se ha abordado un estudio computacional de la unión de varios ligandos de β-galactósido (lactosa, LacdiNAc y tres glicanos que contienen el motivo LacNAc: tetrasacárido del grupo sanguíneo A tipo II, y dos compuestos derivados del importante epítopo α-Gal de xenoantigeno) hacia la galectina-3 humana. Los epítopos LacdiNAc y α-Gal contienen el motivo LacNAc, que ha sido propuesto como patrón molecular para la respuesta inmune mediada por galectina-3. Se han caracterizado las interacciones clave necesarias para la unión a la galectina-3, identificando la sutileza de las diferencias en el evento de reconocimiento, mediante técnicas combinadas de modelado molecular y RMN. Utilizando este enfoque combinado, se han descrito a nivel atómico las interacciones ligando-receptor que caracterizan el reconocimiento de la gal-3 por los epítopos LacdiNAc y α-Gal.

Se ha realizado el diseño computacional de nuevos ligandos selectivos de galectinas y con mejor afinidad teórica. En particular, hemos centrado nuestros estudios en las galectinas -1, -3 y -7, involucradas en la adhesión, crecimiento y regulación celular, que además participan en varios procesos tumorales. Nuestro diseño computacional se ha centrado en el bolsillo de unión adyacente al dominio de reconocimiento de carbohidratos. Esto nos han proporcionado una base estructural para llevar a cabo un diseño basado en fragmentos de nuevos ligandos de galectina con selectividad y alta afinidad. Con esta aproximación, identificamos los mejores fragmentos capaces de anclarse a este bolsillo en cada una de las galectinas y diseñamos derivados de lactosa incorporando los mejores fragmentos en su estructura. En colaboración con otros grupos, se han sintetizado algunos compuestos seleccionados y los resultados preliminares de RMN e ITC son muy prometedores. Estos ligandos selectivos pueden contribuir a la comprensión de las funciones biológicas de las galectinas, altamente relevantes, y su papel en el cáncer.

Se ha propuesto un modelo computacional para la interacción del epítopo de podoplanina (PDPN) (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) con las isoformas de lectina de semilla de *Maackia amurensis* (MASL), explicando la acción protectora de estas lectinas en la degradación del cartílago en la osteoartritis (OA). Hemos explorado el sitio de unión de las dos especies moleculares de la MASL, la leucoaglutinina de *Maackia amurensis* (MAL) y la hemaglutinina de *Maackia amurensis* (MAH), y el posible reconocimiento por el epítopo de la glicoproteína podoplanina. La capacidad de MASL de unirse a glicoproteínas sialiladas como la podoplanina, que se inducen durante la osteoartritis y la artritis reumatoide, protege a los condrocitos de insultos que conducen a la degradación del cartílago. Este mecanismo puede suponer nuevas formas de intervención terapéutica y nuevos tratamientos para la artritis que incluyan la regulación de sialilación durante estadios agudos de la enfermedad. En particular, la unión del tetrasacárido (Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc) muestra la preferencia de este ligando para unirse a la isoforma de la MAH, debido a la presencia de dos unidades de acido siálico y a la presencia del enlace glucosídico específico Galβ1-3GalNAc. Los resultados experimentales indican que la MASL preserva la estructura y la función del cartílago bajo diversos insultos artríticos interfiriendo con la función de los receptores transmembrana α-2,3-sialilados, tales como la glicoproteína transmembrana podinaplanina de tipo mucina (PDPN). Se propone la interacción directa de MASL con PDPN como un mecanismo plausible para esta actividad protectora de la lectina, y nuestros estudios computacionales han proporcionado un modelo molecular 3D para dicha interacción.

Se ha estudiado el modo de unión de una serie de miméticos sintéticos de nucleósidos como inhibidores de una glicosiltransferasa modelo, la GalNAc-T2. Estos análogos de nucleótidos son compuestos menos polares derivados de la uridina que contiene sólo el β-fosfato y podrían ser ligandos eficaces para la inhibición de la enzima, como se deduce de los datos experimentales. Hemos propuesto un modelo 3D computacional para el modo de unión de estos ligandos para proporcionar más información sobre el mecanismo del ciclo catalítico de esta familia de enzimas. En particular, a partir de los estudios de modelado molecular se deduce que la sustitución de uno de los fosfatos (el de la posición β respecto a la desoxirribosa), mantiene la capacidad de unión. Los estudios de docking sugieren que estos análogos menos polares son posibles ligandos de GalNAc-T2, donde la parte de la uridina desempeña un papel principal en el enlace, de acuerdo además con los modos de unión observados en las estructuras cristalográficas. El grupo fosfonato y el azúcar establecen interacciones adicionales, no encontrándose diferencias al estudiar otros tipos de azúcar. Además, la simulación de tMD nos ha dado ha permitido estudiar los cambios en la pose de unión a la GT que el UDP-GalNAc sufre en el proceso de cierre/apertura. Las interacciones específicas de GalNAc en este proceso de cierre/apertura pueden ser útiles entender el ciclo del mecanismo catalítico de esta enzima y para el diseño de ligandos selectivos de GT. De hecho, nos han proporcionado una explicación plausible para el papel del grupo N-acetilamido en el enlace, que apoya su mayor afinidad a la enzima en comparación con la UDP-Gal.