Lectin-Glycan Interactions: new NMR insights on the role of dynamics and presentation by using state-of-the-art NMR methodologies

Sara Bertuzzi

Doctoral Thesis

2022

Thesis directors:

Dr. Jesús Jiménez-Barbero

Dr. Cristina Airoldi

University Tutor: Esther Lete

Lectin-Glycan Interactions: new NMR insights on the role of dynamics and presentation by using state-of-the-art NMR methodologies

Sara Bertuzzi

Doctoral Thesis

2022

Thesis directors:

Dr. Jesús Jiménez-Barbero

Dr. Cristina Airoldi

University Tutor: Esther Lete

This doctoral thesis has been performed at the Center for Cooperative Research in Biosciences (CIC bioGUNE) under cotutelle regime of the University of Basque Country (UPV/EHU) and the University of Milano-Bicocc a (Unimib)

CIC bioGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

EXCELENCIA SEVERO OCHOA

Alla mia famiglia

Acknowledgments

The work presented in this Thesis was performed at CIC bioGUNE and at the University of Milano-Bicocca (Unimib) under the supervision of Prof. Dr. Jesús Jiménez-Barbero and Prof. Dr. Cristina Airoldi.

First, I would like to thank both my directors, Jesús Jiménez-Barbero and Cristina Airoldi for giving me the opportunity to perform my Thesis under their supervision. Thank you for your continuous support throughtout this journey and for everything that you have taught me during all these years.

I would also like to thank all my colleagues (past and present) from the Chemical Glycobiology lab. Thank you for all your help and for having shared this wonderful experience with me.

Contents

Abbreviations

BLI Biolayer Interferometry

CHO Chinese hamster ovary

CRD Carbohydrate Recognition Domain

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

CryoEM Cryo Electron Microscopy

CSP Chemical Shift Perturbation

DLS Dynamic Light Scattering

DOSY Diffusion Ordered SpectroscopY

DTT Dithiothreitol

DTT-d_{10} Dithiothreitol-d_{10}

E. coli *Escherichia coli*

ECM Extracellular matrix

ER Endoplasmic Reticulum

ESI ElectroSpray Ionization

EXSY Chemical Exchange SpectroscopY

FACS Fluorescence Activated Cell Sorting

FLP Flippase recombinase

Fuc Fucose

GAG GlycosAminoGlycan

Gal Galactose

GalNAc N-Acetylgalactosamine

GBP Glycan binding protein

Glc Glucose

GlcA Glucuronic Acid

GlcNAc N-Acetylglucosamine

GLS Glycosphingolipid

GT Glycosyltransferase

HaCaT Immortal human keratinocyte

HB Hydrogen Bond

HBGA Histo Blood Group Antigens

HEK Human Embryonic Kidney

HNC Head and neck cancer

HPLC High Performance Liquid Chromatography

HSQC Heteronuclear Single Quantum Coherence

IPTG Isopropyl-β-D-1-thio-galactopyranoside

ITC Isothermal Titration Calorimetry

LacNAc N-Acetyllactosamine

LB Luria-Bertani broth

Man Mannose

MD Molecular Dynamics

MIC Minimal Inhibitory Concentration

MS Mass Spectrometry

MW Molecular Weight

NAc N-Acetyl

NMR Nuclear Magnetic Resonance

NOE Nuclear Overhauser Effect

NOESY NOE SpectroscopY

ORF Open Reading Frame

PDB Protein Data Bank

PHA-L *Phaseolus vulgaris* Leucoagglutinin

PI Propidium Iodide

PMSF phenylmethylsulfonyl fluoride

PNA Peanut Agglutinin

PTM Post-translational modification

ROESY Rotating frame Overhauser Effect SpectroscopY

RT Room temperature

SDS-PAGE Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis

SNFG Symbol nomenclature for glycans

SPR Surface Plasmon Resonance

STD Saturation Transfer Difference

TDG Thiodigalactoside

TOCSY Totally Correlated SpectroscopY

TROSY Transverse Relaxation-Optimized SpectroscopY

UV Ultraviolet

VVA *Vicia villosa* lectin

WT Wild Type

Abstract

Glycans are ubiquitous in Nature. They are mainly found conjugated with other biomolecules giving rise to glycoproteins and glycolipids.

Relevant biological events such as cell-cell, cell-matrix and host-pathogen interactions are orchestrated through the recognition of glycans by a group of specific proteins, called lectins. Several sub-families of lectins can be differentiated, being galectins, C-type lectins, and siglecs the most studied and suitable for biomedical applications.

In fact, the lectin-sugar binding events play a crucial role in a myriad of physiological and pathological processes in all living cells.

For this reason, a major challenge in modern Science is to obtain a comprehensive view of such fundamental interactions. However, individual methodologies only provide a partial picture of the phenomena. Thus, in this Thesis, a multidisciplinary methodology has been adopted to tackle this problem. For this purpose, different lectin-sugar systems have been selected and the fine details of their binding events have been decoded from the structural, conformational, and dynamic points of view. Particular attention to aspects such as conformational motions, multivalency, presentation, and the actual *in-vitro* or cell-like environment were the driving force behind this work.

First, the interactions between galectin-1 and the human blood group antigens have been scrutinized. This study was established in a framework of a global project on the recognition of self-antigens by galectins. Indeed, moved by the interest in detecting different binding features among galectins, a synergistic approach of techniques, including NMR, ITC, and MD simulations, was employed. Exquisite events of conformational plasticity and allostery were detected and described.

In a second section, the impact of the multivalent presentation of sugar epitopes was studied against the dimeric galectin-1 and the monomeric CRD of galectin-3. HMPA copolymers decorated with LacNAc disaccharide had previously displayed affinities for these lectins in the nanomolar range with detectable preference for galectin-1. Our goal here was to unravel and explain the structural details behind these observations, using NMR as well DLS and Cryo-EM.

In a third part, motivated by the described outstanding selectivity displayed by the C-type lectin LSECtin for one configurational isomer over the other on an N-glycan microarray, in solution NMR investigations were performed. Herein, we shed light into how sugar-branch specificity can be dictated by the immobilization on rigid surfaces.

Since the full understanding of lectin-sugar recognition events *in vivo* remains elusive, in the last chapter, a compendium of promising results, that dissect the binding events directly *in-cell* and *on-cell*, is presented. The urgency to elucidate these interactions in a microenvironment closer to the real one, was here answered through the employment of cell-based NMR methodologies and glycogenomic engineering.

Riassunto

I glicani sono ubiquitari in natura. Si trovano prevalentemente coniugati con altre biomolecole, generando così glicoproteine e glicolipidi. Eventi di grande importanza biologica, come interazioni cellula-cellula, cellula-matrice e ospite-patogeno, sono regolati dal riconoscimento dei glicani a opera di un gruppo specifico di proteine, chiamate lectine. Le lectine possono essere differenziate in vari sottogruppi e le famiglie più studiate e con maggiori applicazioni di tipo biomedico sono le galectine, le lectine di tipo C e le siglecs. Infatti, gli eventi di riconoscimento lectina-zucchero giocano un ruolo fondamentale in un vasto numero di processi fisiologici e patologici in tutte le cellule viventi.
Alla luce di ciò, una delle più grandi sfide della scienza moderna è riuscire a ottenere una visione comprensiva di queste interazioni fondamentali. Tuttavia, l'utilizzo di singole tecniche fornisce solo un'immagine parziale del fenomeno. Al fine di arginare questo problema, in questa tesi è stata adottata una metodologia multidisciplinare per lo studio di vari sistemi lectina-zucchero. In questo modo, i dettagli del riconoscimento molecolare sono stati decodificati dal punto vista strutturale, conformazionale e dinamico. La forza trainante di questo progetto è indubbiamente un'attenzione particolare nei confronti di aspetti come multivalenza, presentazione, movimenti conformazionali e ambiente cellulare.

Inizialmente, è stata studiata l'interazione tra galectina-1 e gli antigeni umani del gruppo sanguigno. Questo lavoro si è sviluppato nell'ottica di un progetto globale focalizzato sul riconoscimento di antigeni-self a carico delle galectine. Motivati dall'interesse di identificare differenti caratteristiche di unione confrontando galectine diverse, abbiamo utilizzato un approccio sinergico combinando NMR, ITC e simulazioni MD. In questo modo, sono stati individuati e descritti raffinati eventi di plasticità conformazionale e allosteria.

Successivamente, è stato analizzato l'impatto della presentazione multipla di epitopi zuccherini sul riconoscimento da parte di una galectina dimerica (galectina-1) e di una monomerica (CRD di galectina-3). In precedenza, affinità nel range nanomolare sono state descritte per l'interazione di queste galectine con copolimeri di tipo HMPA

decorati con il disaccaride LacNAc. In particolare, una preferenza tangibile è stata dimostrata per galectina-1 rispetto a galectina-3. Il nostro obiettivo in questo caso è stato giustificare e spiegare i dettagli strutturali celati dietro a queste informazioni preliminari, utilizzando un insieme di tecniche (NMR, DLS e Cryo-EM).

Nel terzo sottocapitolo della tesi, guidati dalla stupefacente selettività di una lectina di tipo C (LSECtin) per determinati isomeri conformazionali, abbiamo effettuato studi di NMR in soluzione. Nello specifico, in precedenti analisi condotte con array, è emerso che LSECtin presenta una spiccata preferenza nei confronti di uno specifico isomero conformazionale di N-glicano rispetto al suo corrispettivo isomero. Con il nostro approccio in soluzione, abbiamo fatto luce su un aspetto fondamentale, ossia che la specificità per un determinato zucchero può essere dettata dalla sua immobilizzazione su una superficie rigida.

Dal momento che la conoscenza completa degli eventi di riconoscimento molecolare lectina-zucchero tramite studi in-vitro può essere considerato elusivo, nell'ultimo capitolo è riassunta una collezione di risultati promettenti ottenuti analizzando le interazioni direttamente in-cell o on-cell. Tramite l'impiego di metodologie NMR con cellule intatte e dell'emergente ingegneria glicogenomica, vengono descritte possibili strategie per rispondere all'urgenza di studiare l'unione proteina-carboidrato in un ambiente più vicino a quello reale.

Resumen

Los glicanos son ubicuos en la Naturaleza. Su presencia se extiende desde los organismos más simples hasta los más complejos. De hecho, son uno de los componentes fundamentales de cualquier tipo de célula.

Las funciones energéticas y estructurales cruciales de los sacáridos son esenciales. solo El exoesqueleto de los artrópodos está formado principalmente por quitina y el glucógeno sirve como almacenamiento energético de la mayoría de los organismos vivos.

Sin embargo, en las últimas décadas, se ha hecho cada vez más evidente su importancia en procesos de reconocimiento molecular. En el ambiente celular, los glicanos son capaces de interaccionar con otras macromoléculas, generando una respuesta biológica.

Cada célula dispone de una variada y única colección de carbohidratos, que se define en su totalidad como el glicoma. En este contexto, los azúcares se encuentran principalmente conjugados con otras biomoléculas como proteínas o lípidos, dando lugar a glicoproteínas y glicolípidos, respectivamente.

La mayoría de los glicanos llevan a cabo su función extracelularmente, a través de interacciones célula-célula, célula-matriz o huésped-patógeno. Estos eventos de reconocimiento orquestan el plegamiento, recambio, tráfico y estabilidad de proteínas y también regulan una miríada de actividades fisiológicas y patológicas de la célula.

Entre las principales entidades que interaccionan con glicanos (virus, bacteria, anticuerpos, hormonas, enzimas, toxinas etc.), en esta tesis nos hemos centrado en un grupo específico de proteínas, denominadas lectinas.

Pueden diferenciarse distintas sub-familias de lectinas, siendo las galectinas, lectinas de tipo C y siglecs las más estudiadas y apropiadas para aplicaciones biomédicas.

En particular, las galectinas se distinguen en base a su arquitectura supramolecular en: prototípicas (formadas por dos dominios idénticos de reconocimiento de carbohidrato unidos -CRD- no covalentemente), quiméricas (constituidas por un CRD y un conector peptídico) y de tipo tándem (formadas por dos CRD distintos unidos covalentemente a través de un conector peptídico).

Dada su implicación biológica, el estudio de los eventos de interacción lectina-azúcar es de importancia esencial no solo para la investigación básica, sino también para el desarrollo de moléculas terapéuticas. Debido a ello, uno de los mayores retos de la ciencia moderna es obtener una visión completa de estas interacciones fundamentales.

Para el análisis de las interacciones lectina-azúcar pueden utilizarse diversas herramientas, siendo una de la más adecuadas la Resonancia Magnética Nuclear (RMN), que permite de estudiar las particularidades del reconocimiento a nivel atómico.
Sin embargo, el uso de metodologías individuales solo proporciona una imagen parcial del fenómeno. Por lo tanto, en esta tesis se ha adoptado una estrategia multidisciplinar para el estudio de varios sistemas lectina-azúcar. De este modo, los detalles del reconocimiento molecular han sido descodificados desde el punto de vista estructural, conformacional y dinámico.
Además, en esta Tesis hemos prestado atención particular a aspectos como la multivalencia, la presentación, el aspecto conformacional y el ambiente celular.

Inicialmente, se ha estudiado la interacción entre galectina-1, una lectina dimérica, y los antígenos humanos de los grupos sanguíneos. Estos antígenos se encuentran en la superficie de los glóbulos rojos y su composición molecular define el grupo serotípico de un individuo.
El estudio de esta interacción no solo es importante porque es un evento que ocurre en la naturaleza, sino también por estar ligado a fenómenos de infección de patógenos que utilizan estrategias de mimetismo molecular.
Este trabajo se encuentra en el marco de un proyecto global focalizado en el reconocimiento de los antígenos por las galectinas. Las galectinas son capaces de reconocer β-galactosidos, presentes en los antígenos de los grupos sanguíneos, pero cada miembro de esta familia muestra preferencias diferentes en base a las ulteriores modificaciones alrededor del epítopo de reconocimiento.
Para este estudio se ha empleado un enfoque sinérgico, combinando distintas técnicas experimentales, como RMN y la calometría -*isothermal titration calorimetry*- (ITC) con simulaciones de dinámica molecular DM. De este modo, se han detectado y

analizado, de manera precisa, eventos de plasticidad conformacional que ocurren en la proteína tras la unión con el ligando. Este tipo de movimientos concertados, sufrido por algunos residuos situados a lo largo del dímero de la proteína, está relacionado con fenómenos de alostería.

Estos fenómenos son únicos para la galectina-1 y no han sido detectados en el caso de otras galectinas (tanto monoméricas como de tipo tándem). Es decor, cada una de ella se uned a los mismos epítopos con mecanismos perfectamente regulados y dependientes también de su arquitectura.

Posteriormente, se ha analizado el impacto de la presentación multivalente de epítopos de azucares frente una galectina dimérica (galectina-1) y otra monomérica (galectina-3). Se había descrito previamente que copolímeros de tipo HMPA, decorados con el disacárido LacNAc, mostraban afinidades en el rango nM frente a estas galectinas.

Nuestro objetivo en este proyecto fue desentrañar y explicar los detalles estructurales que gobernaban estas observaciones, usando una combinación de métodos experimentales: RMN, dispersión dinámica de luz (DLS) y crio-microscopía electrónica (Cryo-EM). De esta manera, se ha proporcionado una imagen detallada del evento de reconocimiento en diferentes grados de resolución: desde la perspectiva atómica hasta la supramolecular. Se ha confirmado que la presentación múltiple de los epítopos por parte de los glicopolímeros desencadena el aumento de la afinidad en comparación con sus contrapartes monovalentes.

También se ha evaluado el efecto de diferentes tipos de presentación múltiple (mono-, bi- y trivalente) en la misma rama del gliopolímero.

En general, se ha detectado una clara preferencia de unión para galectina-1 frente a la galectina-3, debido a la formación de diferentes entidades supramoleculares dictadas por su arquitectura intrínseca.

Esta selectividad, basada en la estructura tanto de la molécula multivalente como de la galectina, es excepcional y allana el camino al desarrollo de compuestos de interés biomédico, considerando también las características compatibles de los polímeros HMPA con aplicaciones terapéuticas.

En tercer lugar, abordamos la comparación de los resultados experimentales encontrados usando técnicas que estudian fenómenos en disolución, como el RMN, frente a otras que lo hacen en superficies, como los *microarrays*.
En concreto, el estudio ha sido motivado por la selectividad excepcional mostrada por LSECtin, una lectina de tipo C, por un N-glicano biantenario asimétrico respecto a su análogo posicional usando *microarrays* de N-glicanos.
Se han llevado a cabo estudios detallados de RMN en disolución para poder comparar los datos obtenidos por las dos técnicas. De esta manera, se ha demostrado que la especificidad observada estaba dictada por la inmovilización de los N-glicanos en una superficie rígida y no tenía lugar en disolución.
En la superficie, la presentación del epítopo de azúcar que interactúa es esencial, y, por lo tanto, tanto la longitud y la naturaleza química de los enlazadores utilizados para unir los ligandos a la superficie como la composición del soporte sólido en sí misma, son elementos importantes que pueden influir en el resultado final.
Por el contrario, cuando el análisis se realizó por RMN en disolución, los actores disfrutan de la libertad de movimiento que no tienen en un entorno natural.
Con este ejemplo se ha demostrado la dificultad que entraña traducir los resultados obtenidos *in vitro* al ambiente *in vivo* y la necesidad de complementar metodologías diferentes para poder obtener imágenes precisas de los sistemas de interés.

En el último capítulo de resultados, y teniendo en cuenta que el conocimiento sobre el reconocimiento molecular lectina-azúcar *in vivo* es todavía escaso, se ha presentado un compendio de resultados prometedores obtenidos para distintos sistemas lectina-glicano, analizando los eventos de interacciones directamente dentro de la célula, *in-cell*, o sobre la célula, *on-cell*, con el propósito de elucidar los detalles de estas interacciones en un ambiente más cercano al natural.
Así, se han utilizado metodologías de NMR basadas en células y experimentos de ingeniería glico-genómica.
Especificadamente, se han utilizado ovocitos de pez cebra (*Danio rerio*) para detectar la unión de galectina-7 al tiodigalactosido (TDG) directamente en las células (*in-cell*) a través de experimentos bidimensionales de RMN.

Consecuentemente, se han presentado dos estrategias similares de RMN sobre la célula (*on-cell*) utilizando células bacterianas, así como células de mamífero con el fin de estudiar el reconocimiento de glicanos seleccionados por los receptores expuestos.
En el primer caso, se ha observado FimH, una proteína que reconoce manosa y se expresa naturalmente en el extremo apical del *pilus* de cepas uropatogénicas de *Escherichia coli*. Su unión a dendrímeros funcionalizados con manosa ha sido examinada en un entorno celular a través de experimentos de RMN.
Por otro lado, las células HEK293 han sido manipuladas para sobreexpresar el receptor siglec-10 en su superficie y se ha analizado su reconocimiento con un glicomimético sintético que es selectivo para esta lectina, midiendo la interacción por RMN directamente en la célula (*on-cell*).

Finalmente, motivados por los recientes y rápidos desarrollos en el campo de la ingeniería glico-genómica, se presentan resultados preliminares de la modificación estable de genes relacionados con la glicosilación (glicosiltransferasas) en células de mamífero (HEK293 y HaCaT). Las líneas celulares modificadas presentan un tipo de glicosilación homogénea y definida (solo N-glicanos, solo O-glicanos o solo glicoesfingolípidos). Estas líneas han sido utilizadas en ensayos FACS para estudiar cómo las galectinas exógenas son capaces de reconocer sus glicanos en un contexto celular.

CHAPTER 1

GENERAL INTRODUCTION

1.1 Glycans

Glycans are everywhere. They are fundamental components of intra- and extra-cellular compartments of organisms in all domains of life and the most abundant biomolecules in nature. [1] Indeed, it has been estimated that sugars constitute the 70% of the total biomass on Earth. [2]

Saccharides (glycans, sugars, carbohydrates) have long been known to play major structural and energetic roles. Chitin, for instance, is the primary component of cell wall in fungi and forms the exoskeleton of all arthropods, while cellulose provides structural support to every plant cell. Likewise, starch and glycogen are examples of polysaccharides able to serve as energy storage for the metabolism of living organisms. However, in the last decades, the essential role of carbohydrates in molecular recognition events has attracted the attention of many researchers, with diverse background. In the cellular microenvironment, glycans bind to other partners, such as proteins, and generate crucial biological outcomes.

Each different cell type displays a unique and rich battery of complex carbohydrates, which is called the *glycome*. Different types of glycoconjugates can be found in the intracellular area or, predominantly, exposed on the cell surface. Glycans are typically found covalently linked to proteins or lipids, generating glycoproteins or glycolipids, respectively. More recently, the glycosylation of the small RNAs displayed on the surface of living cells has also been postulated. [3]

The study of sugar synthesis, structure, conformation, interactions, biology, and evolution is the fundament of glycosciences, joining glycochemistry, glycobiology, glycobiotechnology, and glycobiomedicine. [4]
Since sugar-mediated binding events are essential for cell-cell, cell-matrix, and cell-pathogen communication and meticulously modulate the physiology of the cell in health and disease, this research field is constantly expanding. In light of this, the understanding of the basis of carbohydrates' structural diversity, linkage to other biomolecules, and cellular distribution is of paramount importance.

1.1.1 Structural diversity

Carbohydrates seldom occur in nature as single units. In fact, they are generally organized into polymeric structures in which monosaccharides constitute the basic building blocks. The complexity of glycans not only arises from the existing plethora of different monosaccharides. Indeed, even if hundreds of building blocks have been discovered, most of the mammalian *glycome* consists of only ten members.

The most recurrent 9 monosaccharides are D-glucose (Glc), N-Acetyl-D-glucosamine (GlcNAc), D-galactose (Gal), N-Acetyl-D-galactosamine (GalNAc), D-mannose (Man), L-fucose (Fuc), N-Acetylneuraminic acid (Neu5Ac), D-xylose (Xyl), and D-glucuronic acid (GlcA). Once incorporated into a glycan structure, these building blocks may suffer further modifications; this is the case of GlcA that can be epimerized at carbon 5 and generate the tenth monosaccharide: iduronic acid (IdoA).

The most common text abbreviation for monosaccharides is the one reported above using three letters and concerning their visual representation, the pictorial notation universally recognized is called Symbol Nomenclature for Glycans (SNFG), and assigns to each component a precise shape and a colour (Figure 1.1).

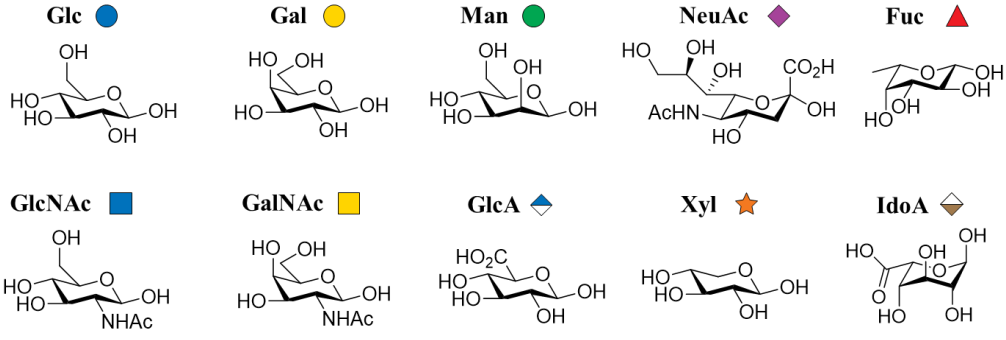


Figure 1.1. Structure of the ten main recurrent monosaccharides in the mammalian *glycome* with their three-letter abbreviation and the SNFG symbol.

Thus, the observed variety in nature does not strictly depend on the available number of sugar building blocks, but it is rooted, instead, in the chemical complexity of the monosaccharides and in the different nature of the linkages that join them together (Figures 1.2 and 1.3).

A **monosaccharide** is formally composed by carbon atoms and water molecules and has the general empirical formula $C_x(H_2O)n$, where *n* ranges between 3 and 9, even if the most represented members are pentoses (5) and hexoses (6).
The carbon chain can cyclize giving rise to rings of different sizes (hemiacetals). Once cyclized, the sugar can experience conformational flexibility, depending on the type of cycle. The adopted ring conformation determines the orientation of the functional groups in the three-dimensional space. [5]

Moreover, the orientation of the hydroxyl group in the anomeric position (C1) can be axial (α) or equatorial (β) with respect to the plane of the sugar ring, generating the possibility of two different epimers. The stereogenic anomeric center can undergo an interconversion of epimers in solution, a phenomenon dubbed mutarotation: the monosaccharide ring opens up (aldehyde, more unstable) and then recloses adopting the other anomeric configuration (Figure 1.2). [6,7] Generally, the β epimer is the most populated conformation, favoured by steric effects. However, in many cases the α epimer has a significant representation, due to the anomeric effect. [8,9]

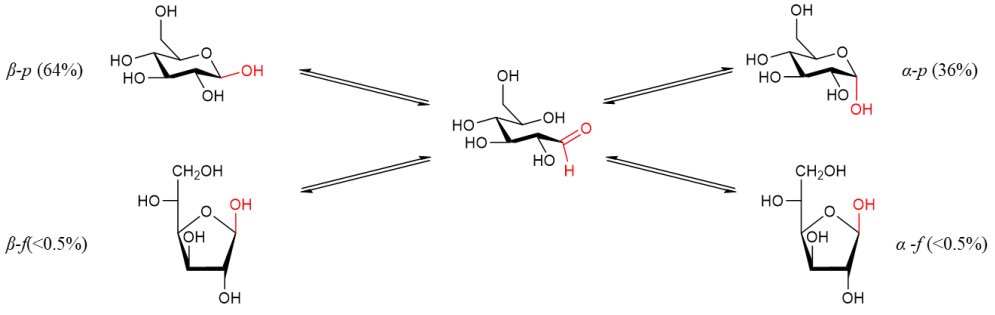


Figure 1.2. Elements of structural variability of saccharides: cyclization process for D-Glc and anomeric configurations.

More monosaccharides are covalently joined together through the glycosidic bond. This fundamental linkage is formed between the anomeric carbon of one monosaccharide and a hydroxyl group of another. A single sugar ring can be attached, in different positions, to multiple other sugar units, allowing not only the formation of linear molecules, but also of branched structures. Additionally, each anomeric carbon can generate both α and β configuration of the glycosidic linkage, thus further increasing the chemical and structural complexity (Figure 1.3).

The type of linkage strongly influences the spatial orientation of each sugar moiety and consequently tunes the physical and chemical properties of the whole molecule as well as its functional activities.

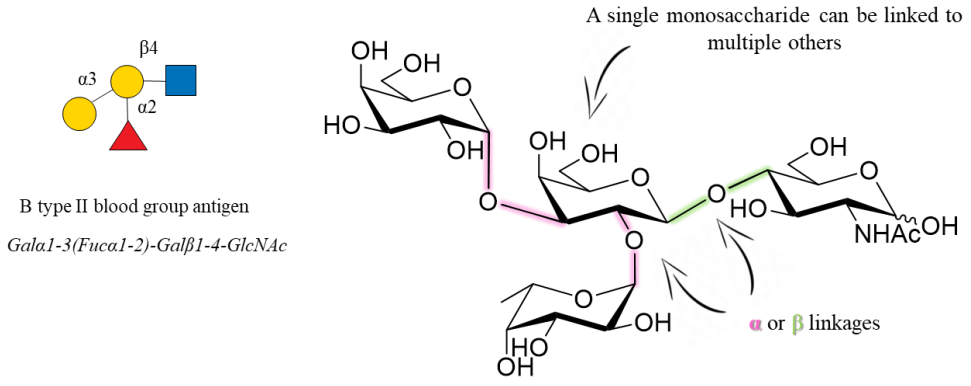


Figure 1.3. Elements of structural variability of saccharides: different types of glycosidic linkages. Example of the B type-II blood group antigen. The monosaccharides can be connected together throug α or β linkages and one unit can be connected to multiple other ones through linkages at different positions.

All the aforementioned variables generate an ideally unlimited assortment of possible oligo- and poly-saccharides in term of shapes and geometry, making the glycans much more variegated than other biopolymers (proteins, nucleic acids). [10]

1.1.2 Mammalian glycosilation

As already mentioned, glycans in cells are mostly found covalently conjugated to other biomolecules. In particular, carbohydrate chains of different length and complexity can be attached to proteins (**glycoproteins** and **proteoglycans**) or to lipids, usually sphingolipids (**glycosphingolipids**). The chemical process that leads to the final glycoconjugate is called glycosylation.

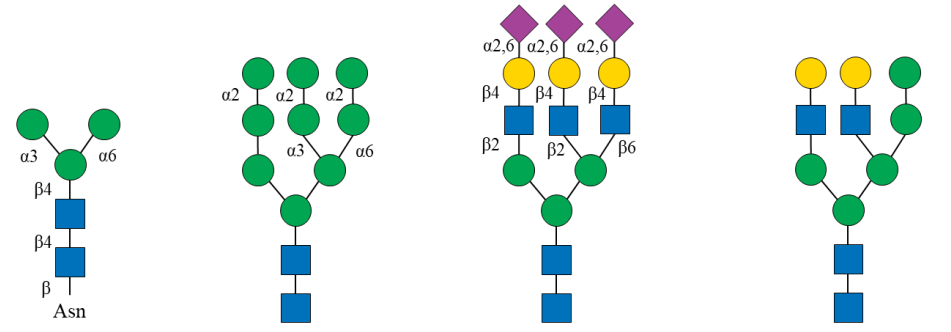
Protein glycosylation is one of the most abundant post-translational modifications (PTMs) found in eukaryotic cells. It has been estimated that a range from 50% to 70% of the entire mammalian proteome is glycosylated. [11] The addition of sugar chains in one or multiple sites of a protein is not just an ornament and has various implications. In fact, glycans contribute in regulating the physicochemical properties of a protein, such as its folding, stability, and solubility and, moreover, they also modulate the interaction of the protein with other partners. [12]

The study of glycosylation is extremely complicated since the structures generated by this event are intrinsically heterogeneous and strongly depend on the availability of the building blocks, cell type, and physio-pathological conditions. This aspect is a consequence of the fact that the attachment of sugars to proteins is a non-template-driven biosynthesis. [13]

Two main classes of glycosylation occur in mammalian cells: N- and O-glycosylation. They take place in diverse cellular compartments and are composed of different steps, but their nomenclature derives from the protein residues to which the sugar is covalently attached. In fact, N-linked glycans target the nitrogen atom of asparagine (Asn) residues, while the O-linked ones target the oxygen atom of serine/threonine (Ser/Thr) residues. The knowledge of mammalian N-linked and S/T-linked glycosylation mechanisms is the most advanced, however other types of glycosylation have been found. O-mannosylation, O-fucosylation, O-glucosylation, C-mannosylation (which targets the carbon of a tryptophan (Trp) residue) and glypiation are some other examples of the 14 distinct pathways present in mammalians, which are distinguished on the basis of the sugar–protein linkage and the initial monosaccharide. [14,15]

N-glycosylation starts in the lumen of the endoplasmatic reticulum (ER) and targets an Asn residue contained within a defined consensus sequence (or sequon): Asn-X-Ser/Thr, in which X is not proline (pro). This sequon is the minimal requirement for N-glycosylation, but its presence in the protein sequence does not necessarily mean that the site will be subjected to glycosylation. [16–18]
As first step, a sugar precursor of 14 monosaccharides ($Glc_3Man_9GlcNAc_2$) is covalently attached to the Asn side chain. At this moment the protein is still unfolded. In the ER, the original N-linked glycan structure is modified by removal of the terminal GlcNAc residues (trimming). During this process, which is an important checkpoint of protein quality control, folding is accomplished. [19] Subsequently, the complex is translocated in the Golgi apparatus where suffers re-glycosylation (processing) with additional sugar residues, such as GlcNAc, Man, Fuc, Gal, or Neu5Ac. [20] This final decoration generates the exquisite glycan diversity. All eukaryotic N-glycans share a

common core of 5 residues, Manα1-3(Manα1-6)Manβ1-4GlcNAcβ1–4GlcNAcβ1–Asn-X-Ser/Thr, and are classified into three types: high mannose (high-Man), complex-, and hybrid-type N-glycans (Figure 1.4). For *high-Man N-glycans*, the core is only extended with Man residues. In the case of *complex-type*, a variable number of antennae (branches, arms) initiate with a GlcNAc residue from the main core. For the *hybrid-type*, a Man residue starts the elongation of the Manα1-6 arm, while the Manα1-3 arm starts its extension with a GlcNAc moiety.



Conserved core structure **High mannose** **Complex structure** **Hybrid structure**

Figure 1.4. From the left: the core structure conserved among N-glycans and the structures of the different types of N-glycans found in mammalian cells (high-Man, complex-, hybrid-type).

In the case of **O-glycosylation**, there is no specific sequon. The most common type of O-glycosylation is the **mucin-type** (Muc-type), in which the first residue to be attached to hydroxyl group of Ser/Thr is GalNAc. O-GalNAc glycans are added to proteins in the Golgi apparatus. Unlike N-glycosylation, there is no trimming processes, but the residues are linked in sequence moving from the *cis*-, *medial*-, and *trans*-Golgi compartments. [21] The O-GalNAc glycans of mucins have *eight major core structures* (Figure 1.5). Each core can be extended by a variety of sugar residues (GalNAc, Gal, GlcNAc, Fuc, and Neu5Ac, whereas Man, Glc, or Xyl are never present herein) to generate linear or branched chains. Moreover, Neu5Ac sugars can be modified by O-acetylation and Gal and GlcNAc, by sulfation. The total length of mucin-type glycans varies from a single O-GalNAc (the Tn-antigen) to more than 20 sugars. The blood group antigens exposed on the surface of human erythrocytes are commonly found in mucins and are examples of extended cores. [22]

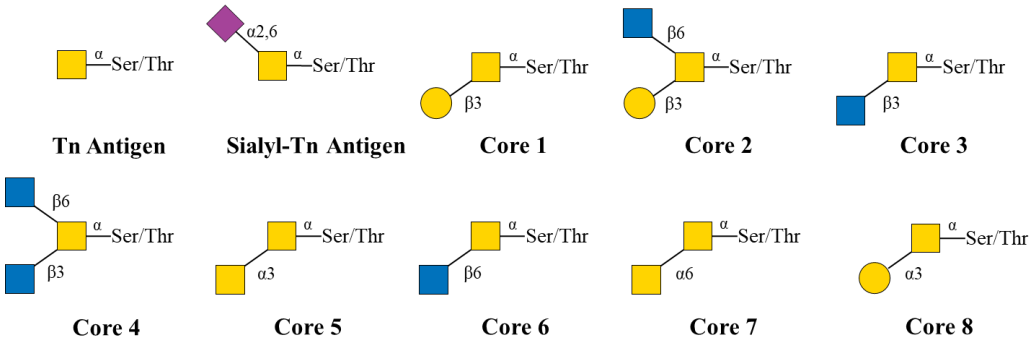


Figure 1.5. Mucin type O-glycosylation: the core structures.

Proteoglycans are a special class of glycoproteins prevalently found in the extracellular matrix and on cell surfaces that, in addition to canonical N-glycans and O-glycans, they also bear long sugars chains attached via O-linked glycosylation. These motives, called **glycosaminoglycans (GAGs)**, can easily contain more than 80 sugars (until 200) and are constructed with disaccharide repetitions. On the basis of their composition, they can be classified into three major groups: heparin/heparan sulfate, chondroitin/dermatan sulfate, or keratan sulfate (Figure 1.6). Since GAGs are often N- or O-sulfated, they are negatively charged entities, a characteristic that not only strongly drive their interactions with the surroundings but, due to their size and density, also modifies the chemical-physical features of the microenvironment. [23]

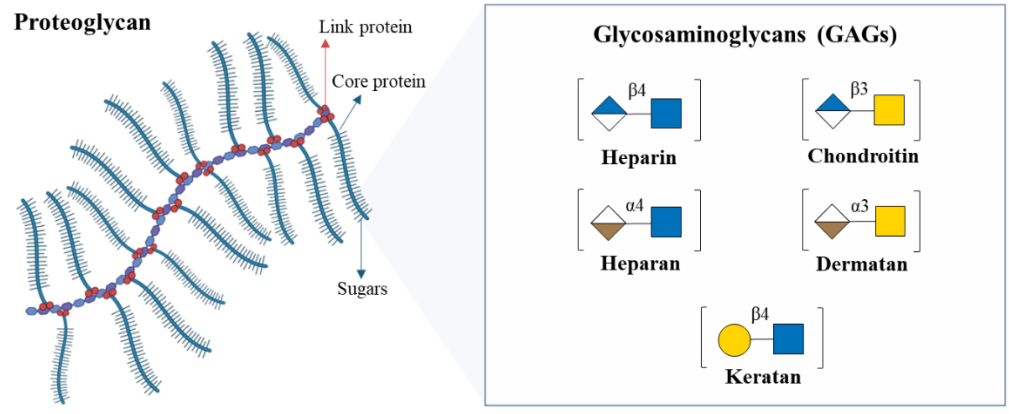


Figure 1.6. On the left, representation of a proteoglycan full structure. On the right, GAG motifs of the three major groups (heparin/heparan sulfate, chondroitin/dermatan sulfate, and keratan sulfate)

Sugar residues can also be liked to lipids, in particular to sphingolipids, generating **glycosphingolipds (GSLs)**. They are located in the cellular membrane with their glycans facing the external *milieu* and the lipid embedded into the membrane. GSLs are synthesized in the ER and Golgi by sequential attachment of single sugars units to a ceramide lipid core (sphingosine linked to fatty acids). [24]

The first monosaccharide to be added is generally Glc (GlcCer), followed by Gal, to give LacCer. The type of linkage and sugar that decorate the initial LacCer unit is used to define the classification of GSL. There are five main subtypes: ganglio, lacto, neolacto, globo and isoglobo (Figure 1.7). [25]

GSLs are amphiphilic molecules, since are constituted by a hydrophobic lipid moiety attached to the hydrophilic carbohydrate portion.

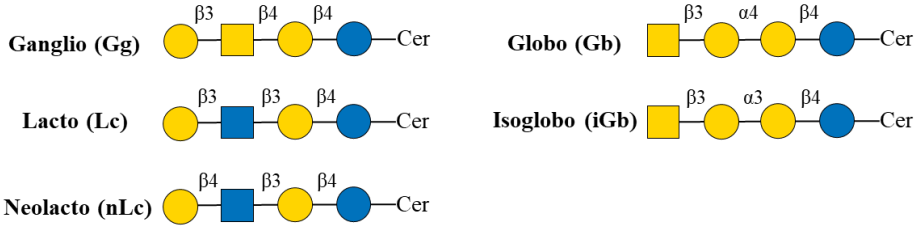


Figure 1.7. The five main subtypes of glycosphingolipids: ganglio, lacto, neolacto, globo, and isoglobo.

More recently, the glycosylation of small RNAs has also been postulated. The glycoRNA has been proposed to be exposed on the surface of cells and to be glycosylated by the same machinery in charge of N-glycosylation. [3] Nevertheless, further investigations have to be performed in order to decipher many aspects, such as the chemistry that links the RNA to the carbohydrate part, the detailed biosynthetic pathway, and the biological significance of this new glycoconjugate.

The glycosylation machinery of a given cell determines the ensemble of glycan structures and the types of possible glycoconjugates.

The biosynthesis of the enormous assortment of glycans found on mammalian glycoproteins, proteoglycans, and glycolipids, involves more than 200 different enzymes operating in 16 major pathways (14 for the generation of glycoproteins and 2 for glycolipids). [26] Of these enzymes, at least 173 are **glycosyltransferases** (**GTs**).

A representation of the current atlas of their roles in glycan biosynthesis is reported in Figure 1.8. [27]

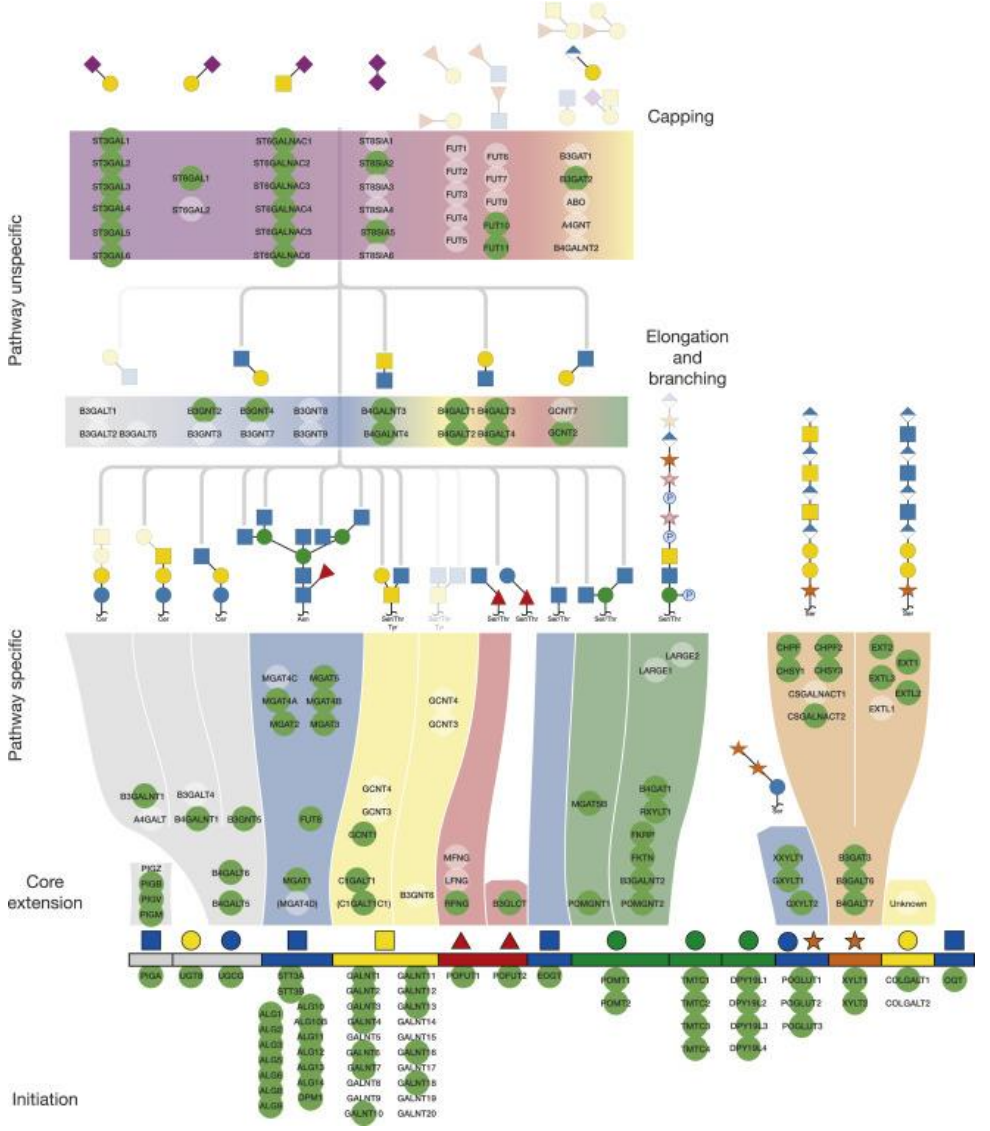


Figure 1.8. Atlas of human glycosylation pathway maps with the functions of 169 GTs. Rainbow depiction of the human glycosylation pathways, from left to right: GPI-anchor, glycolipids (2), N-linked glycans, O-GalNAc mucin-type, O-Fuc type (2), O-GlcNAc type, O-Man type (2), C-Man type, O-Glc type, O-Xyl type (proteoglycans), O-Gal type and O-GlcNAc type (cytosolic). Glycosyltransferase genes are arranged in the pathway-specific steps (initiation and core extension) and in the pathway-nonspecific steps (elongation and capping, 52 genes). Figure taken from Narimatsu et al., 2019. [27]

The glycosylation capacities of different cell types vary primarily through changes in the expression of the GTs. Some steps in the glycosylation pathways are regulated by

a single GT and the loss of function of such gene leads to global changes in the glycan structures produced. These steps are dubbed non-redundant steps. An example is the core α6-fucosylation of N-glycans originated by Fut8 GT. Generally, non-redundant GTs covers the initial core extension steps.

On the contrary, some other glycosylation steps are regulated by multiple isoenzymes with at least partially redundant functions (redundant steps). In this case, the loss of function of a single isoenzyme gene leads to subtle effects (or none) on the final glycan structure. Elongation, branching, and capping phases are covered by partial redundancies. For instance, sialylation is covered by partial redundancies and combinatorial KOs of three genes are required to selectively eliminate α3-sialylation on N-glycans (KO of *st3gal3/4/6*). In contrast, KOs of just two genes are required to selectively eliminate α3-sialylation of core1 O-glycans (KO *st3gal1/2*). [28,29]

The glycosylation patterns are altered in several human diseases, such as cancer, auto-immune, infectious, and chronic inflammatory diseases. [30,31] For instance, the extension of O-GalNAc glycans beyond the first sugar is blocked in some cancer cells, generating an aberrant glycosylation phenotype of the mucins exposed on the cell surface. [32] Similarly, aberrant sialylation (hyper-sialylation) is an established hallmark of several types of cancer directly linked to tumour invasion, migration, and immune evasion. [33]

Moreover, genetic defects in genes involved in the glycosylation (glycogenes) are often lethal at embryonic stages, highlighting the vital role of the glycoconjugates. Sometimes these mutations are compatible with life but generate acute pathologies, globally defined as congenital disorders of glycosylation (CDGs). [34,35]

1.1.3 Distribution

Glycans are as universal and present in all life kingdoms (as nucleic acids, proteins and lipids) and their functions are essential to the existence of all known living organisms. [36]

Focusing on mammalian cells, the location of such glycoconjucates is ubiquitous: they have been found in the nucleus and in the cytoplasm, where they are not only mere

structural components, but modulate complex interplays such as the activation-inactivation switch of proteins dictated by phosphorylation. [37–41]

However, the majority of glycoconjugates is located in the cell surface and in the extracellular space. As a matter of fact, every mammalian cell is coated with a dense layer of membrane-bound proteoglycans and glycoproteins, the *glycocalyx*. [42] This hydrated gel-like structure also includes extracellular components, such as hyaluronic acid and soluble proteoglycans (Figure 1.9). [43]

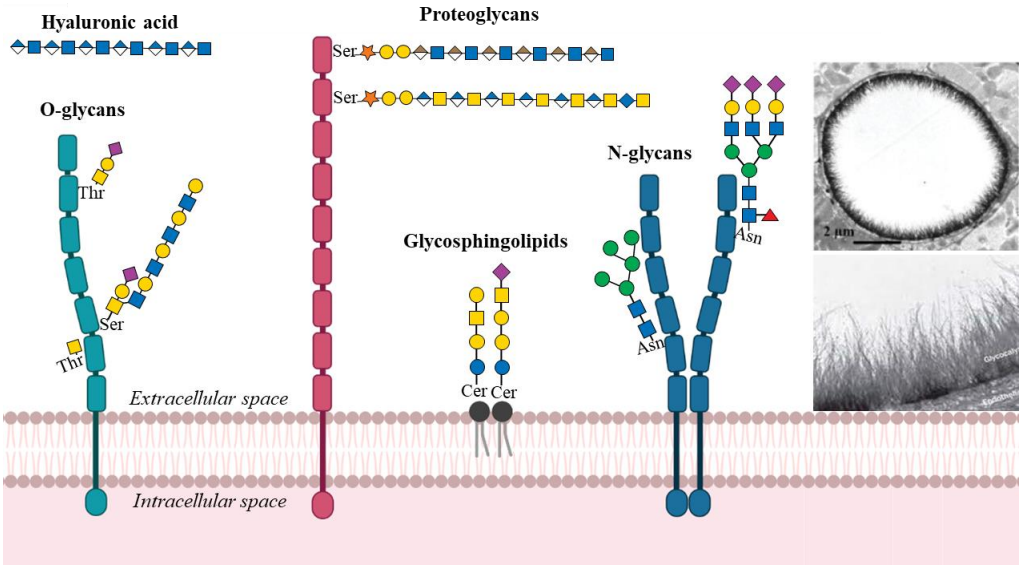


Figure 1.9. Representation of the main components of the *glycocalyx*. On the right: electron microscopic pictures of the glycocalyx of myocardial capillary endothelial surface of rat (taken from Van den Bert et al., 2003) [44]

In homology with the skin, which is the body's first level of defence, the *glycocalix* constitutes the very first physical barrier for cells. Pathogens are not free to penetrate inside the cells, since they found this dense structure as initial obstacle.

The idea that the *glycocalyx* was just a protective and passive coat with no regulative activity was installed for many decades. With the advent of new techniques to visualize the organization and decipher the roles of this complex structure, it became clear that the *glycocalyx* could be considered as an organelle actively involved in processes related to cell communication. [43,45,46] The possibility to visualize this structure at increasing levels of precision and under different physiological conditions remains a hot topic of investigation. [47,48]

Indeed, the *glycocalyx* is directly implicated in vital cellular processes, since its components are continuously interacting with the surrounding. Surface interactors belonging to the *glycocalyx* finely regulate cell-cell, cell-matrix, and cell-pathogen interactions. In fact, surface glycoproteins (and glycolypids) interact with receptors located in **neighbour cells** as well as with circular **antibodies** and **lectins**, a class of proteins in charge of recognizing sugar epitopes (Figure 1.10). Moreover, *cis*-communication, with receptors located at the same cell surface, is also a possibility.

All these molecular interactions trigger a plethora of different biological responses, such as regulation of the immune system, inflammation, changes in cell morphology, development, apoptosis, and disease insurgence. [49] Indeed, the Nobel laureate in 2022 has been awarded to Prof. Carolyn Bertozzi for developing exquisite biorthogonal chemical tools to investigate molecular interactions involving glycans. Additionally, membrane-anchored glycoproteins also interact with external **pathogens** (bacteria, viruses) and facilitate or inhibit the invasion of the host cell (Figure 1.10). [49]

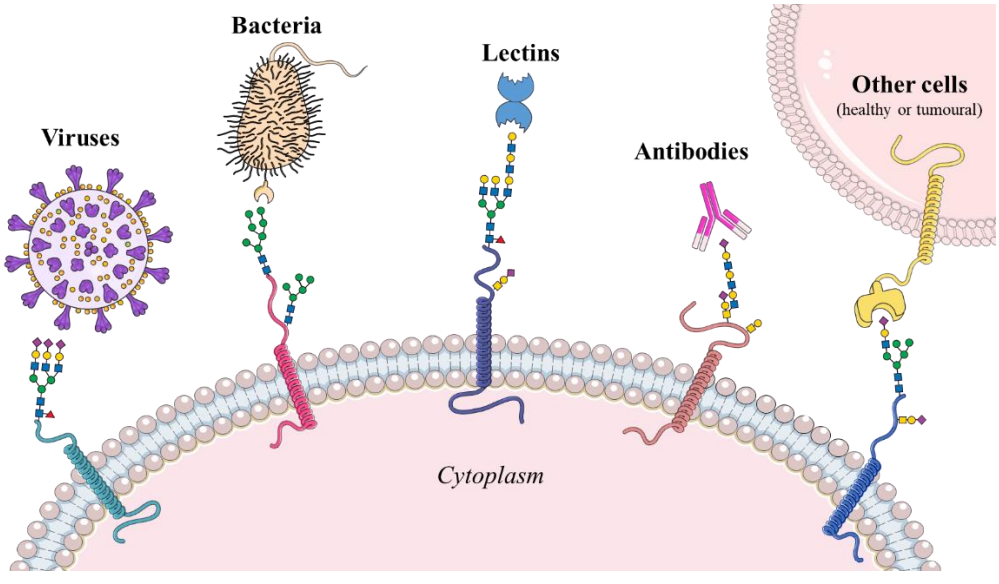


Figure 1.10. Cartoon representation of the main interactions mediated by glycoconjugates on the cell surface.

Any change in the glycocalyx composition in response to a different cell state can promote dissimilar recognition processes. [50] Considering this complex interplay easily linked to disease, it is evident that the interactions and actors involving glycans and their molecular counterparts can be chosen as targets in diverse therapeutic contexts. [51]

1.2 Glycan-lectin interactions

Glycans interact intra- and extra-cellularly with many types of proteins including enzymes, antibodies, and lectins. The recognition of glycans mediated by glycan-binding proteins (GBPs) promotes crucial biological responses. These proteins display no catalytic activity, as enzymes do, and are not linked with a direct immune response, like antibodies are.

GBPs can be divided into two large families: lectins and sulfated GAG-binding proteins. This last class is extremely heterogeneous and does not share a common evolutionary origin or structural features. [52] The discussion of this chapter will only be focused on lectins, which have been the topic of this Thesis.

1.2.1 General aspects of sugar-protein molecular recognition

There are some recurrent weak forces and thermodynamic parameters that regulate the binding of glycans to the usually solvent exposed and well-organized binding sites of lectins. The knowledge of their characteristics is of paramount importance, not only for understanding the mechanism behind the recognition event, but also for developing mimetics of sugars with enhanced affinity. [53,54]

The interactions established between sugars and lectins are of non-covalent nature and, therefore, intrinsically weak (Figure 1.11). Glycans are extremely polar molecules. Indeed, they possess several hydroxyl groups that can establish **hydrogen bonds** (**HB**s) with the residues of the protein, acting as donors, as acceptors, or both (cooperative). Among the amino acids, there are some that have been found more often involved in HBs; this is the case of residues with planar sidechains capable to bifurcate the HB or of residues able to act both as acceptor and donor (Asp, Gln, Arg, His). On the contrary, amino acids with more flexible geometry (Ser, Lys) are less represented, due to fact that the entropy associated to the binding is increased: the rotamers of the sidechains have to be fixed, in addition to the conformational selection of the sugar hydroxyl rotamer. [55] The energy of a single HB interaction typically ranges between 1 and 2 kcal/mol. Moreover, stabilizing HBs can also be mediated by water molecules located at the binding site.

Furthermore, some lectins, like the C-type lectins, show one or more **bivalent cations (Ca^{2+}, Mn^{2+})** in their binding sites, which not only stabilizes the protein in its *apo* form, but also coordinate the vicinal hydroxyl groups of the sugar ligand. The presence of these ions is a necessary condition for the binding to occur.

Although sugars are universally defined as polar compounds, depending on the spatial orientation of the hydroxyl groups in the ring two different faces can be defined: a highly polar one and a less polar one. For this reason, carbohydrates can also be described as amphiphilic molecules, with a dual character. Accordingly, they can generate hydrophilic-based contacts with the surrounding (as HBs), but also non-polar interactions. In line with this aspect, additional contributions stabilizing sugar-lectin binding events can arise from **stacking interactions** (CH-π stacking) between the less polar face of the sugar ring and aromatic sidechains of the protein. CH-π stacking interactions occur when the carbohydrate ring is parallel to the plane of the aromatic ring of the amino acid. [56–58] The residues found more often involved in these interactions are Trp, Tyr and Phe. [57] **Van der Waals interactions** also involve non-polar sidechains (Ile, Leu, Ala, Val, Phe) in close contact with non-polar areas of the sugars. The contribution of one CH-π stacking or van der Waals interaction is ca. 1.5 kcal/mol. [59]

Ionic electrostatic interactions can be established between sialic- and sulfate-bearing sugars and positively charged residues (Arg, Lys, His), giving rise to **salt bridges**. These contacts are strongly stabilizing since may provide more than 2 kcal/mol and. For instance, siglecs possess an extremely conserve Arg residue in their binding pockets that establishes a crucial salt bring with the carboxylate of the sialic acid.

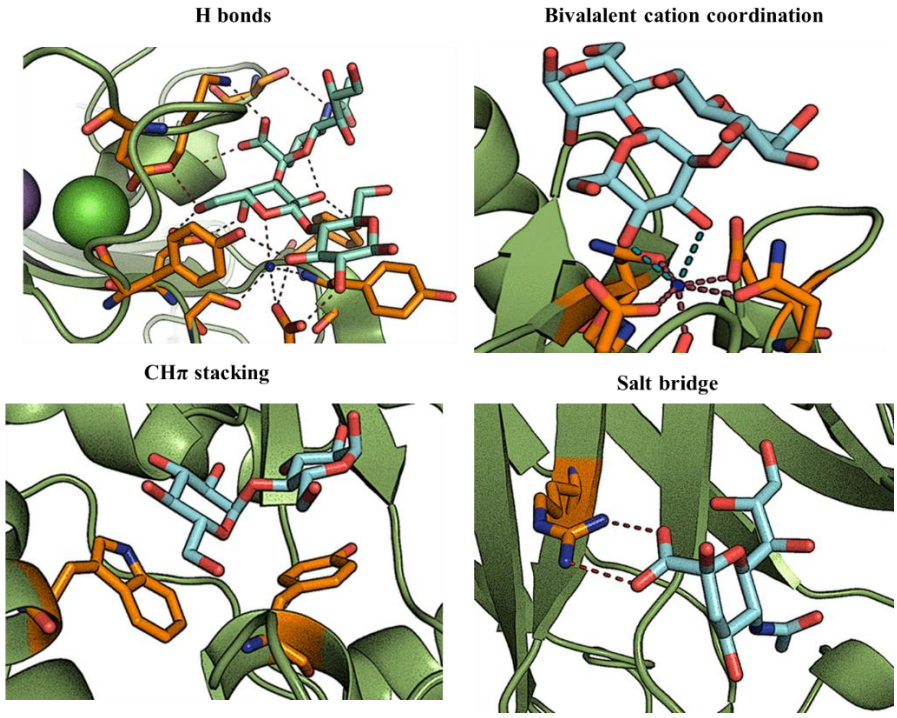


Figure 1.11. Key forces and interactions occurring in lectin-sugar binding events

The thermodynamic behind glycan-lectin recognition is usually governed by the paradigm of **enthalpy/entropy compensation**. This mechanism is observed only for intermediate-weak interactions (such as those occurring between sugars and proteins), where the entropy term value is similar to the enthalpy one. On the contrary, for strong interactions, where the enthalpy contribution is far greater than the entropy term, compensation is no longer observed. The enthalpy/entropy compensation event is very complex and does not easily fit in models of general validity. Preferably, a case-by-case approach should be adopted to interpret this phenomenon, since too many variables depends on the precise features of the system under study. [60–62] However, some common rules can be considered. In fact, since glycans are generally flexible molecules, a typical entropy penalty is observed arising from the loss of degrees of freedom upon ligand conformational selection. A pre-organization of the glycan can lead to decreased entropic cost. In some cases, the increase of protein flexibility upon ligand binding has been described to correlate with a favourable entropy.

The enthalpy can be considered a quantitative indicator of the changes in intermolecular bond energies (hydrogen bonding, van der Waals interactions, salt

bridges etc) occurring during the binding process. Moreover, both enthalpy and entropy terms are affected by changes in the solvation. Generally, water desorption (desolvation) is linked to unfavourable enthalpy and favourable entropy. The enthalpic penalty in this case is considerable, since sugars are highly solvated polar molecules and the binding site of lectin is shallow and exposed to water. Since the favourable interactions established between sugars and lectins provide low energy, their contribution to enthalpy is rather low and insufficient to compensate the enthalpic desolvatation penalties. Consequently, the sugar-protein association process is weak and characterized by **low affinity**. Indeed, K_Ds usually ranges from mM to µM.

A question arises: how can those weak interactions so finely determine critical biological responses?
The answer can be found in the engagement of simultaneous interactions involving the receptor and/or the ligand, a phenomenon called **multivalency**. [63] Multivalency can be described as the contemporary availability of several replicates of one or two components of the system (lectins and sugars) so that they can interrelate in a cooperative manner. In this way, individual weak affinities are simultaneously engaged in a synergic way, thus enhancing the global affinity by orders of magnitude. This event is defined **multivalence effect** and the accumulated strength resulting from individual affinities performing concurrently is known as **avidity**.[64]

Multivalent receptors interacting with multivalent ligands generate an extremely complex binding event and often a combination of different effects are responsible of the observed affinity increase (Figure 1.12)
However, discerning the contribution of each mechanism of interaction is complicated. When multiple lectin binding sites binds simultaneously to multimeric ligands, the event is defined as **chelate association,** which highly increases the avidity. Furthermore, in the local microenvironment in which these interactions take place, the resident concentration of monomeric receptors on the cell surface is high: the local density extremely favours **clustering effects**, strengthening the binding event.
The **subsite association** occurs when a second binding site with different affinity and specificity also interacts with a multimeric ligand with diversified epitope presentation.

Multivalent sugars can also interact with monovalent lectin triggering affinity improvement through a simple higher density of ligands available in close proximity of the binding site. This phenomenon is known as **statistical rebinding**. [64–67] Additionally, lectins can experience internal diffusion between multivalent glycans before dissociation, a phenomenon which, again, increases the affinity and is known as **bind and slide mechanism**. [68]

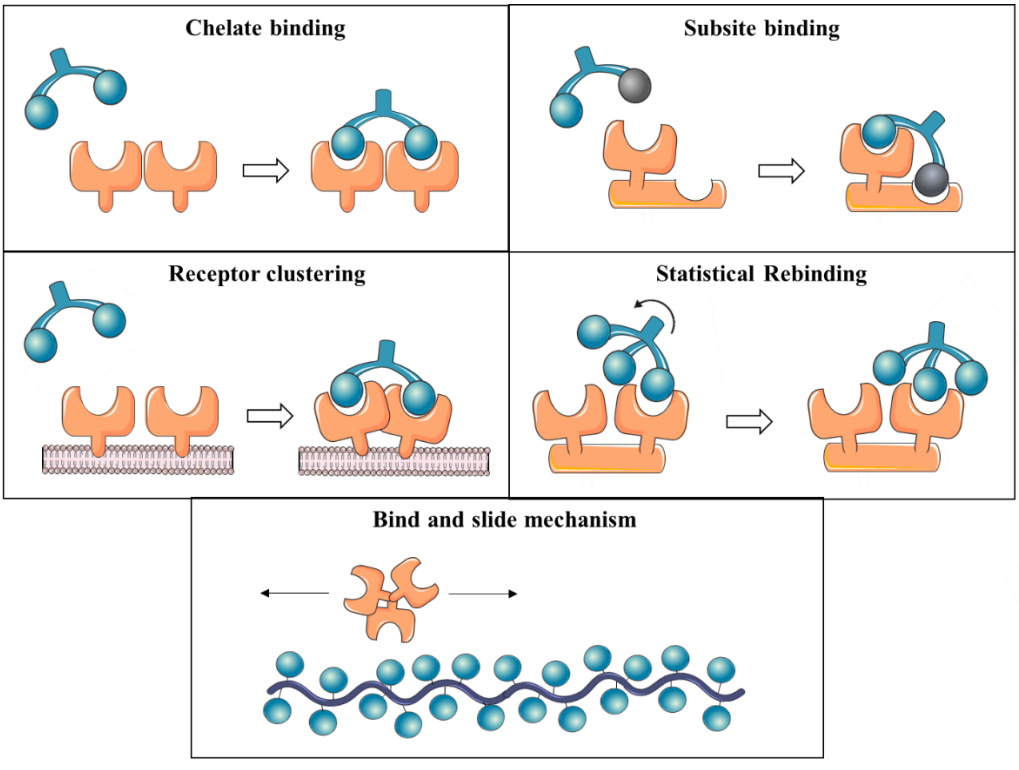


Figure 1.12. Multivalency in action: cartoon representation of the major mechanisms occurring with multiple lectin and/or sugar presentations.

The multivalent effect has also intermolecular implications. In fact, aggregative binding catalysed by **cross-linking effect** can generate large noncovalent three-dimensional networks. If the number of epitopes increases, two- or three- dimensional networks can be produced, which are often poorly soluble and undergo irreversible precipitation (Figure 1.13).

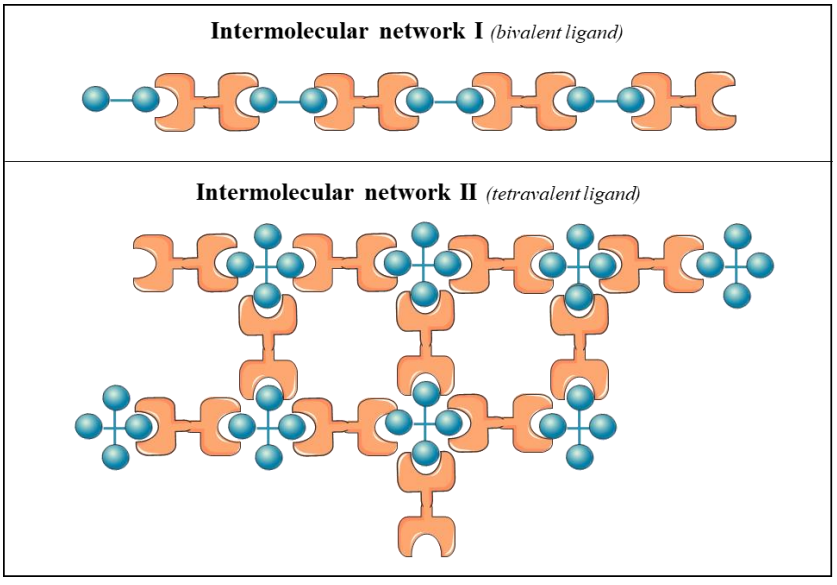


Figure 1.13. Intermolecular networks generated by cross linking effects.

The deep understanding of carbohydrate-protein binding processes is certainly important for fundamental research, but also for biomedical applications. Undeniably, the design of compounds able to target lectins with high affinity and selectivity can benefit the knowledge about basic sugar-protein binding events and exloit it to build inhibitors. For instance, lectins inhibitors with hydrophobic substituents have been synthetized in order to overcome the entropic barrier. [65] Also, a large and diverse number of synthetic multivalent glycans (neoglycopeptides, neoglycoproteins, glycodendrimers, glycopolymers, glyconanoparticles, and glycoliposomes) have been proposed. [67,69]

1.2.2 The human lectins

Lectins were discovered for the first time in plants in 1888, and only in the late 1960s the first animal lectin was isolated. They are found ubiquitously in virus, microorganism, plants, and animals. They can be classified into evolutionarily-related families based on structural similarities. At least 13 structural families are known to exist in animals. Although their main binding partners are self or non-self glycans, many animal lectins also establish protein–protein, protein–lipid, or protein–nucleic acid interactions. [70,71] Their cellular and extracellular localization as well as their biological roles are manifold and will be discussed case-by-case, with particular emphasis on the three major human lectins: galectins, C-type lectins, and siglecs.

1.2.2.1 Galectins

Galectins, earlier named S-type lectins, are the most widely expressed class of lectins in all organisms. [72] In mammalians, 16 member have been isolated up to now, numbered consecutively by order of discovery. They can be broadly be defined as β-galactoside binding lectins, since the main epitope recognized is the β-Gal ring present into most complex glycan structure. Galectins display either one or two conserved carbohydrate recognition domains (CRDs) of ca. 130 amino acids (≈ 15 kDa). On the basis of the supramolecular organization of CRD, galectins can be distinguished into:

- **Prototype**, when a single CRD spontaneously forms a non-covalently bound homodimer (human galectins-1, -2,- 5, -7, -10, -11, -13, -14, -15, -16).
- **Tandem-repeat type**, when two distinct CRDs are connected by a short peptide linker of up to 70 residues long (human galectin-4, -6, -8, -9, -12).
- **Chimera type**, when the CRD is connected to an N-terminal amino acidic non-lectin tail, which is rich in Pro, Gly, and Tyr residues and promotes oligomerization. The only member of this class is galectin-3. (Figure 1.14). [73,74]

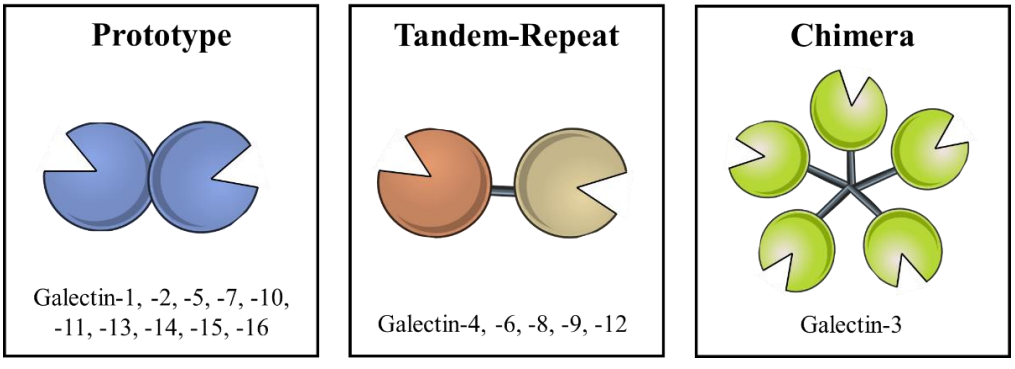


Figure 1.14. The three main type of galectins: prototype, tandem-repeat and chimera. The member of each group are reported.

This classification also includes some galectins that are species-specific, such as galectin-5 and galectin-6, galectin-11, and galectin-15 which have only been found in rodents, sheep, and goats, respectively.

The crystal structure of the majority of the individual CRD of galectins has been solved, both in the *apo* form and bound to different glycans.

The general **structure of the CRD** is conserved among the members of the family and is composed of six (from S1 to S6) and five (from F1 to F5) antiparallel β-strands organized in the jellyroll folding pattern, which is slightly bent and forms a groove on the concave side. This groove, located in the S-face, constitutes the recognition site for sugars and can be sectioned into 5 minor sub-sites (A-E), being the bound β-Gal moiety always bound on sub-site C (Figure 1.15). The canonical binding site for sugars is shallow and solvent exposed. Besides sub-site C, other the sub-sites interact with the long saccharides, thus defining the fine specificity for larger sugars.

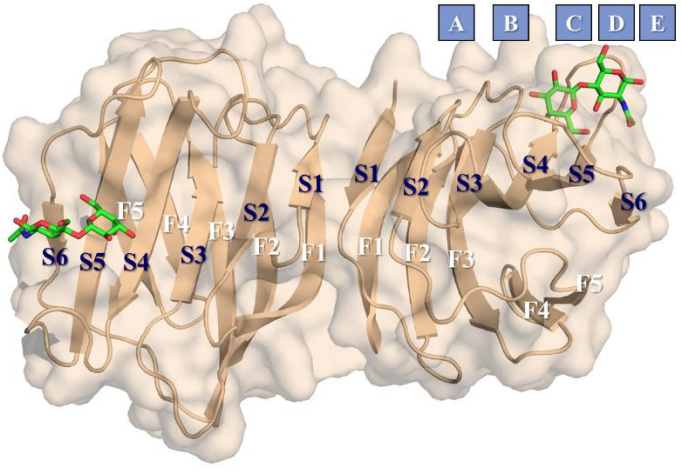


Figure 1.15. Structure of the CRD of galectin-1 in complex with N-acetyllactosamine (PDB ID: 1W6P) with sub-sites (A-E) and beta strands (S1-S6, F1-F6) highlighted.

The **sequence identity** between galectins is moderately high. Even though some key residues in the binding site are highly conserved, the neighbouring sub-sites (A, B, D, E) display more differences, thus modulating the glycan preferences among galectins (Figures 1.16). Many galectins contain a variable number of free Cys residues, whose redox state is related to their stabilities and their binding activities

The **binding** to glycans occurs through HBs, van der Waals interactions, and through a key CH-π stacking interaction between the non-polar face of the β-Gal moiety and an extremely conserved Trp residue (located in the S6 strand). Other conserved amino acids are His, Asn, and Arg moieties at the S4 strand, which establish HBs with the hydroxyl group at position 4 of the bound β-Gal, while a key Asn on the S5 strand interact with the hydroxyl group at position 6 (Figure 1.16). The canonical interaction in the case of galectin-1, for instance, occurs through HBs involving the hydroxyl

groups of the Gal moiety and His44, Asn46, Arg48, Val59, Asn61, Glu71, and Arg73, as well as through the CH-π stacking interaction between the β-Gal ring and Trp68 (Figure 1.16). Moreover, galectin-1 also displays a loop in the binding site (L4 loop, in sub-site D), which connects strands S4 and S5. This is a unique feature of galectin-1, as example of the variance among galectins in the neighbouring sub-sites. [75]

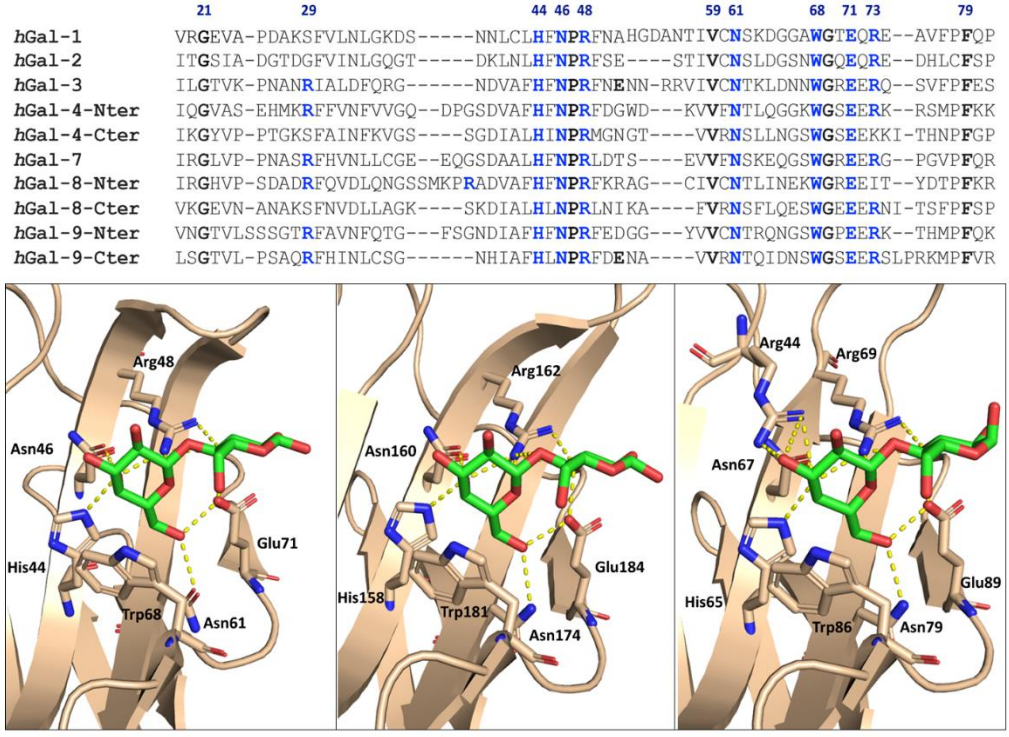


Figure 1.16. Above: sequence alignment of the binding site of galectins. The conserved residues are highlighted. Below: binding site of galectin-1 (PBD ID: 1GZW), galectin-3 (PBD ID: 1NN8) and galectin-8-N-terminal (PDB ID: 5T7S) in complex with lactose. Focus on the main interactions that stabilize the binding: the residues of the protein establishing contacts are displayed as sticks, while hydrogen bonds are displayed as yellow dashed lines. Figure taken by Bertuzzi et al., 2020. [76]

Galectins binds to Gal with **K_Ds** in the mM range, while the **affinity** for simple β-galactosides, such as the N-acetyl-lactosamine, disaccharide (LacNAc) reaches μM values. Since the natural multivalency of these lectins and their oligomeric states strongly influence their binding features, the affinity for natural glycoconjugate is in the sub-μM range.

Although the panel of binders for each galectin in the natural microenvironment is far from be completely understood, a large number of studies employing different

methodologies have characterized their molecular recognition features with fragments or isolated glycans. [77–80] β-Gal motives, and especially LacNAc (Galβ1-4GlcNAc), are commonly found at the end of mammalian glycoproteins and glycosphingolipids. All galectins bind to LacNAc; however, subtle modifications of this structure have dissimilar impact on the recognition event, depending on the type of galectin. In fact, each member displays exquisite **preferences**. [77]

A selection of key biologically relevant binding preferences of galectins is given here:

- **Sialylation** is one of the elements that generates binding diversity among galectins. Glycans are often capped with either α2,3 or α2,6 linked sialic acid. α2,6 sialylation prevent galectin binding without relevant distinctions among members; indeed, it is considered as a "switch off" mechanism for galectins. Only galectin-3 seems to tlightly tolerate this modifications, probably due to its propensity to bind internal LacNAc units. [80,81] On the contrary, α2,3-sialylated β-galactosides are weakly recognized by galectin-1 and -9C, while display a particularly high affinity only for galectin-8N. [80,82–84].
- Galectin-8N is the only galectin with preference for **Galβ1-4GalNAc** over Galβ1-4GlcNAc, leading to a unique recognition of core-1 O-glycans. Furthermore, the α2,3 sialylation of this epitope allows to reach affinity in the low μM range. [82,84,85]
- The **histo blood group antigens** are recognized with different affinities by galectin-1, -3, -4, -7 and -8. [85–89] This class of antigens are good binders (low μM range) of galectin-3, -4, -7, and -8 and it has been suggested that pathogens which use molecular mimicry for infection can be recognized by galectins as first step to eradicate the infection. [90] Contrarily, galectin-1 is the worst binding partner of these epitopes. The details of these interactions are unravelled in Chapter 3 of this Thesis.
- Epitope repetition also strongly influences the galectin binding. Since in nature polymeric glycans are widely present, the interaction of galectins with **poly-LacNAc** moieties has been scrutinized, revealing different preferences for each member. While galectin-1 prefers terminal LacNAc epitopes, galectin-3 only recognize internal positions (similarly to galectin-9N). On the other hand, galectin-7 slight prefers the terminal positions, although with less selectivity

than galectin-1, whereas galectin-8 shows no discrimination for the different LacNAc positions. [80,81]

Galectin-1, -3, and -9 display no preferential **cell-type expression**, since they are heterogeneously expressed throughout the body, including cells of the immune system. Contrariwise, other galectins show some cell-type preferences: galectin-7 is largely expressed by stratified squamous epithelia, galectin-4 and -6 are primarily expressed in the gastrointestinal tract, while galectin-12 levels are high in adipocytes. Galectin-10, is an example of site-specific galectins, since it has been exclusively been found in cells of the immune system. [91]

Concerning their **cellular localization**, galectins can be found in the nucleus, cytosol, outer plasma membrane, and the extracellular matrix. They are synthesized in the cytoplasm and secreted by a non-classical exocytic pathway, that is probably an exosome-mediated secretory route. [92] On the protein-surface, multimeric galectins can crosslink with multivalent glycosylated ligands to build a dynamic scaffold or lattice. The so-called **galectin lattice** is a dynamic, extracellular planar gel-like polymer with increased local concentration of lectins and restricted mobility of glycoproteins anchored to the cell membrane. [93] Generally, tandem-repeat galectins have been shown to generate biological responses more potent than those generated by the prototype galectins. The structural explication likely resides in the presence of the linker, which probably favours crosslinking effects. In fact, it has been demonstrated that an artificial galectin, formed by two CRD of galectin-1 linked with random-coil or rigid α-helical linker, facilitates the formation of larger galectin complexes and increases the affinity for glycans, thus triggering stronger biological effect. Contrarily, a short rigid linker does not produce the same effect. [75]

Although some galectins are involved in intracellular pathways mediated by protein-protein or protein-sugar interactions, it is in the extracellular space where galectins play their major **biological role** interacting with glycans. [94,95] Here, they modulate the *cis*-signalling and the communication of the cell with other cells, with the extracellular matrix or with pathogens. In fact, this class of lectins exhibits pleiotropic

physiological functions that ranges from homeostasis, cell-migration, apoptosis, autophagy to vascular embryogenesis. [96,97]

Consequently, galectins are also involved in a panoply of **pathological events**. [98] To start, they are mediators of acute and chronic inflammation associated to multiple diseases. [99,100] Moreover, they are linked to autoimmune disorders and immune response in general, since they interact with immune cells. [91] Frequently, galectins are associated with the insurgence, development, and maintenance of diverse cancers, being involved in processes such as apoptosis, adhesion, migration, cell transformation, invasion, metastasis, immune escape, and angiogenesis. [101,102]
Host-pathogen interactions are also mediated by galectins, which recognize the sugars epitope exposed on the surface of the exogenous entities and trigger or block the infection. [103]
However, it is difficult to draw general conclusions about the involvement of the entire family in diseases or even in healthy biological events, since each galectin behaves differently. Moreover, the same galectin occasionally appears to act contradictorily depending on the context. Therefore, the discussion on the biological impact will continue mainly focusing on galectin-1, since it has been deeply studied in this Thesis.

The prototype **galectin-1** display a multifaceted and contrasting roles in **cancer**.
In the majority of the cases, this protein positively modulates cell growth, cell adhesion, and cell migration, thereby affecting the process of tumour metastasis. Galectin-1 promotes adhesion through cross-linking with integrin receptors, or cell migration and invasion by recognizing glycans on transmembrane proteins, such as laminin or fibronectin. [104]
Furthermore, it can be defined as a glyco-checkpoint that links tumour immunity and angiogenesis, since it induces apoptosis of activated T cells, thus contributing to tumour-immune escape. [105,106] Changes in its expression are correlated with migratory phenotype and aggressiveness of several tumour types (colon, breast, ovarian, lung, head and neck, among others). [107,108] For instance, high concentrations of circulating galectin-1 is associated with poor clinical outcomes after specific antitumor therapy in head and neck cancer patients. [109]

Controversially, in other contexts, galectin-1 displays anti-cancer activity: the administration of exogenous galectin-1 downregulates proliferation in trophoblastic and neuroblastoma tumour cells. [110,111]

Regarding **inflammation**, galectin-1 controls unresolved inflammation and limits immunopathology. Therefore, its action positive regulates chronic inflammation linked to autoimmune, neurodegenerative, and metabolic diseases. [112]

Oxidized galectin-1, which is no capable to bind sugars, plays an important role in promoting **axonal regeneration**, working as a kind of cytokine, not as a lectin. In fact, the administration of oxidized galectin-1 effectively promoted functional recovery after sciatic nerve injury *in vivo*, demonstrating how the redox state of the microenvironment can have an enormous impact on the galectin function. [113]

In summary, some inflammation-related diseases, neurodegenerative pathologies and muscular dystrophies would benefit of galectin-1 overexpression or selective delivery, whereas its inhibition would be required to fight some types of cancer.

To conclude, the discussed heterogeneous roles of galectins are motivating the rational design of fine-tuned inhibitors with precise selectivity and specificity and therefore, the characterization at the molecular level of the interaction of galectins with their natural binders has paved the structural basis for the design of potent antagonists. Affinities in the nM scale are required for a good drug candidate and to achieve this, both monomeric and multimeric strategies have been exploited, playing in the first case with the addition of hydrophobic substituent to increase the affinity, and in the second case, with scaffolds presenting multiple epitopes to generate avidity enhancement. To date, some molecules are in clinical trials, as the TD139 glycomimetic (Galacto Biotech), which targets galectin-3 to treat Idiopathic Pulmonary Fibrosis. [76,114–116]

1.2.2.2 C-type lectins

C-type lectins are a family of proteins that share structural homology in their carbohydrate recognition domains. The most important element of similarity is the presence of calcium ions (from one up to four atoms) in the binding site, which are fundamental for protein stability as well as for the interaction with carbohydrates. [117] Not all the sites dedicated to Ca^{2+} ions are necessarily occupied and the only one directly involved in the glycan binding is the ion at site 2.

Furthermore, the members of this family share a particular structural organization defined **CTLD domain** (C-type lectin-like domain), which is composed by a central core of five/six/seven β-strands flanked by two α-helices. In the β_2 strand, a conserved WIGL motif is often used as fingerprint for sequence analysis. In addition, the canonical CTLD also presents a long loop region of around 30 amino acids, which take part in sugar recognition and is stabilized by calcium ion sites 1 and 3 (Figure 1.17).

Lastly, another central and highly conserved structural element found among C-type lectins is the presence of strategic Cys residues that establish disulphide bridges, conferring structural stability to the lectin's 3D shape. The disulphide bridges encountered in all the CTLD domains are two: one between α_1-β_5 and other one between β_3-β_4 (Figure 1.17). [118]

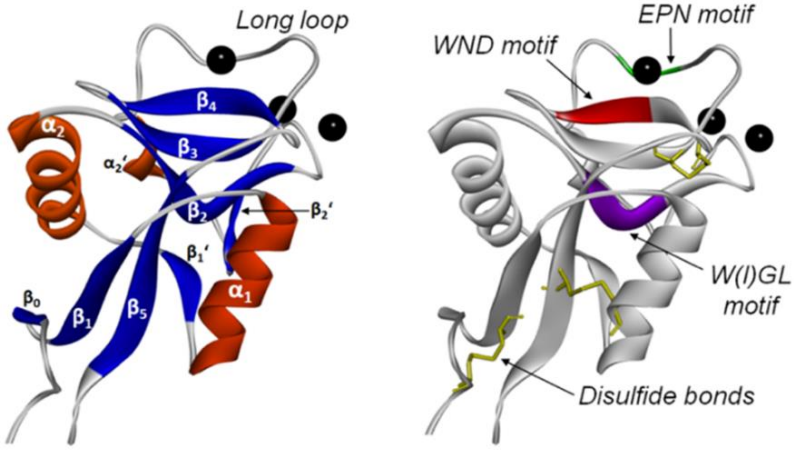


Figure 1.17. Common structural elements (loop, α-helices and β-strands, on the left) and motifs (WDN, WIGL, EPN and disulphide bridges, on the right) conserved in the canonical CTLD domain among C-type lectins. Case of example: DC-SIGN (PDB ID: 1SL5). Figure taken from Valverde et al., 2020. [118]

The sugar **binding site** is shallow and solvent exposed and includes a portion of the loop and other residues from the β_4-strand. Some amino acids participating in the metal

coordination are well conserved (WND motif on the β_4-strand and EPN/QPD motives in the loop). The EPN sequence is generally found in lectins with preference for sugars with equatorial hydroxyl groups at C3 and C4 (Man, Glc).

Some examples are LSECtin, DC-SIGN, and Langerin. The lectins containing the QDP motif (MGL, ASGR) prefer saccharides with axial hydroxyl group at C4, like Gal. [117–120]

The main interaction with saccharides is established through the coordination of two hydroxyl groups of the sugars with the calcium ion and further stabilizations of the union derive from HBs. A common characteristic of this class of lectins is the **promiscuity in ligand specificity** and weak affinity interactions.

C-type lectins can be found both soluble or as membrane-bound receptors. The CTLDs usually oligomerize creating multimeric structures, thus improving the global avidity.

The biological roles of C-type lectins are multiple: they play key roles in homeostasis, pathogen recognition, innate immunity, development and progression of autoimmune diseases, and certain cancer types. [121] The selective targeting of a specific C-type lectin for therapeutic purposes is a difficult task, due to the broad recognition profile and similar preferences among the family members. However, some glycomimetic ligands of DC-SIGN and langerin have been developed and employed to decorate nanoparticles and liposomes. These conjugates have been used to target them with relative success, even achieving certain selectivity. [122–125]

LSECtin, the C-type lectin studied in Chapter 5 of this Thesis, is as a human transmembrane receptor involved in innate immunity.

LSECtin was initially found expressed in liver and lymph node sinusoidal endothelial cells, but it has also been detected in marrow sinusoids, Kuppfer cells, thymic dendritic cells, and in monocyte-derived macrophages. [126–129]

It has been postulated that LSECtin regulates cell adhesion, cell signalling, and glycoprotein clearance through recognition of both self and non-self glycoconjugates (Figure 1.18). [127] Its involvement in immunogenic pathways can be easily deduced from the cellular types where it has been found expressed. Generally, its action is **immunosuppressive**. In fact, after liver injury, LSECtin is capable to negatively modulate T-cell activation and the disease is accelerated in the absence of the lectin.

[130] Moreover, it facilitates the immune escape of cancer and the subsequent tumour progression in melanoma, by inhibiting the T-cell immune response. [131]

LSEctin levels have been found upregulated in tumour-associated macrophages (TAMs) located at the breast cancer tissue. In particular, LSECtin, interacting with receptors on tumour cells, activate pathways that promote cancer stemness. [132] More recently, LSECtin has also emerged as promotor of the adhesion, proliferation, and migration of gastric cancer cells. [133]

The participation of LSECtin in **infection events** has also been studied. The receptor is an attachment factor for the spike protein of SARS coronaviruses, facilitating virus infection. [134] SARS-CoV-2, in fact, binds to LSECtin through specific GlcNAc terminating N-glycans, interfering with the ACE2/spike interaction. [135] LSECtin also interacts with the glycans exposed on the surface of Ebola virus. [136,137] Furthermore, LSECtin enhances the binding of Lassa virus to host cells, promoting the infection lymphocytic choriomeningitis virus (LCMV). [138,139]

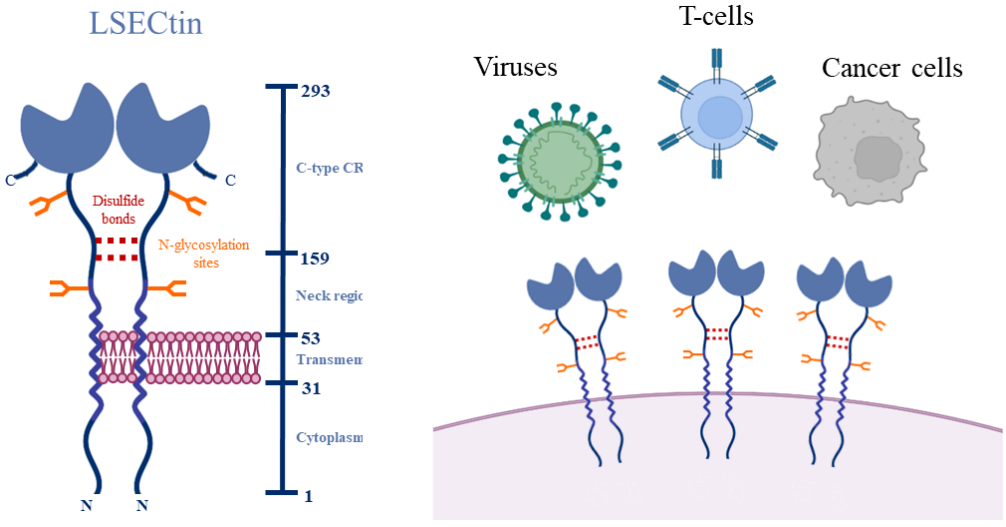


Figure 1.18. On the left: representation of the LSECtin full length receptor. On the right: main biological partner of LSECtin: viruses (exogenous partners, the recognition modulates the infection), T-cells (self-antigens, the interaction controls of the immune response) and cancer cells (self-antigens, the binding event change the tumour spread).

Although it was described for the first time in 2004, its 3D structure is still not available. [126] From sequence alignment, a high degree of sequence identity with DC-SIGN and DC-SIGNR has emerged, and, accordingly, all these lectins are encoded in

the same gene cluster (chromosome 19p13.3). [140] Exploiting this similarity, homology models of the soluble portion (CRD) can be obtained, with conservation of the typical CTLD domain.

From the structural point of view, LSECtin is a type II receptor formed by:

- An **intracellular** N-terminal tail of 31 amino acids,
- A **trans-membrane** portion of 22 amino acids,
- and a long extra-cellular domain (**ECD**), which can be further divided into a **neck** region of 110 residues and a carbohydrate-recognition domain (**CRD**) of 129 residues at C-term (15 KDa) (Figure 1.18).

Regarding molecular recognition, the natural binding partners and the fine details of the interactions of LSECtin with glycans remain poorly understood. Nevertheless, it has been reported that LSECtin recognizes sugars through its extracellular CRD in a Ca^{2+}-dependent manner and is capable to oligomerize on the cell membrane, forming mainly dimers but also tetramers, which are joined together by S-S bonds on the neck region. [126,128,129,136] Two N-glycosylation sites are located in the neck region (positions 77 and 159) and are required for the efficient localization on cell-surface. [126,134]

With respect to its **binding preferences,** LSECtin recognizes **Man** and Glc, accordingly with the presence of the EPN motif. Moreover, it also binds to other monosaccharides, such as Fuc and **GlcNAc**, but does not recognize Gal. [134]
A mass spectrometry analysis on the glycans released from the viral glycoprotein GP1 of the Ebola virus surface showed that N-linked glycans terminating with GlcNAc are recognized by LSECtin, while using microarrays, it was stated that the **GlcNAcβ1-2Man** disaccharide displayed high affinity for the lectin (3.5μM). [136] The GlcNAcβ1-2Man epitope is usually presented on N-glycans. Interestingly the K_D obtained for LSECtin binding is considerably higher than that for DC-SIGN or DC-SIGNR. [132,136]

1.2.2.3 Siglecs

The sialic acid-binding immunoglobulin (Ig)-like lectins (siglecs) family is composed of 15 members in humans. Based on the sequence identity and evolution, two

subgroups can be identified: classic Siglecs and CD33-related Siglecs (Figure 1.19). [141,142] All the siglecs contain an intracellular, a transmembrane, and an **extracellular (ECD) domain**. The ECD is composed by an N-terminal variable-Ig-like (V-Ig) domain, which is in charge of binding sugars, and a flexible number of constant C-type Ig-like (C-Ig) domains. Most siglecs display a tyrosine-based signalling inhibitory motif (ITIM) in the intracellular portion. Instead, Siglecs-14, -15, and -16 contain a positively charged residue within the transmembrane region, which associates with the DAP10/12 (DNAX activation protein-10/12) ITAM-containing adaptor, thereby mediating activating functions. (Figure 1.19). [142]

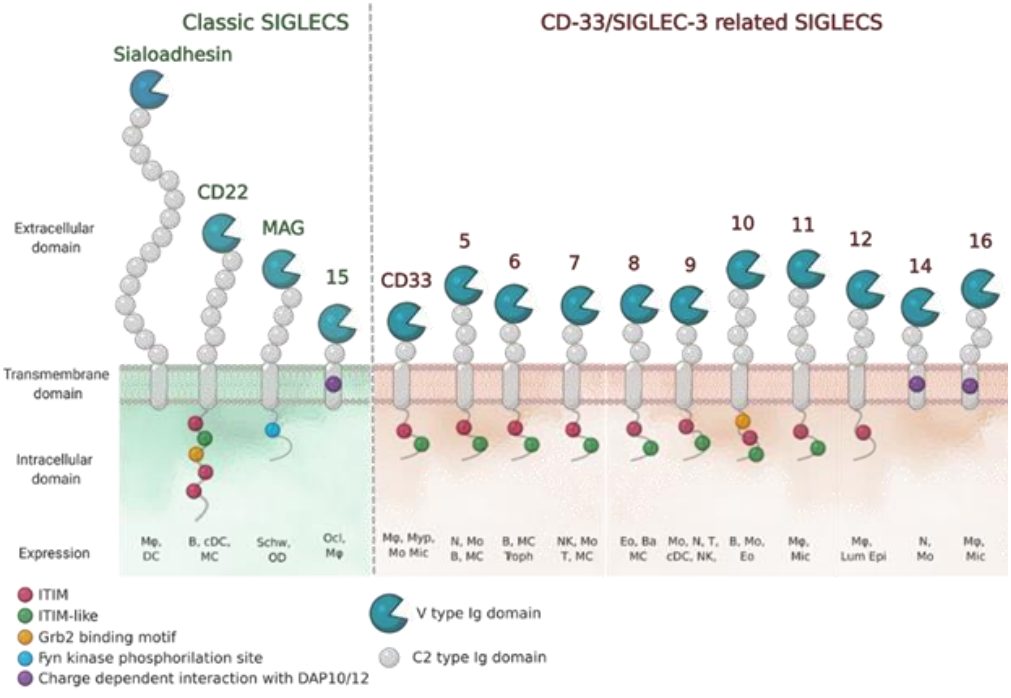


Figure 1.19. Schematic representation of the human siglec family of receptors with their respective domains. The cell types in which each siglec has been found expressed are reported (Mø, macrophages; DC, dendritic cell; B, B cells; MC, mast cells; Schw, Schwann cells; OD, oligodendrocytes; Ocl, osteoclasts; Myp, myeloid progenitor; Mo, monocytes; Mic, microglia; N, neutrophils; Troph, trophoblasts; NK, natural-killer cells; T, T cells; Eo, eosinophils; Ba, basophils; Lum epi, lumen epithelia cells). Figure taken from Lenza et al., 2020. [142]

Indeed, unlike galectins and C-type lectins, which display variegated biological implications, siglecs can be unambiguously defined as receptors of the immune system. In fact, all siglecs, with the exception of siglec-6, are expressed in cells of the innate immune system. Indeed, they play crucial roles in regulating the **immune response** by

mediating both cell–cell interactions and signalling pathways related with immunogenic cascades. [143]

From the structural and molecular recognition perspectives, all the siglecs available structures share the presence of two β-sheets in the V-Ig domain with a variable number of β-strands: A(A´)BED and C(C´)FG(G´), linked through disulphide bridges (Figure 1.20). Siglecs preferentially bind to sialic acid moieties, which can be connected to other sugar units through α2-3, α2-6, and α2-8 linkages. [142,144–147] The sialic acid **binding pocket** is formed by strands F and G and loops C-C´ and C´-D. It always contains a key Arg residue, which establishes a salt bridge with the carboxyl group of the sialic acid. Mutation of this Arg abolishes binding. [141] The affinity for monomeric sialic acid-containing ligands is rather weak. Regarding selectivity, the conformation adopted by C-C´ loop dictates the specificity for the type of sialic acid and linkage that is recognized (Figure 1.20).

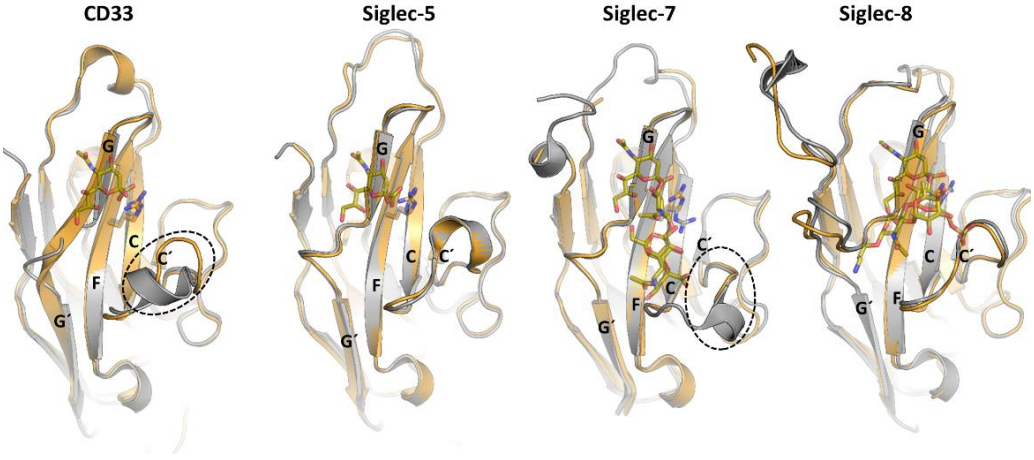


Figure 1.20. Superposition of the *apo* (grey) and bound (orange) structures of the CRD of siglec CD33 (PDB ID: 5IHB and 5J06), Siglec-5 (PDB ID: 2ZG2 and 2ZG3), Siglec-7 (PDB ID: 1O7S and 2HRL) and Siglec-8 (PDB ID: 2N7A and 2N7B).The nomination of the strands and the major conformational changes are indicated. Figure adapted from Lenza et al., 2020. [142]

Significant efforts have been employed to develop chemically modified sialylated ligands able to bind siglecs with elevate affinity and selectivity and then modulate their biological actions. Since detailed structural and molecular recognition information for many members remains elusive, a high-throughput strategy of synthesis and screening

of sialic acid analogue libraries has been proposed and has allowed to identify potent and selective ligands. [148]

Siglec-10 has been chosen as target of the experiments described in Chapter 6. It is a CD33-related siglec with four Ig-C domains. In the intracellular portion, it contains three similar intracellular inhibitory tyrosine-based motifs: a Grb2-binding motif, an ITIM domain, and a membrane-distal ITIM-like domain.

Like in the other family members, siglec-10 mediates the interaction to exogenous sialic acid moieties, which can be masked by *cis*-interactions. [149] Concerning its binding features, a minor preference for **α2-3 linked sialosides** over their α2-6 counterparts has been described, a lower specificity when compared to other siglecs. [150] It has been demonstrated that siglec-10 binds to its major cellular partner, CD24, a highly glycosylated glycoprotein anchored to membrane through GPI, in a sialyl-dependent manner. [151] This lectin has been found expressed on immune cells of the myeloid lineage as well as on B cells, where it regulates innate and adaptive immune responses to tissue injury, sepsis, and viral invasion. [142,152–155] One of its main cellular binders is CD24, Moreover, siglec-10 is also expressed by tumour-associated macrophages, which are in charge detecting tumorigenic cells and orchestrate their clearance through phagocytosis. In this context, it promotes immune evasion of CD24-expressing tumour cells. [152] In fact, CD24, identified as a 'don't-eat-me' signal of cancer cells, interacts with siglec-10, generating an inhibitory signalling cascade that blocks the TAM-mediated clearance, thus evading the immune response. This check point displays evident therapeutic potential, since the blockade of CD24-siglec10 interaction is a target for cancer immunotherapy. [156]

1.3 Techniques to unravel glycan-lectin interactions

The comprehensive study of the fine details that govern the binding events occurring between glycans and lectins is one of the major aim of this Thesis.
Along the years, a collection of complementary techniques has been described that allow characterizing these systems at different levels of resolution. [157]

X-ray crystallography has been for decades the most employed biophysical technique to elucidate the structure of lectin-sugar complexes at atomic resolution. [158] Crystals of protein–carbohydrate complexes can be obtained by co-crystallizing the two molecules together or by soaking the ligand into an existing lectin crystal. Nevertheless, this technique has some general limitations, since the initial trials to find the correct crystallization conditions can require enormous effort in terms of time and, notably, not all proteins are able to crystalize. Moreover, other limitations arise from the intrinsic flexibility of glycans, since they can generate crystal heterogeneity, thus limiting the quality (resolution) of the crystal. The larger the glycan, the harder is the obtainment of structures with good resolution. In addition, it has been demonstrated that deposited structures of sugar-lectin complexes may need further validation and refinement because of erroneous conformations or wrong linkages of the carbohydrate counterpart. [159]

Recent advances in **Cryo electron microscopy** (**cryo-EM**) also provide the possibility to study protein-glycan recognition at atomic level, overcoming the problem of the generation of crystals. The samples can be prepared at low concentrations and in soluble buffers before the vitrification. A great advantage of cryo-EM with respect to X-ray crystallography is that the conformational flexibility is not an obstacle and full glycoproteins can be, in principle, analysed, thus determining the 3D arrangement of their atoms at good resolution.
Cryo-EM can be used not only for defining 3D structures, but also for quality control of the macromolecular composition of the samples. However, one limitation of the technique is the size; in fact, complexes smaller than ≈ 40 kDa are hardly detected. This is not a problem for highly glycosylated entities (such as viruses, glycoproteins or glycolipids) or for the study of multimeric assemblies or lattices (lectins generating

trimers, dimers, tetramers in solution), but prevents the detailed analysis of small lectin-carbohydrate complexes. [160,161]

Surface plasmon resonance (**SPR**) and **biolayer interferometry** (**BLI**) are two biophysical techniques that measure association and dissociation kinetic constants. In both cases, the ligand or the receptor have to be immobilized on a sensor chip, while no additional tags are required for the detection. They can be used to study lectin-sugar complexes, providing thermodynamic and kinetic parameters of the interaction. However, the optimization of the protocol can be challenging. [162,163] Moreover, one of the components is immobilized on a surface, what can affect the proper presentation of the partners, depending on the particular system.

High-throughput **microarrays** also involve the immobilization on solid surfaces. They are a fundamental tool used in Glycoscience to rapidly detect binding hits and guide the rational design of lectin binders. Both glycan and lectin arrays have been developed in a way that hundreds of glycans or lectins can be simultaneously tested in a single experiment. Focusing on glycoarray, glycans are usually covalently linked to a solid surface through reaction with N-hydroxysuccinimide (NHS)-esters on a glass slide. [164–167] Only a few nL of the sample containing the glycans (concentrated 1–100 μm) are deposited by a robotic printer on the glass surface in spots with a diameter of around 100 μm. The support is then: i) incubated to permit the covalent bounding, ii) blocked to avoid nonspecific binding events and iii) incubated with the lectin sample. Several washes follow to remove the unbound lectins. Finally, the binding events are detected through fluorescence measurements. Lectins can be directly labelled with a fluorescent tag or the revelation can be performed by exploiting a fluorescent-tagged anti-lectin antibody. The binding hits are visualized as coloured spots against a dark background and then plotted in terms of relative fluorescence units (RFU).
However, there are some limitations of the technique. At first, high amounts of glycans are densely packed in the spots, a condition that may not occur in nature. Furthermore, the type and length of the linker used, as well as the chemical characteristics of the surface can also affect binding, mainly influencing orientation, packing, and presentation of the glycan. [168,169]

Isothermal titration calorimetry (**ITC**) offers the unique possibility to obtain the thermodynamic parameters that regulate the binding event. No labelling is required for neither the ligand nor the lectin and the growing advances in the detection methods are allowing to perform the experiments with rather low quantity of the sample. Moreover, the data are collected in solution with no need to immobilize one component of the complex. [170,171]

The calorimeter is composed by two identical cells located in an adiabatic jacket. One of them is the reference and contains the solvent in which the samples are dissolved (generally water). The other cell contains the solution with the receptor (the lectin) in its proper buffer (Figure 1.21). The temperature difference (ΔT) between the two cells is always forced to be null. During the experiments, small amounts of the sugar solution (from 1 up to 3 μL) are directly titrated into the cell containing the lectin. The released (exothermic) or absorbed (endotermic) heat generated by the interaction is measured at each point of the titration by the instrument. Specifically, the ΔT between the cells is always zero because a heating power is applied to maintain the absence of temperature differences and this applied power is used estimate the heat released by the interaction.

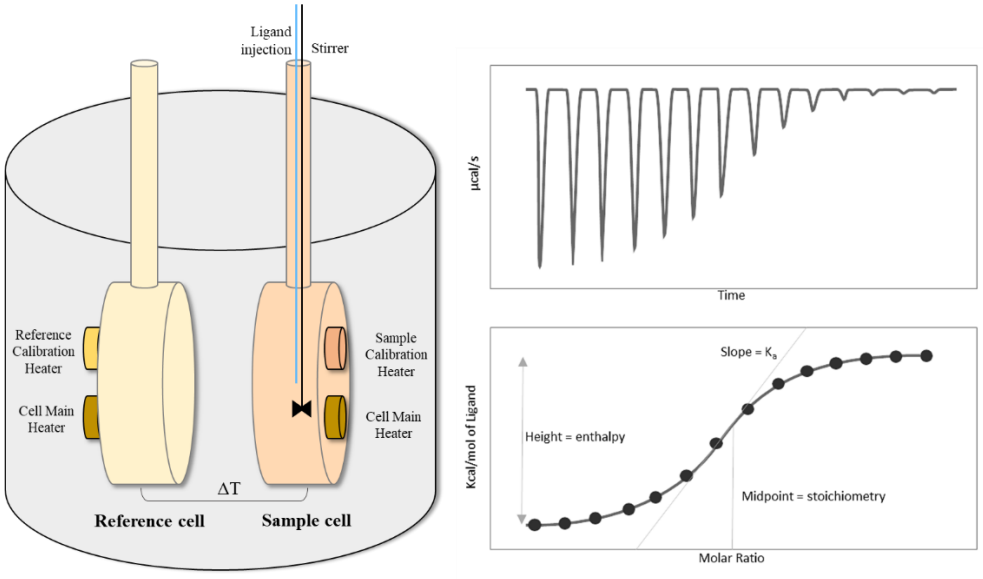


Figure 1.21. A) Representation of the two cells of the calorimeter. B) Thermogram plot of the ITC experiments (above) and sigmoid curve from which the binding parameters are extrapolated.

The heat is then plotted against the ligand concentration and the binding constants (K_D), the reaction stoichiometry (n) and the enthalpy (ΔH) of the recognition process can be accurately determined. In addition, also the entropy (ΔS) can be indirectly estimated. (Figure 1.21). The K_D is obtained from the slope of the curve, which has to be sigmoidal in order to achieve reliable values. Interactions with affinity constants of nM-μM values can be safely studied with ITC and most of sugar-lectin associations falls in this range. The protein concentration can be increased in case of systems with low affinity to optimize the experiment. However, affinity in the mM range usually results incompatible with the acquisition of reliable constants.

Passing to a different topic, the revolution introduced by CRISPR-Cas9 genome engineering also affected the glycoscience field. [172] Employing this technique, it is possible to stably inactivate a specific gene of the cell genome, or even to introduce a new gene coding for a protein of interest. Indeed, the rational engineering of glycosylation (**glyco-engineering**) in mammalian cells can now be performed with a high degree of confidence. [169,173] The editing of glycan biosynthesis through targeted knock outs (KOs) or knock-ins (KIs) of glycosyltransferases opens a new panorama of possibilities to study of lectin-sugar interactions at the cell level, as discussed in Chapter 6. [27,29]
However, challenges still exist in this technique. For instance, partial overlaps in the enzymatic reactions performed by GTs need to be considered.

1.3.1 NMR and molecular recognition

In this Thesis, we have mainly focused on the use of NMR spectroscopy to unravel molecular recognition events of lectin-sugar complexes.
The fundamental theory and the mathematical formalisms behind this technique ahas been extensively discussed in the last decades. [174] Moreover, extensive descriptions of the different multidimensional experiments that allow extracting the key NMR parameters (δ, J, NOEs, T_1, T_2, RDC, PCS, etc) are also widely available. [175–179]
Therefore, I will just focus herein on the applications of NMR to molecular recognition events.

NMR is an excellent technique to explore lectin-glycan interactions,providing information at atomic level. [180] In fact, unlike many other techniques, it offers a significant advantage for studying dynamic events at different time scales. Given the intrinsic flexible nature of saccharides, NMR is the best choice to analyse their structure, conformation, and dynamics in solution. Moreover, this technique can also be applied to inspect the receptor, since conformational motions of the protein can be studied with high precision and detail. Technical advances in terms of sensitivity and resolution are continuously in progress, with current access to magnets beyond 1GHz. Furthermore, the development of new methodologies is also in constant evolution.

There is a vast collection of NMR methodologies that can be employed to study binding events. [178,181] They can be divided into two main groups:

- ligand-based methods, in which changes in the NMR signals of the ligand (the sugar) are observed.
- receptor-based methods, in which the information arise from perturbations in the NMR signals of the receptor (the lectin).

Both groups rely on the fact that protein-ligand binding is a chemical exchange event between the free and the bound states, governed by k_{on} and k_{off} constants. Through NMR experiments, this process can be monitored considering different parameters (chemical shifts, relaxation rates, line width, signal intensity, etc.), depending on the specific k_{on} and k_{off} values.

1.3.1.1 Ligand-based NMR methods

NMR methods from the ligand's point of view monitor and track differences in the ligand NMR signals when passing from free in solution to bound to the receptor. [182,183] In this case, the protein counterpart is not a limitation, since just tiny amounts of unlabelled protein can be used without particular size limits.

These experiments take advantage of the change in the rotational motion correlation time of the small ligand upon binding to the receptor. In fact, **small molecules** have fast motions (molecular tumbling), in the ps time scale, which are associated with slow NMR relaxation time (T_2 relaxation) in the s time scale and generate narrow signals.

Contrarily, **large molecules** display slower motions (ns time scale), fast T_2 relaxation time (ms) and give rise to rather broad signals.

The key point is that when the ligand binds to the receptor, its motional properties change. In fact, when both species are in solution and interact, the free ligand is in equilibrium with its **complexed** form to the receptor. Thus, the ligand acquires new properties, which are similar those of the receptor and characteristic of a macromolecule: slow molecular tumbling (ns time scale) and, therefore, fast T_2 relaxation time (ms). This perturbation is reflected on the NMR experiments: when there is interaction, the signals arising from the ligand became much broader and can even disappear.

Another ley NMR parameter that can be used to distinguish between free and bound ligands is the Nuclear Overhauser Effect (NOE), which is inversely proportional to the sixth power of the distance between two nuclei and largely affected by the molecular motion time scale. For a small ligand with fast motions, the NOEs are positive (cross peaks are of opposite sign to the diagonal peaks in NOESY experiments), while they are negative for large molecules, with slow motions. Again, a bound ligand will display the characteristics of the large molecule and display negative NOE signals (Figure 1.22). [184]

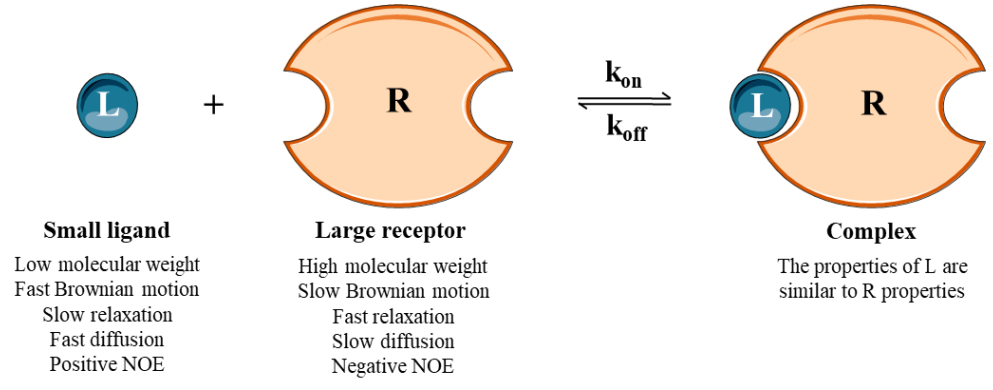


Figure 1.22. The different motional properties of ligand, protein and complex in solution that are exploited for ligand-based methodologies.

The ligand-based NMR experiments are based on:

- exploiting the dramatic change in motional properties of the ligand in the free versus the bound state (trNOESY, trROESY, for instance).
- exploiting the transfer of magnetization (STD-NMR, mainly).

The *geometry of the bound ligand* can be elucidated through **trNOESY experiments**. As mentioned above, the intensity of the NOE cross-peaks in a two-dimensional NOESY spectra depends on the rotational motion correlation time of the molecule in solution, which is related to the molecular size, but also on the inverse sixth power of the distance between two nuclei. Therefore, the NOEs contain key conformational information. Typically, small ligands display NOE signals positive or close to zero in NOESY experiments (short rotational motion correlation times).

It has to be pointed out that the sign and intensity of the NOE cross peaks is also dependent on the strength of the magnetic field of the magnet (Figure 1.23).

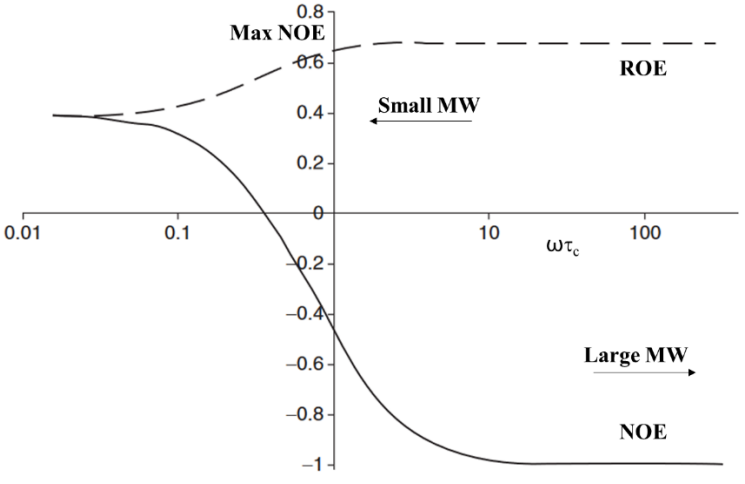


Figure 1.23. Plot of NOE / ROE intensity as a function of the correlation time ($\omega\tau_c$). For small molecules, the NOE builds up very slowly and is positive, while for big molecules the NOE builds up rapidly and is negative. For molecules of medium size the NOE crosses over from positive to negative, and thus NOEs are very small. ROEs are non-zero for all correlation times.

If the ligand interacts with the receptor, the NOE changes from positive to negative, what can be measured in the so-called transferred NOESY (trNOESY) experiment. Fittingly, since the slow motions strongly dominate relaxation features, the NOE (formally, the cross-relaxation rate) for the bound form is much larger than that of the free form, and the observed NOE will be dominated by the bound form. Thus, the observed trNOEs provides information on the bound ligand conformation (Figure 1.24). [184]

The set-up of the methodology is composed by two sequential steps:

- The acquisition of a NOESY spectrum for the ligand in absence of the protein. The mixing time to achieve the maximum intensity of the cross-peaks is between 500 ms and 1 s for a small molecule (di to tetrasaccharides). For longer molecules, the mixing time is decreased at a range of 100-300 ms.
- The acquisition of a trNOESY spectrum with ligand and protein in solution. Medium-low protein:ligand ratios are typically employed (from 1:5 to 1:40, optimised depending on the affinity and on the kinetic parameters). The mixing time is usually set below 200 ms.

Nevertheless, this technique also has some weak points and limitations.

For example, there is a narrow accessible kinetic window: if the k_{off} is too slow, the NOE information arising from the ligand at the protein binding site is lost due to the fast transverse relaxation that take place in the bound state. Furthermore, the existence of spin diffusion effects, typical for large molecules, can generate negative cross-peaks between protons which are relatively far apart, thus complicating the analysis of the geometry in the bound state. This is the reason for which short mixing times are employed in trNOESY experiments.

It has also to be considered that NOE signals for vary large N-glycans, with rotational motion correlation times of several hundreds of ps, are already negative, a characteristic that may hamper the quantitative analysis of the bound conformation.

The combined employment of Rotating-frame Overhauser SpectroscoPY NMR (**trROESY-NMR**) **experiments** can help to discriminate these issues, since in these experiments, positive cross peaks arise for the direct NOEs between a given proton pair, while cross peaks mediated by a third spin are negative (opposite sign to the ROE-derived peaks) (Figure 1.23 and Figure 1.24). [182,185]

Fittingly, if the chemical exchange event is slow in the chemical shift time scale, cross-peaks arising from the chemical exchange (EXSY) between the bound and free form of a proton, can also be detected. These peaks always (either in NOESY or in ROESY) appear with negative sign (Figure 1.24). In this way, the *chemical shift perturbation suffered by a specific proton of the ligand in the bound state* provides additional information about the interacting ligand epitope. For instance, upfield shifts are indication of closeness to an aromatic ring of the protein.

Moreover, the quantitative analysis of the intensities of the exchange cross peaks may allow the calculation of the kinetic exchange rates k_{on} and k_{off} with a detailed EXSY analysis. [186,187]

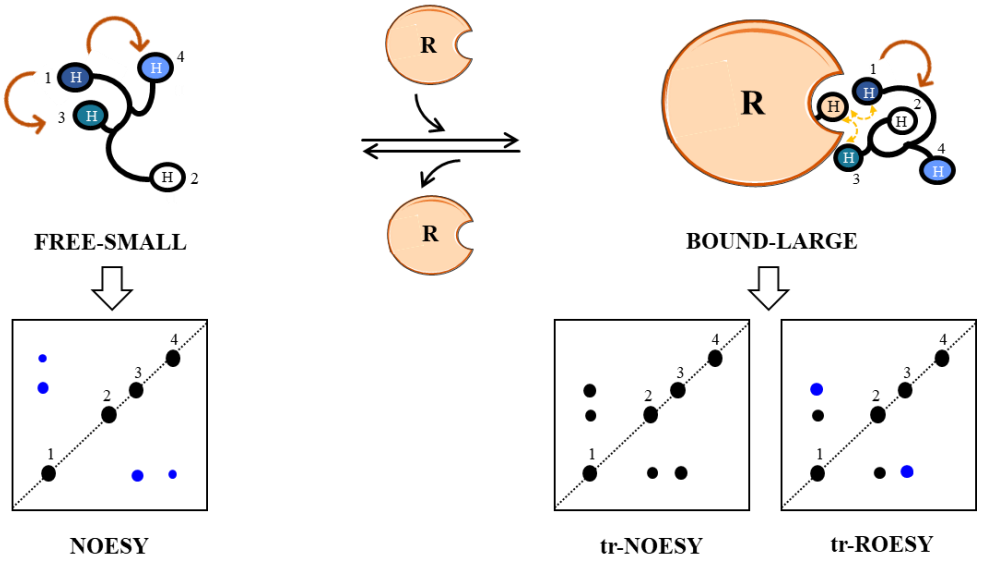


Figure 1.24. Visual representation of a NOESY, trNOESY and trROESY experiment. For a small molecule in solution (NOESY) the cross-peaks are positive and display opposite sign to the diagonal peaks; in the presence of the receptor, the cross-peaks of the bound ligand become negative; in order to differentiate the peaks generated by spin diffusion artefacts (or chemical exchange cross peak), in trROESY they show opposite sign to the diagonal.

The **Saturation Transfer Difference (STD) NMR** experiment is the most versatile and widely used NMR tool to study ligand-receptor interactions. [188–190] STD-NMR not only provides information on the presence or absence of the recognition event, but also permits the extrapolation of a key clue: *the ligand binding epitope*, a map of the protons of the ligand that are in close contact to the protein surface. In particular, two different spectra (*on-resonance* and *off-resonance*) are recorded for the same sample applying selective irradiations:

- In the *on-resonance* spectrum, a train of low power radiofrequency pulses is applied for a few seconds (typically 2, from 1 to 3 s) at a frequency chosen so that no ligand protons are present at least within 1-2 ppm. This saturation is efficiently propagated across the entire receptor through spin-diffusion, and transferred to the ligand, provided that it is contact with the protein receptor. The saturation affects the intensities of the signals of the ligand, depending on the distance to the receptor. Then the ligand, which is transiently bound to the

receptor, dissociates to the solution. The saturated state of the ligand protons persists and the intensity changes can be detected. Since the experiment is carried out using a large excess of the ligand, given the free-bound exchange process, the population of saturated ligands back to solution systematically increase and accumulate (provided that the relaxation is slow relative to k_{off}), due to the continuous turnover of ligands interacting with the protein. As consequence, a spectrum is generated where the NMR signals of the ligand protons display diverse intensities due to their partial saturation, which depend on the relative proximity to the protein protons. A large ligand excess (more than 50 molar equivalents with respect to the receptor) allows maximizing the population of saturated ligand in solution, thus increasing the sensitivity of the experiment. Obviously, the binding event has to be strong enough to permit an efficient saturation, but also weak enough to ensure a fast dissociation in the relaxation time scale and the accumulation of a pool of saturated ligands in solution. The affinity usually ranges from 10^{-8} M to 10^{-3} M.

- In the *off-resonance* spectrum, the reference, the selective irradiation is performed at a frequency where neither ligand nor protein signals appear (usually 100 ppm).

Finally, the *on-resonance* spectrum is subtracted to the *off-resonance* one, generating a difference spectrum in which only the ligand protons that have modified their intensities in the on-resonance experiment (those in close contact to the protein) display a signal. The identification of the STD signals and their relative quantification build the binding epitope of the ligand. (Figure 1.25). In fact, the most intense STD signal is assumed to be 100% and other STD signals are normalized respectively. [180] Additional information on the types of amino acid residues surrounding the ligand in the binding pocket can be obtained comparing STD-NMR experiments acquired with either aliphatic or aromatic irradiation of the protein. [191]

If the overlap of the ligand signals affects the final analysis, 2D STD-NMR experiments are an option to enhance in the spectral dispersion. 2D STD-TOCSY and 2D-STD-HSQC NMR experiments have been employed in the glycan field. However ^{13}C labelled sugars have to be available in this last case. [81,192]

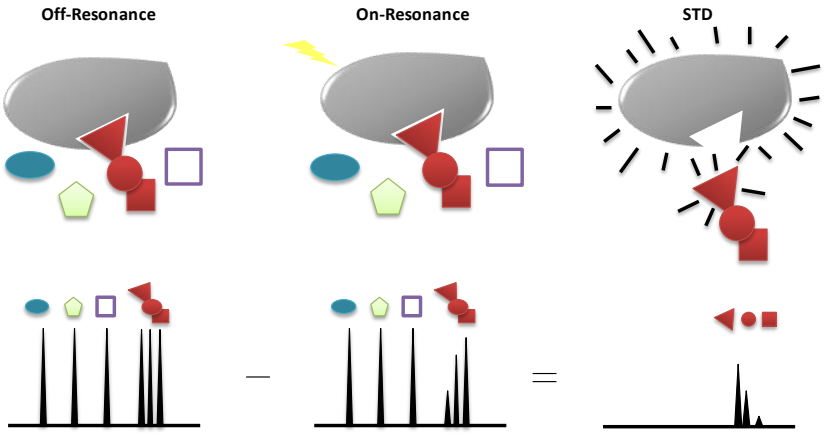


Figure 1.25. Visual representation of the two spectra acquired during a STD experiment (*off-resonance* and *on-resonance*), together with their difference (STD spectrum).

1.3.1.2 Receptor-based NMR methods

The NMR receptor-based methods focus on the changes in the NMR parameters of the receptor in the absence and presence of the ligand of interest.

Simple experiments can be performed to monitor **^{1}H signals of the receptor** in mono- or two-dimensional spectra using unlabelled receptors. However, a tremendous overlap in the proton dimension usually impedes this analysis.

The labelling of the protein with one or more active heteronucleus (^{15}N, ^{13}C, etc) to perform 2D heteronuclear experiments is the preferential choice to carry out these experiments. Many protocols are now available for the expression and purification of proteins in labelled media, although the methodology may be expensive. Moreover, when changing from well-established expression systems (*E. coli*) to complex host cells (mammalian cells, for instance), the optimization of the protocols can be rather time demanding.

The most common strategy, extensively used in this Thesis, is to use ^{15}N labelled protein to perform ^{1}H-^{15}N 2D-experiments (^{1}H-^{15}N heteronuclear single-quantum coherence, HSQC or ^{1}H-^{15}N TROSY, for very large complexes).

^{1}H-^{15}N HSQC and **TROSY experiments** display cross peaks which correspond to all the NH amide groups of the backbone and lateral chains of a protein. The protein sequence, its secondary, tertiary and quaternary organization provide a unique chemical environment for each amino acid. Since the chemical environment finely

modulates the NMR chemical shift, the HSQC spectrum of each protein is distinctive and can be defined as *the protein fingerprint*. [193]

If the protein interacts with a ligand in solution, the chemical environment of the residues directly or indirectly involved in this interaction is strongly perturbed. Based on this, the cross-peaks generated by these amino acids experience a **NMR chemical shift perturbation (CSP)**, which can be monitored.

The general strategy involves the acquisition of a first reference $^{1}H,^{15}N$-HSQC spectrum with the protein in its *apo* form. Subsequently, the ligand is titrated at increasing concentrations into the protein sample, and at each point of the titration, a new spectrum is acquired. The titration proceeds until the receptor reaches the saturation. The CSP data are then analysed and the final results allow to discern the perturbation of each amino acid upon recognition. [194] In homology with the ligand binding epitope obtained through STD NMR, here it is possible to obtain the protein binding epitope of the interaction (Figure 1.26). Interestingly, not only residues in close contact with the ligand are perturbed (located in the binding pocket), but also residues far away from the binding site that may experience conformational motions can be detected and identified in this analysis.

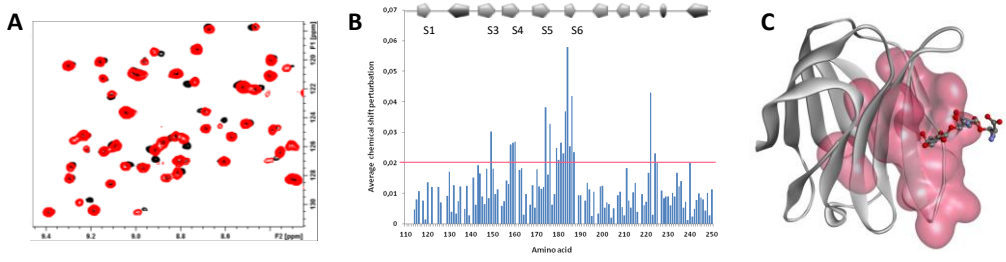


Figure 1.26. A) Superposition of ^{1}H-^{15}N HSQC experiments of *apo* protein (black) and protein at the presence of the ligand of interest (red). B) CSP plot of the amino acid sequence of the protein. C) Epitope map of the protein residues involved in the interaction with the ligand.

Additionally, the kinetics of the binding process can be scrutinized. Indeed, if the chemical exchange rate is fast in the NMR time scale, the cross-peaks gradually shift, while when the exchange rate is slow in the chemical shift time scale, two peaks are simultaneously displayed (free and bound state). Intermediate chemical exchange rates can also occur, generating heterogeneous patterns, where the peaks shift, broaden, and may even disappear (Figure 1.27).

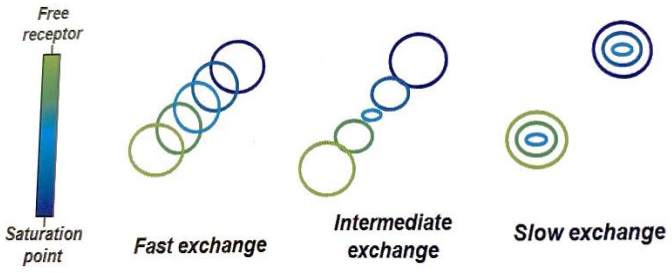


Figure 1.27. representation of the pattern of ^{1}H-^{15}N HSQC-based titration depending on the chemical exchange rate in the NMR time scale of the system in study (fast or slow).

The intensity of the cross-peak is also informative: when the supramolecular complexes formed upon binding are very large, they can barely be detected by NMR. It is worthy to mention that this general class of experiments is accessible only to proteins with a limited size, since entities larger than 100 kDa become blind to NMR. Nevertheless, TROSY experiments on ^{13}C-labelled methyl peaks may become the alternative for such huge systems. [195]

Many other types of experiments from the viewpoint of the protein can be chosen to unravel other key information. For instance, phase-modulated **CLEAN chemical exchange NMR (CLEANEX-PM)** allows the detection of NH residues presenting very fast exchange rates with water in a time range of 5-500 ms. [196,197]

^{15}N CPMG relaxation dispersion NMR experiments are another elegant tool to detect amino acids that show dynamic motions in a particular time scale: μs-ms, which is often correlated with allosterism. [198]

1.4 References

1. Varki, A.; Cummings, R.D.; Esko, J.D.; Stanley, P.; Hart, G.W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N.H.; Prestegard, J.H.; et al. Essentials of Glycobiology. *Essentials Glycobiol.* **2022**, doi:10.1101/9781621824213.
2. Watt, G.D. A New Future for Carbohydrate Fuel Cells. *Renew. Energy* **2014**, *72*, 99–104, doi:10.1016/J.RENENE.2014.06.025.
3. Flynn, R.A.; Pedram, K.; Malaker, S.A.; Batista, P.J.; Smith, B.A.H.; Johnson, A.G.; George, B.M.; Majzoub, K.; Villalta, P.W.; Carette, J.E.; et al. Small RNAs Are Modified with N-Glycans and Displayed on the Surface of Living Cells. *Cell* **2021**, *184*, 3109-3124.e22, doi:10.1016/J.CELL.2021.04.023.
4. Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27*, 3–49, doi:10.1093/GLYCOB/CWW086.
5. Angyal, S.J. The Composition and Conformation of Sugars in Solution. *Angew. Chemie Int. Ed. English* **1969**, *8*, 157–166, doi:10.1002/ANIE.196901571.
6. Kendrew, J.C.; Moelwyn-Hughes, E.A. The Kinetics of Mutarotation in Solution. *Proc. R. Soc. Lond. A. Math. Phys. Sci.* **1940**, *176*, 352–367, doi:10.1098/RSPA.1940.0094.
7. Pigman, W.; Isbell, H.S. Mutarotation of Sugars in Solution: Part I: History, Basic Kinetics, and Composition of Sugar Solutions. *Adv. Carbohydr. Chem.* **1968**, *23*, 11–57, doi:10.1016/S0096-5332(08)60167-8.
8. Lemieux, R. U.; Chu, N.J. Abstract of Paper. *Am. Chem. Soc.* **1958**, *133*.
9. Lemieux, R.U.; Koto, S. The Conformational Properties of Glycosidic Linkages. *Tetrahedron* **1974**, *30*, 1933–1944, doi:10.1016/S0040-4020(01)97324-7.
10. Laine, R.A. A Calculation of All Possible Oligosaccharide Isomers Both Branched and Linear Yields 1.05 x 10(12) Structures for a Reducing Hexasaccharide: The Isomer Barrier to Development of Single-Method Saccharide Sequencing or Synthesis Systems. *Glycobiology* **1994**, *4*, 759–767, doi:10.1093/GLYCOB/4.6.759.
11. An, H.J.; Froehlich, J.W.; Lebrilla, C.B. Determination of Glycosylation Sites and Site-Specific Heterogeneity in Glycoproteins. *Curr. Opin. Chem. Biol.* **2009**, *13*, 421, doi:10.1016/J.CBPA.2009.07.022.
12. Paulson, J.C. Glycoproteins: What Are the Sugar Chains For? *Trends Biochem. Sci.* **1989**, *14*, 272–276, doi:10.1016/0968-0004(89)90062-5.
13. Bagdonaite, I.; Malaker, S.A.; Polasky, D.A.; Riley, N.M.; Schjoldager, K.; Vakhrushev, S.Y.; Halim, A.; Aoki-Kinoshita, K.F.; Nesvizhskii, A.I.; Bertozzi, C.R.; et al. Glycoproteomics. *Nat. Rev. Methods Prim. 2022 21* **2022**, *2*, 1–29, doi:10.1038/s43586-022-00128-4.
14. Schjoldager, K.T.; Narimatsu, Y.; Joshi, H.J.; Clausen, H. Global View of Human Protein Glycosylation Pathways and Functions. *Nat. Rev. Mol. Cell Biol. 2020 2112* **2020**, *21*, 729–749, doi:10.1038/s41580-020-00294-x.
15. Kinoshita, T.; Fujita, M. Biosynthesis of GPI-Anchored Proteins: Special Emphasis on GPI Lipid Remodeling. *J. Lipid Res.* **2015**, *57*, 6–24, doi:10.1194/JLR.R063313.
16. Cao, L.; Diedrich, J.K.; Ma, Y.; Wang, N.; Pauthner, M.; Park, S.K.R.; Delahunty, C.M.; McLellan, J.S.; Burton, D.R.; Yates, J.R.; et al. Global Site-Specific Analysis of Glycoprotein N-Glycan Processing. *Nat. Protoc. 2018 136* **2018**, *13*, 1196–1212, doi:10.1038/nprot.2018.024.
17. Bieberich, E. Synthesis, Processing, and Function of N-Glycans in N-Glycoproteins. *Adv. Neurobiol.* **2014**, *9*, 47, doi:10.1007/978-1-4939-1154-7_3.
18. Schwarz, F.; Aebi, M. Mechanisms and Principles of N-Linked Protein Glycosylation. *Curr. Opin. Struct. Biol.* **2011**, *21*, 576–582, doi:10.1016/J.SBI.2011.08.005.
19. Roth, J.; Zuber, C.; Park, S.; Jang, I.; Lee, Y.; Kysela, K.G.; Le Fourn, V.; Santimaria, R.; Guhl, B.; Cho, J.W. Protein N-Glycosylation, Protein Folding, and Protein Quality Control. *Mol. Cells 2010 306* **2010**, *30*, 497–506, doi:10.1007/S10059-010-0159-Z.

20. Fisher, P.; Thomas-Oates, J.; Wood, A.J.; Ungar, D. The N-Glycosylation Processing Potential of the Mammalian Golgi Apparatus. *Front. Cell Dev. Biol.* **2019**, *7*, 157, doi:10.3389/FCELL.2019.00157/BIBTEX.
21. Jensen, P.H.; Kolarich, D.; Packer, N.H. Mucin-Type O-Glycosylation--Putting the Pieces Together. *FEBS J.* **2010**, *277*, 81–94, doi:10.1111/J.1742-4658.2009.07429.X.
22. King, S.L.; Joshi, H.J.; Schjoldager, K.T.; Halim, A.; Madsen, T.D.; Dziegiel, M.H.; Woetmann, A.; Vakhrushev, S.Y.; Wandall, H.H. Characterizing the O-Glycosylation Landscape of Human Plasma, Platelets, and Endothelial Cells. *Blood Adv.* **2017**, *1*, 429–442, doi:10.1182/BLOODADVANCES.2016002121.
23. Iozzo, R. V.; Schaefer, L. Proteoglycan Form and Function: A Comprehensive Nomenclature of Proteoglycans. *Matrix Biol.* **2015**, *42*, 11–55, doi:10.1016/J.MATBIO.2015.02.003.
24. D'Angelo, G.; Capasso, S.; Sticco, L.; Russo, D. Glycosphingolipids: Synthesis and Functions. *FEBS J.* **2013**, *280*, 6338–6353, doi:10.1111/FEBS.12559.
25. MacCioni, H.J.F.; Quiroga, R.; Ferrari, M.L. Cellular and Molecular Biology of Glycosphingolipid Glycosylation. *J. Neurochem.* **2011**, *117*, 589–602, doi:10.1111/J.1471-4159.2011.07232.X.
26. Mohorko, E.; Glockshuber, R.; Aebi, M. Oligosaccharyltransferase: The Central Enzyme of N-Linked Protein Glycosylation. *J. Inherit. Metab. Dis.* **2011**, *34*, 869–878, doi:10.1007/S10545-011-9337-1/FIGURES/2.
27. Narimatsu, Y.; Joshi, H.J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y.H.; Schjoldager, K.T.; Steentoft, C.; et al. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cell* **2019**, *75*, 394-407.e5, doi:10.1016/J.MOLCEL.2019.05.017.
28. Hansen, L.; Lind-Thomsen, A.; Joshi, H.J.; Pedersen, N.B.; Have, C.T.; Kong, Y.; Wang, S.; Sparso, T.; Grarup, N.; Bech Vester-Christensen, M.; et al. A Glycogene Mutation Map for Discovery of Diseases of Glycosylation. *Glycobiology* **2014**, *25*, 211–224, doi:10.1093/glycob/cwu104.
29. Narimatsu, Y.; Büll, C.; Chen, Y.H.; Wandall, H.H.; Yang, Z.; Clausen, H. Genetic Glycoengineering in Mammalian Cells. *J. Biol. Chem.* **2021**, *296*, 100448, doi:10.1016/J.JBC.2021.100448.
30. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol. 2019 156* **2019**, *15*, 346–366, doi:10.1038/s41581-019-0129-4.
31. Magalhães, A.; Duarte, H.O.; Reis, C.A. The Role of O-Glycosylation in Human Disease. *Mol. Aspects Med.* **2021**, *79*, 100964, doi:10.1016/J.MAM.2021.100964.
32. Burchell, J.M.; Beatson, R.; Graham, R.; Taylor-Papadimitriou, J.; Tajadura-Ortega, V. O-Linked Mucin-Type Glycosylation in Breast Cancer. *Biochem. Soc. Trans.* **2018**, *46*, 779–788, doi:10.1042/BST20170483.
33. Pietrobono, S.; Stecca, B. Aberrant Sialylation in Cancer: Biomarker and Potential Target for Therapeutic Intervention? *Cancers (Basel).* **2021**, *13*, doi:10.3390/CANCERS13092014.
34. Leroy, J.G. Congenital Disorders of N-Glycosylation Including Diseases Associated With O- as Well as N-Glycosylation Defects. *Pediatr. Res. 2006 606* **2006**, *60*, 643–656, doi:10.1203/01.pdr.0000246802.57692.ea.
35. Péanne, R.; de Lonlay, P.; Foulquier, F.; Kornak, U.; Lefeber, D.J.; Morava, E.; Pérez, B.; Seta, N.; Thiel, C.; Van Schaftingen, E.; et al. Congenital Disorders of Glycosylation (CDG): Quo Vadis? *Eur. J. Med. Genet.* **2018**, *61*, 643–663, doi:10.1016/J.EJMG.2017.10.012.
36. Varki, A. Evolutionary Forces Shaping the Golgi Glycosylation Machinery: Why Cell Surface Glycans Are Universal to Living Cells. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*, 1–14, doi:10.1101/CSHPERSPECT.A005462.
37. Holt, G.D.; Hart, G.W. The Subcellular Distribution of Terminal N-Acetylglucosamine Moieties. Localization of a Novel Protein-Saccharide Linkage, O-Linked GlcNAc. *J.*

Biol. Chem. **1986**, *261*, 8049–8057, doi:10.1016/S0021-9258(19)57510-X.

38. Holt, G.D.; Haltiwanger, R.S.; Torres, C.R.; Hart, G.W. Erythrocytes Contain Cytoplasmic Glycoproteins. O-Linked GlcNAc on Band 4.1. *J. Biol. Chem.* **1987**, *262*, 14847–14850, doi:10.1016/S0021-9258(18)48100-8.
39. D'Onofrio, M.; Starr, C.M.; Park, M.K.; Holt, G.D.; Haltiwanger, R.S.; Hart, G.W.; Hanover, J.A. Partial CDNA Sequence Encoding a Nuclear Pore Protein Modified by O-Linked N-Acetylglucosamine. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 9595–9599, doi:10.1073/PNAS.85.24.9595.
40. Hart, G.W.; Haltiwanger, R.S.; Holt, G.D.; Kelly, W.G. Glycosylation in the Nucleus and Cytoplasm. *Annu. Rev. Biochem.* **1989**, *58*, 841–874, doi:10.1146/ANNUREV.BI.58.070189.004205.
41. Chou, T.Y.; Hart, G.W.; Dang, C. V. C-Myc Is Glycosylated at Threonine 58, a Known Phosphorylation Site and a Mutational Hot Spot in Lymphomas. *J. Biol. Chem.* **1995**, *270*, 18961–18965, doi:10.1074/JBC.270.32.18961.
42. Luft, J.H. Fine Structures of Capillary and Endocapillary Layer as Revealed by Ruthenium Red. *Fed. Proc.* **1966**, *25*, 1773–1783.
43. Reitsma, S.; Slaaf, D.W.; Vink, H.; Van Zandvoort, M.A.M.J.; Oude Egbrink, M.G.A. The Endothelial Glycocalyx: Composition, Functions, and Visualization. *Pflugers Arch.* **2007**, *454*, 345–359, doi:10.1007/S00424-007-0212-8.
44. Van den Berg, B.M.; Vink, H.; Spaan, J.A.E. The Endothelial Glycocalyx Protects against Myocardial Edema. *Circ. Res.* **2003**, *92*, 592–594, doi:10.1161/01.RES.0000065917.53950.75.
45. Van den Berg, B.M.; Vink, H.; Spaan, J.A.E. The Endothelial Glycocalyx Protects Against Myocardial Edema. *Circ. Res.* **2003**, *92*, 592–594, doi:10.1161/01.RES.0000065917.53950.75.
46. Möckl, L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front. Cell Dev. Biol.* **2020**, *8*, 253, doi:10.3389/FCELL.2020.00253/BIBTEX.
47. Lauehlin, S.T.; Bertozzi, C.R. Metabolic Labeling of Glycans with Azido Sugars and Subsequent Glycan-Profiling and Visualization via Staudinger Ligation. *Nat. Protoc. 2007 211* **2007**, *2*, 2930–2944, doi:10.1038/nprot.2007.422.
48. Möckl, L.; Pedram, K.; Roy, A.R.; Krishnan, V.; Gustavsson, A.K.; Dorigo, O.; Bertozzi, C.R.; Moerner, W.E. Quantitative Super-Resolution Microscopy of the Mammalian Glycocalyx. *Dev. Cell* **2019**, *50*, 57-72.e6, doi:10.1016/J.DEVCEL.2019.04.035.
49. Purcell, S.C.; Godula, K. Synthetic Glycoscapes: Addressing the Structural and Functional Complexity of the Glycocalyx. *J. R. Soc. Interface Focus* **2019**, *9*, doi:10.1098/RSFS.2018.0080.
50. Tarbell, J.M.; Cancel, L.M. The Glycocalyx and Its Significance in Human Medicine. *J. Intern. Med.* **2016**, *280*, 97–113, doi:10.1111/JOIM.12465.
51. Ushiyama, A.; Kataoka, H.; Iijima, T. Glycocalyx and Its Involvement in Clinical Pathophysiologies. *J. Intensive Care* **2016**, *4*, 1–11, doi:10.1186/S40560-016-0182-Z/TABLES/3.
52. Zhang, L. Glycosaminoglycan (GAG) Biosynthesis and GAG-Binding Proteins. *Prog. Mol. Biol. Transl. Sci.* **2010**, *93*, 1–17, doi:10.1016/S1877-1173(10)93001-9.
53. Hevey, R. Bioisosteres of Carbohydrate Functional Groups in Glycomimetic Design. *Biomimetics* **2019**, *4*, doi:10.3390/BIOMIMETICS4030053.
54. Toone, E.J. Structure and Energetics of Protein-Carbohydrate Complexes. *Curr. Opin. Struct. Biol.* **1994**, *4*, 719–728, doi:10.1016/S0959-440X(94)90170-8.
55. Vennelakanti, V.; Qi, H.W.; Mehmood, R.; Kulik, H.J. When Are Two Hydrogen Bonds Better than One? Accurate First-Principles Models Explain the Balance of Hydrogen Bond Donors and Acceptors Found in Proteins. *Chem. Sci.* **2021**, *12*, 1147–1162, doi:10.1039/D0SC05084A.

56. Asensio, J.L.; Ardá, A.; Cañada, F.J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946–954, doi:10.1021/AR300024D/ASSET/IMAGES/MEDIUM/AR-2012-00024D_0006.GIF.
57. Hudson, K.L.; Bartlett, G.J.; Diehl, R.C.; Agirre, J.; Gallagher, T.; Kiessling, L.L.; Woolfson, D.N. Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **2015**, *137*, 15152–15160, doi:10.1021/JACS.5B08424/ASSET/IMAGES/LARGE/JA-2015-08424N_0006.JPEG.
58. Hsu, C.H.; Park, S.; Mortenson, D.E.; Foley, B.L.; Wang, X.; Woods, R.J.; Case, D.A.; Powers, E.T.; Wong, C.H.; Dyson, H.J.; et al. The Dependence of Carbohydrate-Aromatic Interaction Strengths on the Structure of the Carbohydrate. *J. Am. Chem. Soc.* **2016**, *138*, 7636–7648, doi:10.1021/JACS.6B02879/SUPPL_FILE/JA6B02879_SI_002.PDF.
59. Weis, W.I.; Drickamer, K. Structural Basis of Lectin-Carbohydrate Recognition. *Annu. Rev. Biochem.* **1996**, *65*, 441–473, doi:10.1146/ANNUREV.BI.65.070196.002301.
60. Claveria-Gimeno, R.; Vega, S.; Abian, O.; Velazquez-Campoy, A. A Look at Ligand Binding Thermodynamics in Drug Discovery. *Expert Opin. Drug Discov.* **2009**, *12*, 363–377, doi:10.1080/17460441.2017.1297418.
61. Fox, J.M.; Zhao, M.; Fink, M.J.; Kang, K.; Whitesides, G.M. The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *https://doi.org/10.1146/annurev-biophys-070816-033743* **2018**, *47*, 223–250, doi:10.1146/ANNUREV-BIOPHYS-070816-033743.
62. Peccati, F.; Jiménez-Osés, G. Enthalpy-Entropy Compensation in Biomolecular Recognition: A Computational Perspective. *ACS Omega* **2021**, *6*, 11122–11130, doi:10.1021/ACSOMEGA.1C00485/ASSET/IMAGES/LARGE/AO1C00485_0002.JPEG.
63. Kiessling, L.L.; Pohl, N.L. Strength in Numbers: Non-Natural Polyvalent Carbohydrate Derivatives. *Chem. Biol.* **1996**, *3*, 71–77, doi:10.1016/S1074-5521(96)90280-X.
64. Dam, T.K.; Brewer, C.F. Multivalent Lectin—Carbohydrate Interactions: Energetics and Mechanisms of Binding. *Adv. Carbohydr. Chem. Biochem.* **2010**, *63*, 139–164, doi:10.1016/S0065-2318(10)63005-3.
65. Cecioni, S.; Imberty, A.; Sébastien Vidal, S. Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. **2014**, doi:10.1021/cr500303t.
66. Asensio, J.L.; Cañada, F.J.; Siebert, H.C.; Laynez, J.; Poveda, A.; Nieto, P.M.; Soedjanaamadja, U.; Gabius, H.J.; Jiménez-Barbero, J. Structural Basis for Chitin Recognition by Defense Proteins: GlcNAc Residues Are Bound in a Multivalent Fashion by Extended Binding Sites in Hevein Domains. *Chem. Biol.* **2000**, *7*, 529–543, doi:10.1016/S1074-5521(00)00136-8.
67. Kim, Y.; Hyun, J.Y.; Shin, I. Multivalent Glycans for Biological and Biomedical Applications. *Chem. Soc. Rev.* **2021**, *50*, 10567–10593, doi:10.1039/D0CS01606C.
68. Dam, T.K.; Brewer, C.F. Effects of Clustered Epitopes in Multivalent Ligand-Receptor Interactions. *Biochemistry* **2008**, *47*, 8470–8476, doi:10.1021/BI801208B/ASSET/IMAGES/MEDIUM/BI-2008-01208B_0007.GIF.
69. Bernardi, A.; Jiménez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K.E.; Lahmann, M.; et al. Multivalent Glycoconjugates as Anti-Pathogenic Agents. *Chem. Soc. Rev.* **2013**, *42*, 4709–4727, doi:10.1039/C2CS35408J.
70. Ambrosi, M.; Cameron, N.R.; Davis, B.G. Lectins: Tools for the Molecular Understanding of the Glycocode. *Org. Biomol. Chem.* **2005**, *3*, 1593–1608, doi:10.1039/B414350G.
71. Coelho, L.C.B.B.; Silva, P.M.D.S.; Lima, V.L.D.M.; Pontual, E.V.; Paiva, P.M.G.; Napoleão, T.H.; Correia, M.T.D.S. Lectins, Interconnecting Proteins with Biotechnological/Pharmacological and Therapeutic Applications. *Evidence-based*

Complement. Altern. Med. **2017**, *2017*, doi:10.1155/2017/1594074.

72. Cummings, R.D.; Liu, F.-T.; Rabinovich, G.A.; Stowell, S.R.; Vasta, G.R. Galectins. In *Essentials of Glycobiology*; Cold Spring Harbor Laboratory Press, 2022; pp. 400–420 ISBN 978-1-621824-21-3.
73. Leffler, H.; Carlsson, S.; Hedlund, M.; Qian, Y.; Poirier, F. Introduction to Galectins. *Glycoconj. J.* 2002.
74. Dings, R.P.M.; Miller, M.C.; Griffin, R.J.; Mayo, K.H. Galectins as Molecular Targets for Therapeutic Intervention. *Int. J. Mol. Sci.* **2018**, *19*, doi:10.3390/IJMS19030905.
75. López-Lucendo, M.F.; Solís, D.; André, S.; Hirabayashi, J.; Kasai, K.I.; 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*, 957–970, doi:10.1016/J.JMB.2004.08.078.
76. Bertuzzi, S.; Quintana, J.I.; Ardá, A.; Gimeno, A.; Jiménez-Barbero, J. Targeting Galectins With Glycomimetics. *Front. Chem.* 2020, *8*, 593.
77. Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W.E.G.; et al. Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572*, 232–254, doi:10.1016/S0304-4165(02)00311-2.
78. Sörme, P.; Kahl-Knutsson, B.; Huflejt, M.; Nilsson, U.J.; Leffler, H. Fluorescence Polarization as an Analytical Tool to Evaluate Galectin-Ligand Interactions. *Anal. Biochem.* **2004**, doi:10.1016/j.ab.2004.06.042.
79. Horlacher, T.; Oberli, M.A.; Werz, D.B.; Kröck, L.; Bufali, S.; Mishra, R.; Sobek, J.; Simons, K.; Hirashima, M.; Niki, T.; et al. Determination of Carbohydrate-Binding Preferences of Human Galectins with Carbohydrate Microarrays. *Chembiochem* **2010**, *11*, 1563–1573, doi:10.1002/CBIC.201000020.
80. 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*, 10109–10123, doi:10.1074/jbc.M709545200.
81. Moure, M.J.; Gimeno, A.; Delgado, S.; Diercks, T.; Boons, G.J.; Jiménez-Barbero, J.; Ardá, A. Selective 13C-Labels on Repeating Glycan Oligomers to Reveal Protein Binding Epitopes through NMR: Polylactosamine Binding to Galectins. *Angew. Chemie Int. Ed.* **2021**, *60*, 18777–18782, doi:10.1002/ANIE.202106056.
82. Carlsson, S.; Öberg, C.T.; Carlsson, M.C.; Sundin, A.; Nilsson, U.J.; Smith, D.; Cummings, R.D.; Almkvist, J.; Karlsson, A.; Leffler, H. Affinity of Galectin-8 and Its Carbohydrate Recognition Domains for Ligands in Solution and at the Cell Surface. *Glycobiology* **2007**, *17*, 663–676, doi:10.1093/GLYCOB/CWM026.
83. Ideo, H.; Matsuzaka, T.; Nonaka, T.; Seko, A.; Yamashita, K. Galectin-8-N-Domain Recognition Mechanism for Sialylated and Sulfated Glycans. *J. Biol. Chem.* **2011**, *286*, 11346–11355, doi:10.1074/jbc.M110.195925.
84. Nielsen, M.I.; Stegmayr, J.; Grant, O.C.; Yang, Z.; Nilsson, U.J.; Boos, I.; Carlsson, M.C.; Woods, R.J.; Unverzagt, C.; Leffler, H.; et al. Galectin Binding to Cells and Glycoproteins with Genetically Modified Glycosylation Reveals Galectin-Glycan Specificities in a Natural Context. *J. Biol. Chem.* **2018**, *293*, 20249–20262, doi:10.1074/JBC.RA118.004636.
85. Gómez-Redondo, M.; Delgado, S.; Núñez-Franco, R.; Jiménez-Osés, G.; Ardá, A.; Jiménez-Barbero, J.; Gimeno, A. The Two Domains of Human Galectin-8 Bind Sialyl- and Fucose-Containing Oligosaccharides in an Independent Manner. A 3D View by Using NMR. *RSC Chem. Biol.* **2021**, *2*, 932–941, doi:10.1039/D1CB00051A.
86. Bertuzzi, S.; Gimeno, A.; Núñez-Franco, R.; Bernardo-Seisdedos, G.; Delgado, S.; Jiménez-Osés, G.; Millet, O.; Jiménez-Barbero, J.; Ardá, A. Unravelling the Time Scale of Conformational Plasticity and Allostery in Glycan Recognition by Human Galectin-

1. *Chem. - A Eur. J.* **2020**, *26*, 15643–15653, doi:10.1002/CHEM.202003212.
87. Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M.A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F.J.; et al. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie - Int. Ed.* **2019**, doi:10.1002/anie.201900723.
88. Quintana, J.I.; Delgado, S.; Núñez-Franco, R.; Cañada, F.J.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Ardá, A. Galectin-4 N-Terminal Domain: Binding Preferences Toward A and B Antigens With Different Peripheral Core Presentations. *Front. Chem.* **2021**, *9*, 193, doi:10.3389/FCHEM.2021.664097/BIBTEX.
89. Wu, S.-C.; Kamili, N.A.; Dias-Baruffi, M.; Cummings, R.D.; Stowell, S.R.; Arthur, C.M. Innate Immune Galectin-7 Specifically Targets Microbes That Decorate Themselves in Blood Group-like Antigens. *iScience* **2022**, *25*, 104482, doi:10.1016/j.isci.2022.104482.
90. Arthur, C.M.; Patel, S.R.; Mener, A.; Kamili, N.A.; Fasano, R.M.; Meyer, E.; Winkler, A.M.; Sola-Visner, M.; Josephson, C.D.; Stowell, S.R. Innate Immunity against Molecular Mimicry: Examining Galectin-Mediated Antimicrobial Activity. *Bioessays* **2015**, *37*, 1327–1337, doi:10.1002/BIES.201500055.
91. Thiemann, S.; Baum, L.G. Galectins and Immune Responses-Just How Do They Do Those Things They Do? *Annu. Rev. Immunol* **2016**, *34*, 243–264, doi:10.1146/annurev-immunol-041015-055402.
92. Hughes, R.C. Secretion of the Galectin Family of Mammalian Carbohydrate-Binding Proteins. *Biochim. Biophys. Acta - Gen. Subj.* **1999**, *1473*, 172–185, doi:10.1016/S0304-4165(99)00177-4.
93. Nabi, I.R.; Shankar, J.; Dennis, J.W. The Galectin Lattice at a Glance. *J. Cell Sci.* **2015**, *128*, 2213–2219, doi:10.1242/JCS.151159/259231/AM/THE-GALECTIN-LATTICE-AT-A-GLANCE.
94. Shimura, T.; Takenaka, Y.; Fukumori, T.; Tsutsumi, S.; Okada, K.; Hogan, V.; Kikuchi, A.; Kuwano, H.; Raz, A. Implication of Galectin-3 in Wnt Signaling. *Cancer Res.* **2005**, *65*, 3535–3537, doi:10.1158/0008-5472.CAN-05-0104.
95. Paz, A.; Haklai, R.; Elad-Sfadia, G.; Ballan, E.; Kloog, Y. Galectin-1 Binds Oncogenic H-Ras to Mediate Ras Membrane Anchorage and Cell Transformation. *Oncogene 2001 2051* **2001**, *20*, 7486–7493, doi:10.1038/sj.onc.1204950.
96. Chan, Y.C.; Lin, H.Y.; Tu, Z.; Kuo, Y.H.; Hsu, S.T.D.; Lin, C.H. Dissecting the Structure–Activity Relationship of Galectin–Ligand Interactions. *Int. J. Mol. Sci.* 2018.
97. Johannes, L.; Jacob, R.; Leffler, H. Galectins at a Glance. *J. Cell Sci.* **2018**, *131*, doi:10.1242/JCS.208884/57083.
98. Klyosov, A.A.; Traber, P.G. Galectins in Disease and Potential Therapeutic Approaches. *ACS Symp. Ser.* **2012**, *1115*, 3–43, doi:10.1021/BK-2012-1115.CH001/ASSET/IMAGES/MEDIUM/BK-2012-005234_G005.GIF.
99. Brinchmann, M.F.; Patel, D.M.; Iversen, M.H. The Role of Galectins as Modulators of Metabolism and Inflammation. *Mediators Inflamm.* 2018.
100. Liu, F.T.; Rabinovich, G.A. Galectins: Regulators of Acute and Chronic Inflammation. *Ann. N. Y. Acad. Sci.* **2010**, *1183*, 158–182, doi:10.1111/J.1749-6632.2009.05131.X.
101. Liu, F.T.; Rabinovich, G.A. Galectins as Modulators of Tumour Progression. *Nat. Rev. Cancer* 2005.
102. Girotti, M.R.; Salatino, M.; Dalotto-Moreno, T.; Rabinovich, G.A. Sweetening the Hallmarks of Cancer: Galectins as Multifunctional Mediators of Tumor Progression. *J. Exp. Med.* **2020**, doi:10.1084/jem.20182041.
103. Vasta, G.R. Roles of Galectins in Infection. *Nat. Rev. Microbiol. 2009 76* **2009**, *7*, 424–438, doi:10.1038/nrmicro2146.
104. Hughes, R.C. Galectins as Modulators of Cell Adhesion. *Biochimie* **2001**, *83*, 667–676, doi:10.1016/S0300-9084(01)01289-5.

105. Rabinovich, G.A. Galectin-1 as a Potential Cancer Target. *Br. J. Cancer 2005 927* **2005**, *92*, 1188–1192, doi:10.1038/SJ.BJC.6602493.
106. Camby, I.; Le Mercier, M.; Lefranc, F.; Kiss, R. Galectin-1: A Small Protein with Major Functions. *Glycobiology* **2006**, *16*, 137R-157R, doi:10.1093/GLYCOB/CWL025.
107. 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*, 307–319, doi:10.1016/J.CTRV.2013.07.007.
108. Martínez-Bosch, N.; Navarro, P. Galectins in the Tumor Microenvironment: Focus on Galectin-1. *Adv. Exp. Med. Biol.* **2020**, *1259*, 17–38, doi:10.1007/978-3-030-43093-1_2/COVER.
109. Nambiar, D.K.; Aguilera, T.; Cao, H.; Kwok, S.; Kong, C.; Bloomstein, J.; Wang, Z.; Rangan, V.S.; Jiang, D.; Von Eyben, R.; et al. Galectin-1–Driven T Cell Exclusion in the Tumor Endothelium Promotes Immunotherapy Resistance. *J. Clin. Invest.* **2019**, *129*, 5553–5567, doi:10.1172/JCI129025.
110. Kopitz, J.; Von Reitzenstein, C.; André, S.; Kaltner, H.; Uhl, J.; Ehemann, V.; Cantz, M.; Gabius, H.J. Negative Regulation of Neuroblastoma Cell Growth by Carbohydrate-Dependent Surface Binding of Galectin-1 and Functional Divergence from Galectin-3. *J. Biol. Chem.* **2001**, *276*, 35917–35923, doi:10.1074/jbc.M105135200.
111. Jeschke, U.; Karsten, U.; Wiest, I.; Schulze, S.; Kuhn, C.; Friese, K.; Walzel, H. Binding of Galectin-1 (Gal-1) to the Thomsen-Friedenreich (TF) Antigen on Trophoblast Cells and Inhibition of Proliferation of Trophoblast Tumor Cells in Vitro by Gal-1 or an Anti-TF Antibody. *Histochem. Cell Biol.* **2006**, *126*, 437–444, doi:10.1007/S00418-006-0178-1/FIGURES/4.
112. Sundblad, V.; Morosi, L.G.; Geffner, J.R.; Rabinovich, G.A. Galectin-1: A Jack-of-All-Trades in the Resolution of Acute and Chronic Inflammation. *J. Immunol.* **2017**, *199*, 3721–3730, doi:10.4049/JIMMUNOL.1701172.
113. Horie, H.; Kadoya, T. Galectin-1 Plays Essential Roles in Adult Mammalian Nervous Tissues. Roles of Oxidized Galectin-1. *Glycoconjugate J. 2002 197* **2002**, *19*, 479–489, doi:10.1023/B:GLYC.0000014077.84016.52.
114. Sethi, A.; Sanam, S.; Alvala, R.; Alvala, M. An Updated Patent Review of Galectin-1 and Galectin-3 Inhibitors and Their Potential Therapeutic Applications (2016–Present). *Expert Opin. Ther. Pat.* **2021**, *31*, 709–721, doi:10.1080/13543776.2021.1903430.
115. 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.
116. Blanchard, H.; Yu, X.; Collins, P.M.; Bum-Erdene, K. Galectin-3 Inhibitors: A Patent Review (2008-Present). *Expert Opin. Ther. Pat.* **2014**, doi:10.1517/13543776.2014.947961.
117. Cummings R.D.; Chiffoleau E.; van Kyook, Y.; McEver R.P. C-Type Lectins. In *Essentials of Glycobiology*; Cold Spring Harbor …, 2022; pp. 675–683 ISBN 9784431548416.
118. Valverde, P.; Martínez, J.D.; Cañada, F.J.; Ardá, A.; Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* **2020**, *21*, 2999–3025, doi:10.1002/CBIC.202000238.
119. Drickamer, K.; Taylor, M.E. Recent Insights into Structures and Functions of C-Type Lectins in the Immune System. *Curr. Opin. Struct. Biol.* **2015**, *34*, 26, doi:10.1016/J.SBI.2015.06.003.
120. Arnold, J.N.; Mitchell, D.A. Tinker, Tailor, Soldier, Cell; the Role of C-Type Lectins in the Defense and Promotion of Disease. *Protein Cell* **2022**, doi:10.1093/PROCEL/PWAC012.
121. Brown, G.D.; Willment, J.A.; Whitehead, L. C-Type Lectins in Immunity and Homeostasis. *Nat. Rev. Immunol.* **2018**, *18*, 374–389, doi:10.1038/S41577-018-0004-8.

122. Unger, W.W.J.; Van Beelen, A.J.; Bruijns, S.C.; Joshi, M.; Fehres, C.M.; Van Bloois, L.; Verstege, M.I.; Ambrosini, M.; Kalay, H.; Nazmi, K.; et al. Glycan-Modified Liposomes Boost CD4+ and CD8+ T-Cell Responses by Targeting DC-SIGN on Dendritic Cells. *J. Control. Release* **2012**, *160*, 88–95, doi:10.1016/J.JCONREL.2012.02.007.
123. Chiodo, F.; Marradi, M.; Park, J.; Ram, A.F.J.; Penadés, S.; Van Die, I.; Tefsen, B. Galactofuranose-Coated Gold Nanoparticles Elicit a pro-Inflammatory Response in Human Monocyte-Derived Dendritic Cells and Are Recognized by DC-SIGN. *ACS Chem. Biol.* **2014**, *9*, 383–389, doi:10.1021/CB4008265/SUPPL_FILE/CB4008265_SI_001.PDF.
124. Lepenies, B.; Lee, J.; Sonkaria, S. Targeting C-Type Lectin Receptors with Multivalent Carbohydrate Ligands. *Adv. Drug Deliv. Rev.* **2013**, *65*, 1271–1281, doi:10.1016/J.ADDR.2013.05.007.
125. Mayer, S.; Raulf, M.K.; Lepenies, B. C-Type Lectins: Their Network and Roles in Pathogen Recognition and Immunity. *Histochem. Cell Biol. 2016 1472* **2016**, *147*, 223–237, doi:10.1007/S00418-016-1523-7.
126. Liu, W.; Tang, L.; Zhang, G.; Wei, H.; Cui, Y.; Guo, L.; Gou, Z.; Chen, X.; Jiang, D.; Zhu, Y.; et al. Characterization of a Novel C-Type Lectin-like Gene, LSECtin: Demonstration of Carbohydrate Binding and Expression in Sinusoidal Endothelial Cells of Liver and Lymph Node. *J. Biol. Chem.* **2004**, *279*, 18748–18758, doi:10.1074/jbc.M311227200.
127. Domínguez-Soto, Á.; Aragoneses-Fenoll, L.; Gómez-Aguado, F.; Corcuera, M.T.; Clária, J.; García-Monzón, C.; Bustos, M.; Corbí, A.L. The Pathogen Receptor Liver and Lymph Node Sinusoidal Endothelial Cell C-Type Lectin Is Expressed in Human Kupffer Cells and Regulated by PU.1. *Hepatology* **2009**, *49*, 287–296, doi:10.1002/HEP.22678.
128. Dominguez-Soto, A.; Aragoneses-Fenoll, L.; Martin-Gayo, E.; Martinez-Prats, L.; Colmenares, M.; Naranjo-Gomez, M.; Borras, F.E.; Munoz, P.; Zubiaur, M.; Toribio, M.L.; et al. The DC-SIGN–Related Lectin LSECtin Mediates Antigen Capture and Pathogen Binding by Human Myeloid Cells. *Blood* **2007**, *109*, 5337–5345, doi:10.1182/BLOOD-2006-09-048058.
129. Gramberg, T.; Soilleux, E.; Fisch, T.; Lalor, P.F.; Hofmann, H.; Wheeldon, S.; Cotterill, A.; Wegele, A.; Winkler, T.; Adams, D.H.; et al. Interactions of LSECtin and DC-SIGN/DC-SIGNR with Viral Ligands: Differential PH Dependence, Internalization and Virion Binding. *Virology* **2008**, *373*, 189–201, doi:10.1016/J.VIROL.2007.11.001.
130. Tang, L.; Yang, J.; Liu, W.; Tang, X.; Chen, J.; Zhao, D.; Wang, M.; Xu, F.; Lu, Y.; Liu, B.; et al. Liver Sinusoidal Endothelial Cell Lectin, LSECtin, Negatively Regulates Hepatic T-Cell Immune Response. *Gastroenterology* **2009**, *137*, 1498, doi:10.1053/J.GASTRO.2009.07.051.
131. Xu, F.; Liu, J.; Liu, D.; Liu, B.; Wang, M.; Hu, Z.; Du, X.; Tang, L.; He, F. LSECtin Expressed on Melanoma Cells Promotes Tumor Progression by Inhibiting Antitumor T-Cell Responses. *Cancer Res.* **2014**, *74*, 3418–3428, doi:10.1158/0008-5472.CAN-13-2690/651326/AM/LSECTIN-EXPRESSED-ON-MELANOMA-CELLS-PROMOTES-TUMOR.
132. Liu, D.; Lu, Q.; Wang, X.; Wang, J.; Lu, N.; Jiang, Z.; Hao, X.; Li, J.; Liu, J.; Cao, P.; et al. LSECtin on Tumor-Associated Macrophages Enhances Breast Cancer Stemness via Interaction with Its Receptor BTN3A3. *Cell Res. 2019 295* **2019**, *29*, 365–378, doi:10.1038/s41422-019-0155-6.
133. Zhang, Y.; Feng, Z.; Xu, Y.; Jiang, S.; Zhang, Q.; Zhang, Z.; Wang, K.; Li, X.; Xu, L.; Yuan, M.; et al. Novel Roles of LSECtin in Gastric Cancer Cell Adhesion, Migration, Invasion, and Lymphatic Metastasis. *Cell Death Dis. 2022 137* **2022**, *13*, 1–11, doi:10.1038/s41419-022-05026-x.
134. Gramberg, T.; Hofmann, H.; Möller, P.; Lalor, P.F.; Marzi, A.; Geier, M.; Krumbiegel,

M.; Winkler, T.; Kirchhoff, F.; Adams, D.H.; et al. LSECtin Interacts with Filovirus Glycoproteins and the Spike Protein of SARS Coronavirus. *Virology* **2005**, *340*, 224–236, doi:10.1016/J.VIROL.2005.06.026.

135. Hoffmann, D.; Mereiter, S.; Jin Oh, Y.; Monteil, V.; Elder, E.; Zhu, R.; Canena, D.; Hain, L.; Laurent, E.; Grünwald-Gruber, C.; et al. Identification of Lectin Receptors for Conserved SARS-CoV-2 Glycosylation Sites. *EMBO J.* **2021**, *40*, doi:10.15252/EMBJ.2021108375.
136. Powlesland, A.S.; Fisch, T.; Taylor, M.E.; Smith, D.F.; Tissot, B.; Dell, A.; Pöhlmann, S.; Drickamer, K. A Novel Mechanism for LSECtin Binding to Ebola Virus Surface Glycoprotein through Truncated Glycans *. *J. Biol. Chem.* **2008**, *283*, 593–602, doi:10.1074/JBC.M706292200.
137. Pipirou, Z.; Powlesland, A.S.; Steffen, I.; Pöhlmann, S.; Taylor, M.E.; Drickamer, K. Mouse LSECtin as a Model for a Human Ebola Virus Receptor. *Glycobiology* **2011**, *21*, 806–812, doi:10.1093/GLYCOB/CWR008.
138. Shimojima, M.; Stroher, U.; Ebihara, H.; Feldmann, H.; Kawaoka, Y. Identification of Cell Surface Molecules Involved in Dystroglycan-Independent Lassa Virus Cell Entry. *J. Virol.* **2012**, *86*, 2067–2078, doi:10.1128/JVI.06451-11.
139. Shimojima, M.; Kawaoka, Y. Cell Surface Molecules Involved in Infection Mediated by Lymphocytic Choriomeningitis Virus Glycoprotein. *J. Vet. Med. Sci.* **2012**, *74*, 1363–1366, doi:10.1292/JVMS.12-0176.
140. Zhang, F.; Ren, S.; Zuo, Y. DC-SIGN, DC-SIGNR and LSECtin: C-Type Lectins for Infection. *Int. Rev. Immunol.* **2014**, *33*, 54–66, doi:10.3109/08830185.2013.834897.
141. Paulson, J.C.; MacAuley, M.S.; Kawasaki, N. Siglecs as Sensors of Self in Innate and Adaptive Immune Responses. *Ann. N. Y. Acad. Sci.* **2012**, *1253*, 37–48, doi:10.1111/J.1749-6632.2011.06362.X.
142. Lenza, M.P.; Atxabal, U.; Oyenarte, I.; Jiménez-Barbero, J.; Ereño-Orbea, J. Current Status on Therapeutic Molecules Targeting Siglec Receptors. *Cells 2020, Vol. 9, Page 2691* **2020**, *9*, 2691, doi:10.3390/CELLS9122691.
143. Crocker, P.R.; Paulson, J.C.; Varki, A. Siglecs and Their Roles in the Immune System. *Nat. Rev. Immunol. 2007 74* **2007**, *7*, 255–266, doi:10.1038/nri2056.
144. Duan, S.; Paulson, J.C. Siglecs as Immune Cell Checkpoints in Disease. *https://doi.org/10.1146/annurev-immunol-102419-035900* **2020**, *38*, 365–395, doi:10.1146/ANNUREV-IMMUNOL-102419-035900.
145. Crocker, P.R.; Varki, A. Siglecs in the Immune System. *Immunology* **2001**, *103*, 137, doi:10.1046/J.0019-2805.2001.01241.X.
146. MacAuley, M.S.; Crocker, P.R.; Paulson, J.C. Siglec-Mediated Regulation of Immune Cell Function in Disease. *Nat. Rev. Immunol. 2014 1410* **2014**, *14*, 653–666, doi:10.1038/nri3737.
147. Gonzalez-Gil, A.; Schnaar, R.L. Siglec Ligands. *Cells* **2021**, *10*, 1260, doi:10.3390/CELLS10051260.
148. Rillahan, C.D.; Schwartz, E.; McBride, R.; Fokin, V. V.; Paulson, J.C. Click and Pick: Identification of Sialoside Analogues for Siglec-Based Cell Targeting. *Angew. Chemie Int. Ed.* **2012**, *51*, 11014–11018, doi:10.1002/ANIE.201205831.
149. Whitney, G.; Wang, S.; Chang, H.; Cheng, K.Y.; Lu, P.; Zhou, X.D.; Yang, W.P.; McKinnon, M.; Longphre, M. A New Siglec Family Member, Siglec-10, Is Expressed in Cells of the Immune System and Has Signaling Properties Similar to CD33. *Eur. J. Biochem.* **2001**, *268*, 6083–6096, doi:10.1046/J.0014-2956.2001.02543.X.
150. Forgione, R.E.; Di Carluccio, C.; Guzmán-Caldentey, J.; Gaglione, R.; Battista, F.; Chiodo, F.; Manabe, Y.; Arciello, A.; Del Vecchio, P.; Fukase, K.; et al. Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. *iScience* **2020**, *23*, doi:10.1016/J.ISCI.2020.101231.
151. Chen, G.Y.; Tang, J.; Zheng, P.; Liu, Y. CD24 and Siglec-10 Selectively Repress Tissue Damage - Induced Immune Responses. *Science (80-.).* **2009**, *323*, 1722–1725,

doi:10.1126/SCIENCE.1168988/SUPPL_FILE/CHEN.SOM.PDF.
152. Chen, G.Y.; Tang, J.; Zheng, P.; Liu, Y. CD24 and Siglec-10 Selectively Repress Tissue Damage - Induced Immune Responses. *Science (80-.).* **2009**, *323*, 1722–1725, doi:10.1126/SCIENCE.1168988/SUPPL_FILE/CHEN.SOM.PDF.
153. Chen, G.Y.; Brown, N.K.; Zheng, P.; Liu, Y. Siglec-G/10 in Self–Nonself Discrimination of Innate and Adaptive Immunity. *Glycobiology* **2014**, *24*, 800–806, doi:10.1093/GLYCOB/CWU068.
154. Bandala-Sanchez, E.; Bediaga, N.G.; Goddard-Borger, E.D.; Ngui, K.; Naselli, G.; Stone, N.L.; Neale, A.M.; Pearce, L.A.; Wardak, A.; Czabotar, P.; et al. CD52 Glycan Binds the Proinflammatory B Box of HMGB1 to Engage the Siglec-10 Receptor and Suppress Human T Cell Function. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 7783–7788, doi:10.1073/PNAS.1722056115/SUPPL_FILE/PNAS.1722056115.SAPP.PDF.
155. Yin, S.S.; Gao, F.H. Molecular Mechanism of Tumor Cell Immune Escape Mediated by CD24/Siglec-10. *Front. Immunol.* **2020**, *11*, 1324, doi:10.3389/FIMMU.2020.01324/BIBTEX.
156. Barkal, A.A.; Brewer, R.E.; Markovic, M.; Kowarsky, M.; Barkal, S.A.; Zaro, B.W.; Krishnan, V.; Hatakeyama, J.; Dorigo, O.; Barkal, L.J.; et al. CD24 Signalling through Macrophage Siglec-10 Is a Target for Cancer Immunotherapy. *Nat. 2019 5727769* **2019**, *572*, 392–396, doi:10.1038/S41586-019-1456-0.
157. Angulo, J.; Zimmer, J.; Imberty, A.; Prestegard, J.H. Structural Biology of Glycan Recognition. *Essentials Glycobiol.* **2022**, doi:10.1101/GLYCOBIOLOGY.4E.30.
158. Krengel, U.; Imberty, A. Crystallography and Lectin Structure Database. *Lectins Anal. Technol.* **2007**, 15–50, doi:10.1016/B978-044453077-6/50003-X.
159. Lütteke, T.; Frank, M.; Von Der Lieth, C.W. Data Mining the Protein Data Bank: Automatic Detection and Assignment of Carbohydrate Structures. *Carbohydr. Res.* **2004**, *339*, 1015–1020, doi:10.1016/j.carres.2003.09.038.
160. Atanasova, M.; Bagdonas, H.; Agirre, J. Structural Glycobiology in the Age of Electron Cryo-Microscopy. *Curr. Opin. Struct. Biol.* **2020**, *62*, 70–78, doi:10.1016/J.SBI.2019.12.003.
161. Bai, X. chen; McMullan, G.; Scheres, S.H.W. How Cryo-EM Is Revolutionizing Structural Biology. *Trends Biochem. Sci.* **2015**, *40*, 49–57, doi:10.1016/J.TIBS.2014.10.005.
162. Duverger, E.; Lamerant-Fayel, N.; Frison, N.; Monsigny, M. Carbohydrate-Lectin Interactions Assayed by SPR. *Methods Mol. Biol.* **2010**, *627*, 157–178, doi:10.1007/978-1-60761-670-2_10/COVER.
163. Wallner, J.; Sissolak, B.; Sommeregger, W.; Lingg, N.; Striedner, G.; Vorauer-Uhl, K. Lectin Bio-Layer Interferometry for Assessing Product Quality of Fc- Glycosylated Immunoglobulin G. *Biotechnol. Prog.* **2019**, *35*, e2864, doi:10.1002/BTPR.2864.
164. Blixt, O.; Collins, B.E.; van den Nieuwenhof, I.M.; Crocker, P.R.; Paulson, J.C. Sialoside Specificity of the Siglec Family Assessed Using Novel Multivalent Probes. *J. Biol. Chem.* **2003**, *278*, 31007–31019, doi:10.1074/jbc.m304331200.
165. Alvarez, R.A.; Blixt, O. Identification of Ligand Specificities for Glycan-Binding Proteins Using Glycan Arrays. *Methods Enzymol.* **2006**, *415*, 292–310, doi:10.1016/S0076-6879(06)15018-1.
166. Gao, C.; Wei, M.; McKitrick, T.R.; McQuillan, A.M.; Heimburg-Molinaro, J.; Cummings, R.D. Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* **2019**, *7*, 833, doi:10.3389/FCHEM.2019.00833/BIBTEX.
167. Song, X.; Heimburg-Molinaro, J.; Cummings, R.D.; Smith, D.F. Chemistry of Natural Glycan Microarrays. *Curr. Opin. Chem. Biol.* **2014**, *18*, 70–77, doi:10.1016/J.CBPA.2014.01.001.
168. Grant, O.C.; Smith, H.M.K.; Firsova, D.; Fadda, E.; Woods, R.J. Presentation, Presentation, Presentation! Molecular-Level Insight into Linker Effects on Glycan

Array Screening Data. *Glycobiology* **2014**, *24*, 17–25, doi:10.1093/GLYCOB/CWT083.
169. Wisnovsky, S.; Bertozzi, C.R. Reading the Glyco-Code: New Approaches to Studying Protein–Carbohydrate Interactions. *Curr. Opin. Struct. Biol.* **2022**, *75*, doi:10.1016/J.SBI.2022.102395.
170. Velázquez-Campoy, A.; Ohtaka, H.; Nezami, A.; Muzammil, S.; Freire, E. Isothermal Titration Calorimetry. *Curr. Protoc. Cell Biol.* **2004**, *23*, 17.8.1-17.8.24, doi:10.1002/0471143030.CB1708S23.
171. Takeda, Y.; Matsuo, I. Isothermal Calorimetric Analysis of Lectin-Sugar Interaction. *Methods Mol. Biol.* **2014**, *1200*, 207–214, doi:10.1007/978-1-4939-1292-6_18/COVER.
172. Sorek, R.; Kunin, V.; Hugenholtz, P. CRISPR--a Widespread System That Provides Acquired Resistance against Phages in Bacteria and Archaea. *Nat. Rev. Microbiol.* **2008**, *6*, 181–186, doi:10.1038/NRMICRO1793.
173. Marinova, I.N.; Wandall, H.H.; Dabelsteen, S. Protocol for CRISPR-Cas9 Modification of Glycosylation in 3D Organotypic Skin Models. *STAR Protoc.* **2021**, *2*, 100668–100668, doi:10.1016/J.XPRO.2021.100668.
174. James Keeler *Understanding NMR Spectroscopy*; Second Edition.; Wiley, 2010;
175. Ven, F.J.M. van de *Multidimensional NMR in Liquids : Basic Principles and Experimental Methods*; 1st edition.; Wiley-VCH, 1995; ISBN 978-0-471-18594-9.
176. Sattler, M.; Schleucher, J.; Griesinger, C. Heteronuclear Multidimensional NMR Experiments for the Structure Determination of Proteins in Solution Employing Pulsed Field Gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158, doi:10.1016/S0079-6565(98)00025-9.
177. Duus, J.; Gotfredsen, C.H.; Bock, K. Carbohydrate Structural Determination by NMR Spectroscopy: Modern Methods and Limitations. *Chem. Rev.* **2000**, *100*, 4589–4614, doi:10.1021/CR990302N/ASSET/CR990302N.FP.PNG_V03.
178. Ardá, A.; Jiménez-Barbero, J. The Recognition of Glycans by Protein Receptors. Insights from NMR Spectroscopy. *Chem. Commun.* **2018**, *54*, 4761–4769, doi:10.1039/C8CC01444B.
179. Valverde, P.; Quintana, J.I.; Santos, J.I.; Ardá, A.; Jiménez-Barbero, J. Novel NMR Avenues to Explore the Conformation and Interactions of Glycans. *ACS Omega* **2019**, *4*, 13618–13630, doi:10.1021/ACSOMEGA.9B01901/ASSET/IMAGES/LARGE/AO9B01901_0009.JPEG.
180. Atxabal, U.; Gimeno, A.; Jiménez-Barbero, J. Nuclear Magnetic Resonance Techniques for the Study of Glycan Interactions. *Compr. Glycosci. Second Ed.* **2021**, 329–345, doi:10.1016/B978-0-12-819475-1.00015-8.
181. Gimeno, A.; Valverde, P.; Ardá, A.; Jiménez-Barbero, J. Glycan Structures and Their Interactions with Proteins. A NMR View. *Curr. Opin. Struct. Biol.* **2020**, *62*, 22–30, doi:10.1016/J.SBI.2019.11.004.
182. Poveda, A.; Jiménez-Barbero, J. NMR Studies of Carbohydrate–Protein Interactions in Solution. *Chem. Soc. Rev.* **1998**, *27*, 133–144, doi:10.1039/A827133Z.
183. Unione, L.; Galante, S.; Díaz, D.; Cañada, F.J.; Jiménez-Barbero, J. NMR and Molecular Recognition. The Application of Ligand-Based NMR Methods to Monitor Molecular Interactions. *Medchemcomm* **2014**, *5*, 1280–1289, doi:10.1039/C4MD00138A.
184. Peters, T. Transfer NOE Experiments for the Study of Carbohydrate-Protein Interactions. *Carbohydrates Chem. Biol.* **2008**, *2–4*, 1003–1023, doi:10.1002/9783527618255.CH35.
185. Arepalli, S.R.; Glaudemans, C.P.J.; Daves, G.D.; Kovac, P.; Bax, A. Identification of Protein-Mediated Indirect Noe Effects in a Disaccharide-Fab′ Complex by Transferred ROESY. *J. Magn. Reson. Ser. B* **1995**, *106*, 195–198, doi:10.1006/JMRB.1995.1033.

186. Perrin, C.L.; Dwyer, T.J. Application of Two-Dimensional NMR to Kinetics of Chemical Exchange. *Chem. Rev.* **2002**, *90*, 935–967, doi:10.1021/CR00104A002.
187. Latham, M.P.; Zimmermann, G.R.; Pardi, A. NMR Chemical Exchange as a Probe for Ligand-Binding Kinetics in a Theophylline-Binding RNA Aptamer. *J. Am. Chem. Soc.* **2009**, *131*, 5052–5053, doi:10.1021/JA900695M/SUPPL_FILE/JA900695M_SI_001.PDF.
188. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788, doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.
189. Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, doi:10.1021/ja0100120.
190. Viegas, A.; Manso, J.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, doi:10.1021/ed101169t.
191. Monaco, S.; Tailford, L.E.; Juge, N.; Angulo, J. Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein–Ligand Contacts. *Angew. Chemie Int. Ed.* **2017**, *56*, 15289–15293, doi:10.1002/ANIE.201707682.
192. Ardá, A.; Blasco, P.; Varón Silva, D.; Schubert, V.; André, S.; Bruix, M.; Cañada, F.J.; Gabius, H.J.; Unverzagt, C.; Jiménez-Barbero, J. Molecular Recognition of Complex-Type Biantennary N-Glycans by Protein Receptors: A Three-Dimensional View on Epitope Selection by NMR. *J. Am. Chem. Soc.* **2013**, *135*, 2667–2675, doi:10.1021/JA3104928.
193. Mcintosh, L.P.; Dahlquist, F.W. Biosynthetic Incorporation of 15N and 13C for Assignment and Interpretation of Nuclear Magnetic Resonance Spectra of Proteins. *Q. Rev. Biophys.* **1990**, *23*, 1–38, doi:10.1017/S0033583500005400.
194. Williamson, M.P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16, doi:10.1016/J.PNMRS.2013.02.001.
195. Schütz, S.; Sprangers, R. Methyl TROSY Spectroscopy: A Versatile NMR Approach to Study Challenging Biological Systems. *Prog. Nucl. Magn. Reson. Spectrosc.* **2020**, *116*, 56–84, doi:10.1016/J.PNMRS.2019.09.004.
196. Hwang, T.L.; Van Zijl, P.C.M.; Mori, S. Accurate Quantitation of Water-Amide Proton Exchange Rates Using the Phase-Modulated CLEAN Chemical EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *J. Biomol. NMR* **1998**, *11*, 221–226, doi:10.1023/A:1008276004875.
197. Hwang, T.L.; Mori, S.; Shaka, A.J.; Van Zijl, P.C.M. Application of Phase-Modulated CLEAN Chemical EXchange Spectroscopy (CLEANEX-PM) to Detect Water - Protein Proton Exchange and Intermolecular NOEs. *J. Am. Chem. Soc.* **1997**, *119*, 6203–6204, doi:10.1021/JA970160J/ASSET/JA970160J.FP.PNG_V03.
198. Singh, A.; Purslow, J.A.; Venditti, V. 15N CPMG Relaxation Dispersion for the Investigation of Protein Conformational Dynamics on the Ms-ms Timescale. *J. Vis. Exp.* **2021**, *2021*, doi:10.3791/62395.

CHAPTER 2

OBJECTIVES

The global objectives set for this Thesis were focused both on the training and scientific perspectives. The work presented in this manuscript has been performed primarily in the centre CIC bioGUNE (Bilbao), under regime of cotutele between the University of Basque Country (UPV/EHU, main institution) and the University of Milano Bicocca (Unimib), and this aspect has contributed to enrich both formative and scientific goals.

From the training point of view, the key objective was to acquire knowledge on the factors that govern lectin-carbohydrate interactions from different perspectives and using diverse methodologies, not only *in vitro* but also in the cell-like or cellular contexts. To achieve this goal, a panel of different techniques have been learned and employed, with major focus on NMR-based protocols. In the last year of the Thesis a great effort has been placed into the design and subsequent application of strategies that provide information on the recognition events in natural-like contexts. Apart from the training in the study of protein-sugar binding events, the formation was enriched with the achievement of skills in lectin expression, purification, cellular culture and cellular engineering. Soft skills on writing scientific papers, preparing and delivering poster and oral presentations, participation in scientific discussions, and sharing resources within multidisciplinary laboratories and Institutes were also developed.

Moving to the scientific point of view, the main goal of this Thesis was to decipher the fine structural and dynamic details of the interaction between different lectins and their corresponding natural or synthetic ligands. In that regard, the objective was to apply, in a synergic manner, different experimental methodologies and combine them with theoretical predictions.

Moreover, one key aim was to find ways to move from the standard reductionistic approach employed in chemistry labs with *in vitro* experiments towards the natural-like context, trying to analyse the lectin/glycan interaction events in systems related to the cell environment and eventually, to finish with intact cells. The final aim was to find avenues to break the limits of current knowledge in carbohydrate recognition, trying to open new possibilities for Glycosciences.

CHAPTER 3

THE MOLECULAR RECOGNITION OF THE HISTO BLOOD GROUP ANTIGENS BY GALECTIN-1

The time scale of conformational plasticity and allostery

3.1 Introduction

A synergistic combination of different NMR methodologies, molecular dynamics (MD) simulations, and isothermal titration calorimetry (ITC) has been employed to provide the structural, conformational, dynamic and energy details of the recognition of the histo blood group antigens by human galectin-1.
The results presented in this chapter have been successfully published in 2020. [1] The histo blood group antigens and their constituting fragments have been employed as recognition targets within the framework of a global project on the interaction between self-antigens by galectins. [2–4]

The MD simulations discussed in this work were performed by the Computational Chemistry Lab at CIC bioGUNE (https://www.cicbiogune.es/people/gjoses) and additional information on the protocols regarding the MD simulations can be found in the publication. [1]

3.1.1. General context

As described in the general Introduction, galectin-1 is a member of the galectin family. [5] This lectin is ubiquitously distributed throughout the body and it participates in the regulation of a wide range of biological phenomena through its binding to glycoproteins and glycolipids. [6] It has been described that galectin-1 participates in B-cell development and signalling, in T-cell immunity and in the regulation of diverse inflammatory responses. [7–9] Moreover, changes in its expression have been correlated with the progression of different cancers. [10] For example, a mechanism of immune evasion mediated by galectin-1 has been proposed to take place through the endothelial compartment that negatively affects antitumor immune responses in head and neck cancer (HNC) mouse models [11]. Indeed, a high concentration of circulating galectin-1 is associated with poor clinical outcomes after specific antitumor therapy in HNC patients. [12]
In the tumour microenvironment, galectin-1 recognizes aberrant mucin-1 glycans presented on the surface of cancer cells. [13] Moreover, it interacts with glycosylated receptors on endothelial cells, triggering tumorigenic cell signalling pathways that induce their proliferation, migration, and activation. [14,15]

Considering its implication in pathogenic events, there is an increasing interest in exploiting galectin-1 for therapeutic purposes. [16–18] In light of this, protein-ligand binding studies, shedding light into the structural details of the interaction, may provide unique elements for the design of inhibitors.

Concerning the protein's architecture, galectin-1 is a prototype galectin formed by two identical monomers of 14.5 kDa each (Figure 3.1). Both homodimers are capable of binding sugar epitopes and the binding sites are located in opposite directions. The overall structure is composed by two anti-parallel β-sheets (five F and six S strands), with the so called jellyroll folding pattern, conserved among galectins. [5,6]
Galectin-1 dimerizes in a non-covalent manner in solution with an equilibrium constant in the low micromolar range. [19] The integrity of the dimer is maintained by interactions between the monomers at the interface (F1 and S1 strands), as well as by hydrophobic residues that cluster to form a common hydrophobic core. [6] It has been postulated that its intrinsic ability to cross-link glycoconjugates may be strongly connected to the biological activity. [20,21] In fact, galectin-1 mutants with modified dimerization properties display biological altered functions. [22–24]

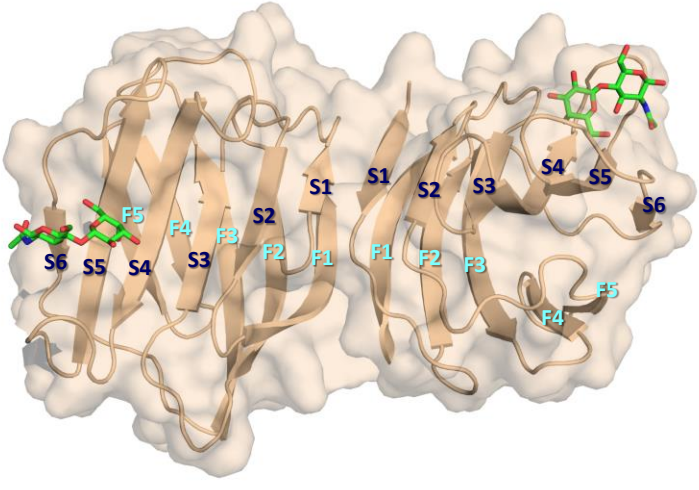


Figure 3.1. Galectin-1 homodimer structure in complex with lactose (PDB ID: 1GZW). F and S sheets are indicated in light blue and black respectively.

β-galactosides are one of the most common epitopes presented at the end of mammalian glycans; however, subtle modifications on their structure can drastically

impact on the affinity depending on the galectin. Galectin-1, in particular, besides the basic disaccharide ligands such as lactose,(Galβ1-4Glc), LacNAc (Galβ1-4GlcNAc), and LacdiNAc (GalNAcβ1-4GlcNAc), has shown affinity to epitopes based on (poly-)N-acetyllactosamine and blood-group oligosaccharides. [25,26]

The carbohydrate binding site of galectin-1 is very similar to that observed in other members of the galectin family, and comprises S4, S5 and S6a/S6b strands. The canonical interaction with the β-Gal moiety occurs through several hydrogen bonds involving the hydroxyl groups of β-Gal and several lectin residues (His44, Asn46, Arg48, Val59, Asn61, Glu71, and Arg73), as well as through the CH-π stacking interaction between the non-polar face of the β-Gal and Trp68. [5,27,28]

The L4 loop, that connects strands S4 and S5, is a unique feature of this galectin and it has been described to be extremely flexible in the absence of ligand, populating two main conformations called close and open. [29]

Most of the knowledge on how galectin-1 binds to carbohydrates derives from X-ray structures. However, many **dynamic aspects** of the interaction remain elusive when using this technique. Conformational plasticity and allostery are, among others, crucial events that can be investigated exclusively with techniques that allow the detection of dynamic features, such as NMR.

In line with this issue, the binding of galectin-1 to lactose has been recently scrutinized through NMR-based hydrogen-deuterium exchange experiments. [30] The results of the study highlighted how residues located in an opposite way respect to the binding-site display increased exchange rates upon binding that correlate with increased backbone dynamics. [30] Such an outcome can be explained by the presence of allosteric communication along the monomers.

Thus, we decided to pay particular attention to the dynamic details of the molecular interaction. In fact, despite the binding preference of galectins have been extensively studied, few works have focused on the dynamic aspects of the interaction.

Our investigation followed a similar approach to that used in a previous work conducted in our laboratory that explored the recognition of galectin-3 (a monomer) with the A and B blood group antigens and their fragments. [2] Specifically, the employed ligands to unravel the recognition features of galectin-1 (dimer) are N-

acetyllactosamine (**LacNAc**), the **Galili** and blood **H type-II** trisaccharides and the blood **B type-II** tetrasaccharide (Figure 3.2).

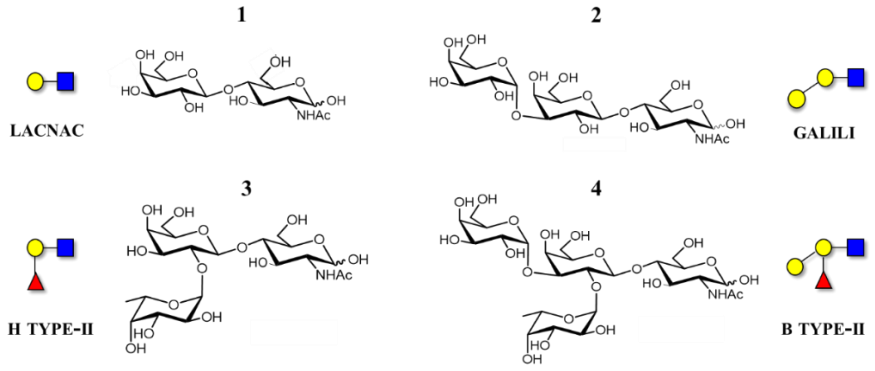


Figure 3.2. Structure, symbol (SNFG) representation, and nomenclature of the oligosaccharides employed in this chapter. The ligands' numeration (**1**-**4**) has been adopted only throughout this chapter.

3.2 Isothermal titration calorimetry experiments

As starting point, the thermodynamic profiles of the interaction of galectin-1 with all the ligands were deduced through ITC measurements. The binding isotherms measured from the titration experiments were adjusted to two different binding models: a single-site and a sequential binding model, considering two identical or different binding events at each site of the dimer. [31] The obtained parameters are presented in Tables 3.1 and 3.2, respectively. The comparison of the χ^2 value (Table 3.2), representing the quality of the fitting, revealed that the sequential binding model provided the best quality or equal to that of the single-site.

Ligand	**ΔG** (kcal/mol)	**ΔH** (kcal/mol)	**-TΔS** (kcal/mol)	**K_D** (μM)
LacNAc (**1**)	-5.5	-5.3	-0.186	99
Galili (**2**)	-5.5	-5.2	-0.385	95
H type-II (**3**)	-4.3	-4.3	-0.469	319
B type-II (**4**)	-4.7	-4.3	-0.388	379

Table 3.1. Thermodynamic parameters from ITC data fitted to single-site binding model

	One-site model		Sequential model		
Ligand	K_D	χ^2	K_{D1}	K_{D2}	χ^2
LacNAc (**1**)	99	1365	17	264	317
Galili (**2**)	95	2309	34	536	1151
H type-II (**3**)	319	169	196	1100	173
B type-II (**4**)	379	69	116	422	74

Table 3.2. Comparison of the affinity parameters obtained by fitting of the ITC data to one-site (left) and sequential binding models (right).

The **binding constants (K_D)** deduced using the single-site model were in the high-medium micromolar range, being **LacNAc** (99 μM) and **Galili** (95 μM) the best binders. Moreover, the affinity obtained for LacNAc was in line with that already published. [32] The affinity and enthalpy values of **Galili** for galectin-1 are very similar to those observed for LacNAc, suggesting that the additional α-Gal residue does not establish additional favourable contacts. In contrast, the **H type-II** and **B type-II** antigens displayed lower binding affinities and enthalpy values compared with the disaccharide, suggesting that the introduction of the fucose moiety decrease the affinity by destabilizing some interactions.

The K_D estimates derived by sequential fitting show a trend for negative cooperativity. Indeed, the first binding event to one monomer of galectin-1 is more favourable than the second one. This phenomenon has already been proposed for lactose recognition. [29] However, for **LacNAc** and **Galili**, this tendency is even more marked: K_{D1} is more than an order of magnitude better than K_{D2}. Again, LacNAc and Galili displayed comparable affinities, whereas **H type-II** and **B type-II** are weaker binders (Table 3.2).

Focusing on the **thermodynamic aspects**, the binding enthalpy remained unaltered passing from the disaccharide to the trisaccharide, suggesting the absence of additional stabilizing intramolecular contacts between the lectin and the α-Gal epitope (Figure 3.3, above). The enthalpy decreased with the fucosylated ligands, lowering the binding free energy. Overall, the enthalpy contribution did not correlate with the entropy, as usually is observed in sugar binding events, which follow the enthalpy-entropy paradigm. [33]

An unusual positive entropy contribution was observed for all the ligands tested. This atypical result differs from the common trend observed for protein-sugar interactions that usually display large and positive binding enthalpies counteracted by unfavourable entropic contributions. [33] The gain from the entropic term is rather moderate (below 0.5 kcal mol^{-1}, Table 3.1), but the key point is not the absolute value, as much as that the entropy term is not opposing the binding event. In line with these results, a previous report highlighted that lactose binds to galectin-1 with a small positive binding entropy and with negative cooperativity. [29]

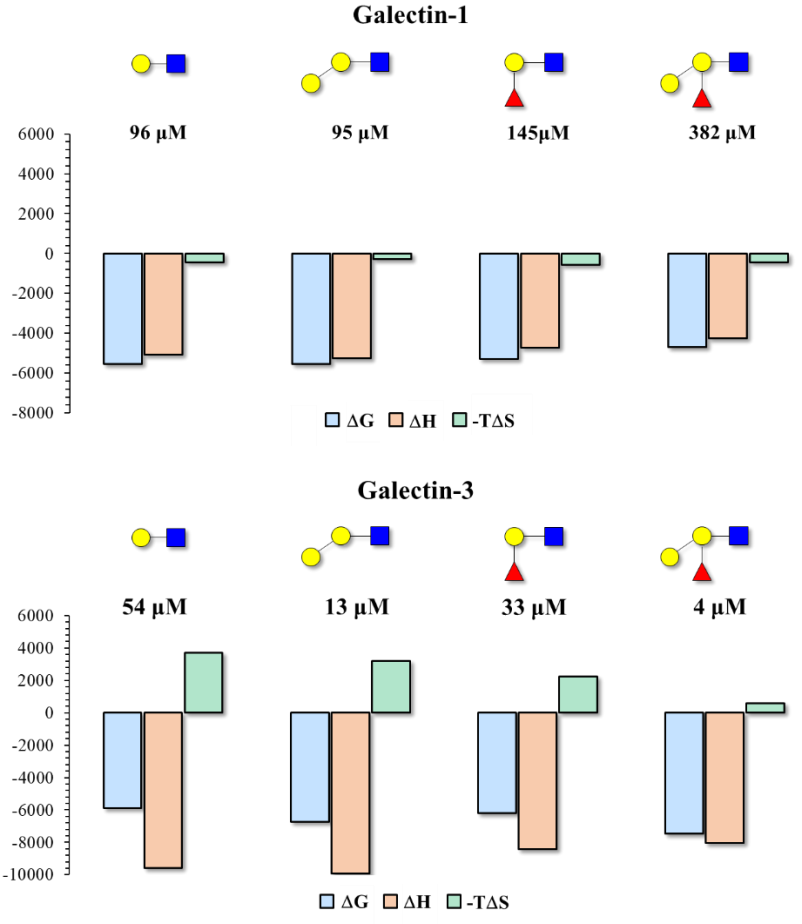


Figure 3.3. Thermodynamic profile of ligands **1-4** obtain by ITC measurements fitted to one-site binding model of galectin-1 (above) and galectin-3 (below).

Intriguingly, the ITC data strongly differ from our previous results obtained for the interaction of the same antigens with the monomer galectin-3, thus highlighting the

different molecular recognition mechanisms employed by the two lectins of the same family (Figure 3.3). [2] In particular, for galectin-3, the higher affinities observed only for the fucosylated tetrasaccharide antigens were justified by the existing pre-organization of these ligands that bind in their free-state conformation to the lectin, facilitating the binding event through a very favourable k_{on} kinetic constant, a decrease of the energy barrier, with the concomitant impact on the binding entropy. [2] However, the comparison of the thermodynamic profiles (Figure 3.3) evidences that this is not the case for galectin-1. Herein, independently of the size of the ligand, the entropy term is basically identical. Such a discrepancy strongly suggests that different conformational dynamics are occurring with galectin-1.
Thus, further investigations were carried out.

3.3 The predictions: Generation of 3D models through MD simulations

The X-ray crystallographic structure of the galectin-1 dimer complexed with lactose (PDB ID: 1GZW) was used as starting point to build the 3D models of the sugar/lectin complexes. [27] Firstly, a standard manual docking protocol using lactose as fixed scaffold to fit the Galβ4GlcNAc core of ligands **1**-**4** was employed. With the generated complexes, short MD simulations of 100 ns were performed by the Computational Chemistry Lab (CIC bioGUNE). The motions around the glycosidic linkages as well as the intermolecular hydrogen bonds (HB) were scrutinized along the MD simulations (Figure 3.4).

The obtained models for galectin-1 and **LacNAc** nicely matched with that obtained by X-ray crystallography with lactose: the key CH-π interaction is provided by the indole moiety of Trp68 stacking with the β-Gal ring, while stable HB are formed between His44, Arg48, Asn61 and Glu71 lectin residues with Gal O4, Gal O5 and GlcNAc O3 of the disaccharide. [28] These interactions are kept in all complexes, with the obvious exception of those involving the positions substituted by additional sugar moieties.
The L4 loop is kept into the close conformation, folded towards the ligand. His52, in the L4 loop, is transiently involved in HBs with β-Gal 2-OH (25% of the MD trajectory). This interaction is also present in the X-ray structure with lactose.

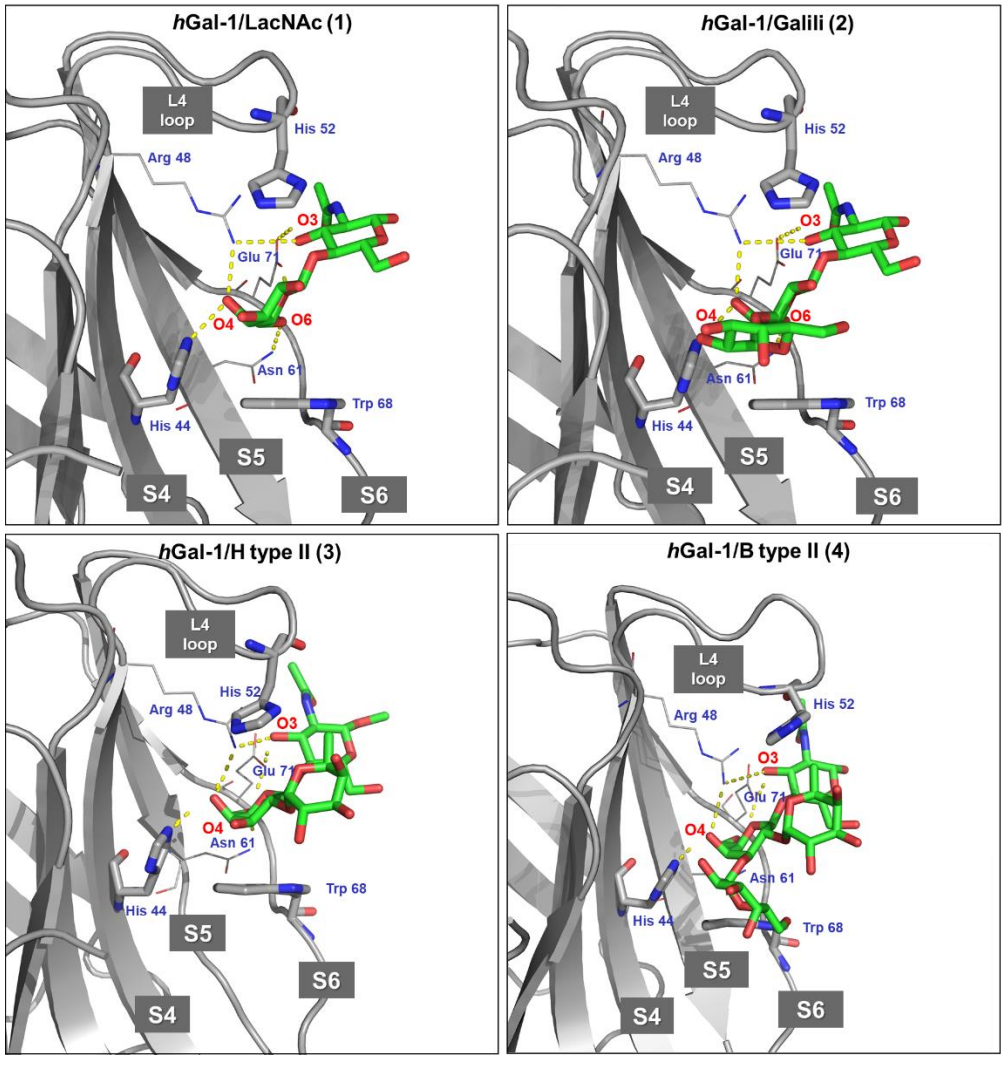


Figure 3.4. Proposed model of the complexes: from left to right and top to bottom: galectin-1/LacNAc, galectin-1/Galili, galectin-1/H type-II and galectin-1/B type-II. The key amino acids involved in the interaction are displayed as well as the most representative intramolecular HBs (yellow dots).

The **Galili**/galectin-1 complex showed that the additional α-Gal residue was close to the protein surface, but without providing additional interactions, in full agreement with the ITC data. Other interactions were identical to those described for LacNAc.

For both **H type-II** and **B type-II** antigens**,** the Fuc moiety is close to the L4 loop; in particular, Fuc O5 and/or Fuc 4-OH established transient HB with His52 (25% of the MD trajectory). Again, the interactions with the LacNAc scaffold are predicted.

Thus, the MD simulations predict that all the saccharides bind galectin-1 in a similar fashion, with only the LacNAc core involved in stabilizing interactions with the lectin.

3.4 The experimental observations: NMR

Two different NMR approaches were employed: from the point of view of the ligand (STD-NMR experiments) and from the point of view of the lectin (CSP-NMR, CLEANEX, and relaxation dispersion experiments).

3.4.1 The ligand's perspective: Saturation transfer difference NMR experiments (STD-NMR)

The initial structural experimental information related to the interaction between human galectin-1 and ligands **1-4** was obtained through ^{1}H STD-NMR experiments. [34–36] As described in the Introduction, this methodology allows deducing the binding epitope of the ligand. In all cases, two sets of STD-NMR experiments were registered, irradiating the aromatic and the aliphatic protons of the lectin for well-defined ligand:lectin ratios of each complex.

Clear STD intensities were observed, confirming the existence of interaction between galectin-1 and all the ligands (Figure 3.5 and supporting information). Important STD signals for H4, H5, and H6 of the central **β-Gal** were detected. This evidence is in accordance with the typical pattern found for the recognition of saccharides by galectins. [5] For the ligands containing the **α-Gal** moiety (Galili and B type-II), clear STD contributions arising from H1, H2 and H3 protons of this sugar were also visible. In addition, considerable STD signals were found for H1 of the **α-Fuc** residue of ligands H type-II and B type II ligands. Interestingly, upon aromatic irradiation (δ 7.67 ppm), a stronger signal for Fuc H1 was observed (See top panel of Figure 3.5) than for the aliphatic irradiation (δ 0.55 ppm). This experimental result can be justified by the proximity of the Fuc moiety to the L4 loop and, in particular, to the aromatic His52, as also predicted by the MD-based 3D models (Figure 3.4). Nevertheless, it has to be remembered that the introduction of the Fuc did not increase the affinity respect to the disaccharide (Figure 3.3).

Therefore, even if the β-Gal ring is the main binding epitope, also α-Gal/α-Fuc moieties are also in close vicinity to the lectin surface. These observations provide the experimental validation of the proposed 3D complexes, where the α-Gal and α-Fuc residues are adjacent to the binding site, providing few transient interactions. This

recognition feature differs from those reported for human galectin-3, where the α-Fuc was exposed to the solvent without taking any contact with the protein surface. [2] Thus, the STD NMR data corroborate the model obtained by docking and MD simulations, contributing to enrich the general picture of the interaction.

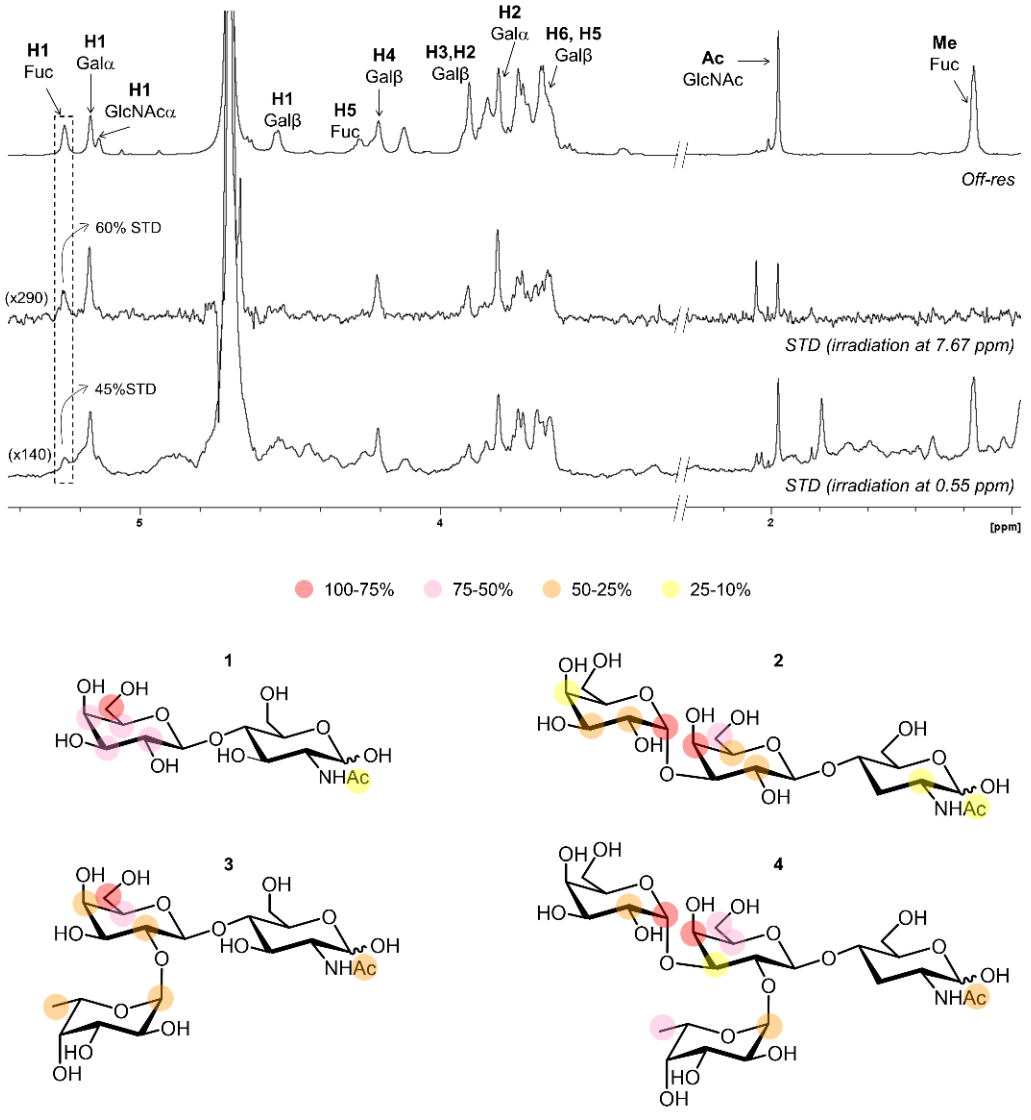


Figure 3.5. ^{1}H-STD-NMR results. *Above*: NMR spectra registered for unravelling the interaction of galectin-1 with the B type-II antigen. On top, reference spectrum with annotations of the ^{1}H signals showing STD effect. Middle: STD NMR spectrum with lectin irradiation at the aromatic region (δ 7.67 ppm). Below: STD NMR spectrum with irradiation at the aliphatic region (δ 0.55 ppm). The relative STD amplification factor is indicated and the signal arising from Fuc H1 is highlighted. *Below*: Epitope mapping obtained from the analysis of the STD NMR experiments (under aliphatic irradiation) with **1**, **2**, **3**, and **4**.

3.4.2 The lectin's perspective: Chemical shift perturbation analysis (CSP-NMR)

In a first step to characterise the interactions from the protein's perspective, ^{1}H-^{15}N HSQC NMR experiments were employed and the chemical shift perturbations (CSP) of the amide signals of the lectin upon ligand addition were monitored. [37]

Briefly, to a sample containing the ^{15}N-labelled protein alone, increasing equivalents of ligands **1-4** were respectively added, and ^{1}H-^{15}N HSQC spectra were acquired at each point. The presence of the ligand modifies the chemical environment of the amino acids directly or indirectly involved in the binding. This perturbation induces a chemical shift perturbation (CSP) of the cross peaks of the protein that are analysed and translated into a unique information on the recognition event at atomic level. Indeed, with this NMR protein-based approach, each subtle environment modification around the lectin's residues upon binding can be identified. The CSP plots obtained for **1-4** are gathered in Figure 3.6.

For the titration with **LacNAc** (black histogram in Figure 3.6), most of the perturbed residues were included in the region 46Asn-76Val (strands S4-S6 and L4 loop). These data agree with the proposed binding mode described above (Figure 3.4) and with the available X-ray crystallography structures (PDB ID: 4Y1U, 4Q26 and 1W6P). Curiously, perturbations on several amino acids at the F-sheets, especially F3 and F4 and in loop L1 (connecting S1 and F2) were also observed (Figure 3.6 and supporting information). These results indicated that the interaction with LacNAc induces changes on the whole structure of the protein, as previously reported for its binding with lactose and with lacto-*N*-neotetraose. [29,30,38]

The titration with **Galili** provided a similar CSP pattern to that obtained with LacNAc: both ligands bind to the protein in a similar fashion (Figure 3.6 and supporting information) with no additional interactions of the αGal epitope, as suggested by the ITC experimental data and the MD simulations.

Remarkably, the addition of **H type-II** and **B type-II** produced additional perturbations for amino acids Ala51-Ala55 at the L4 loop (Figure 3.6 and Figure 3.7). This

experimental evidence is in agreement with the STD NMR data and with the proposed binding pose, where the Fuc residue is positioned in close contact with this L4 loop.

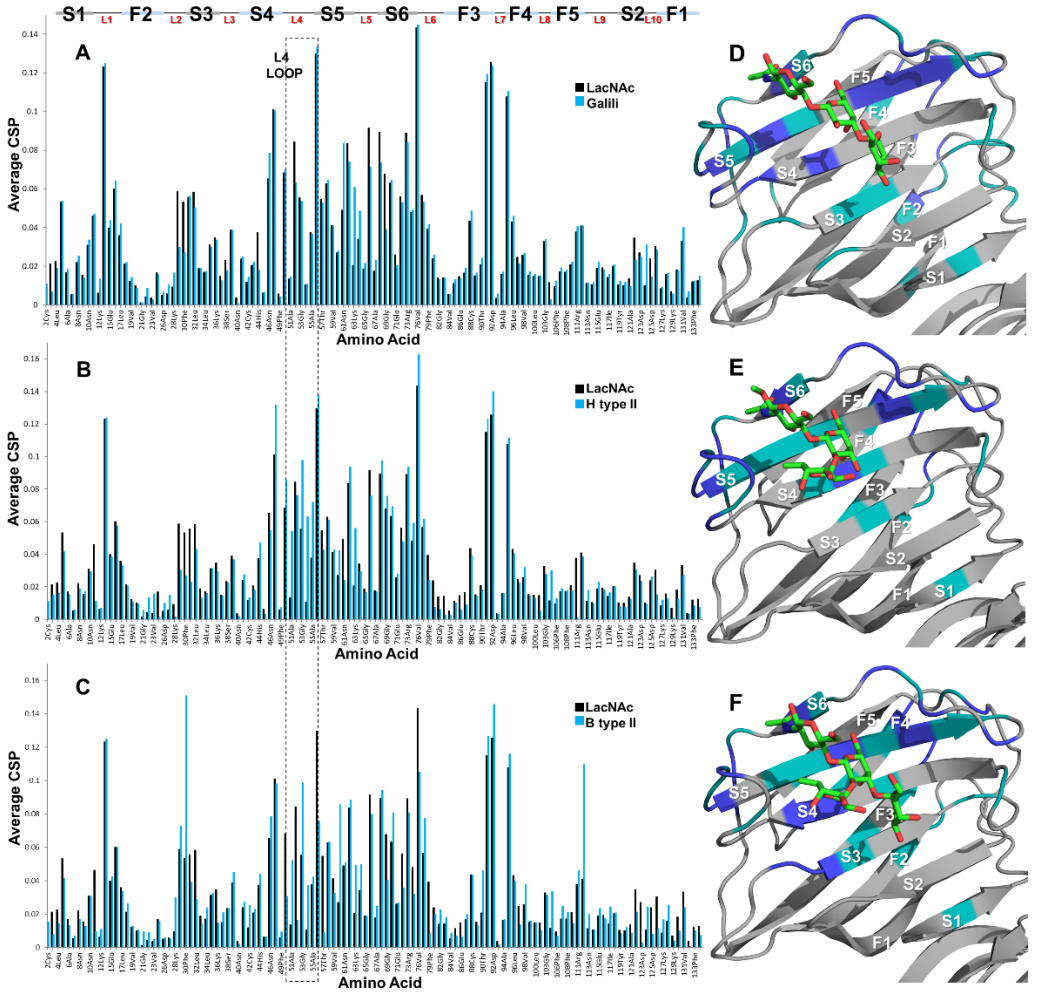


Figure 3.6. CSP plot of the backbone amide signals of galectin-1 obtained upon addition of LacNAc (in black, 10 equivalents) or, in blue, of A) Galili (10 eq), B) H type-II (15 eq), and C) B type-II (15 eq). The most perturbed residues are mapped into the MD-derived 3D models (D-F). In the 3D models, the dark blue coloration refers to CSP value over 2σ, whereas the light blue one to values between 1σ and 2σ.

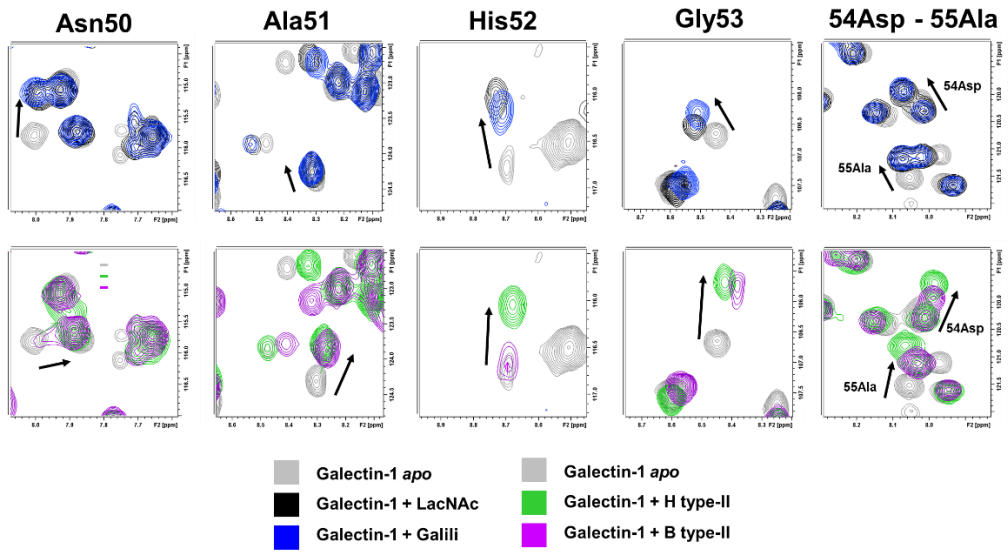


Figure 3.7. Expansions of ^{1}H-^{15}N HSQC spectra of apo galectin-1 (grey) upon addition of 10 equivalents of LacNAc (black),10 equivalents of Galili (blue), 15 equivalents of H type-II (green) and 15 equivalents of B type-II (purple).

3.4.3 Long range ^{1}H-^{15}N HSQC experiments: the behaviour of His residues

Additional structural rationale regarding the role of the Fuc moiety in the interaction was obtained through long-range ($^{2}J_{NH}$) ^{1}H-^{15}N HSQC experiments. In particular, the analysis of the signals corresponding to the His side chains was carried out to decipher the involvement of these residues in the binding to the fucosylated and non-fucosylated ligands.

Galectin-1 shows two His: **His44** and **His52**. His44, located in the S4 strand, is a strictly conserved residue among galectins and consistently involved in the interaction with β-Gal: acts as a key HB acceptor from β-Gal 4-OH. Contrarily, His52 is located in the L4 loop, an unique element for galectin-1 (Figure 3.8).

Following the long range ^{1}H-^{15}N correlations of the imidazole ring of His residues, it is possible to decipher the major tautomer (Nδ1-H or Nε2-H), since when the nitrogen of the ring is protonated, its chemical shift is ca. δ 170 ppm, while the non-protonated form appears at ca. δ 250 ppm (Figure 3.9, panel A).

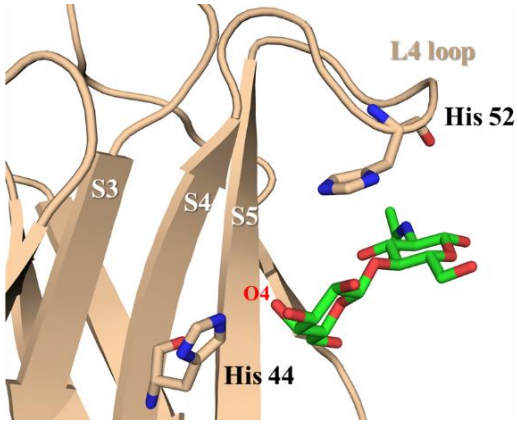


Figure 3.8. 3D model of galectin-1 bound to LacNAc pointing out the position of His44 and His52, displayed as sticks.

Moreover, the NMR signals for the two ring nitrogens (Nδ1 and Nε2) tend to get closer in the spectra when a tautomer equilibrium is taking place. [39,40]

Thus, a set of long-range ^{1}H-^{15}N HSQC experiments were recorded for galectin-1 alone and in the presence of 10 equivalents of **LacNAc**and the **H type-II** antigen.

In the ***apo* form** of the lectin, only the NH cross-peaks for one imidazole ring were observed (Figure 3.9, panel B). The presence of Nε2-Hδ2 signal at δ 187.5 ppm was consistent with a larger population of the Nε2-H tautomer, whereas the ^{15}N chemical shift deviation from the expected values for the pure Nε2-H tautomer also indicated the contribution of the Nδ1-H tautomer. The addition of 10 equivalents of **LacNAc** did not provide substantial changes on these resonances but, in contrast, a new set of NH cross-peaks appeared with a significant differentiation between $\delta_{N\varepsilon 2}$ and $\delta_{N\delta 1}$ and with a pattern characteristic for the Nδ1-H tautomer (Figure 3.9, panel C). This set of peaks, whose chemical shift indicated that the Nε2-H protonated form was marginal, were assigned to His44, given the key role of this residue for galactose binding through HB to Gal OH4, which should involve the Nδ1-H tautomer. Consequently, the other set of peaks was assigned to His52.

Upon addition of the **H type-II** antigen**,** His44 signals were clearly detected, displaying the same chemical shift as for LacNAc (Figure 3.9, panel D). Contrarily, the signals for His52 became broader, even displaying multiple peaks, strongly suggesting the presence of multiple states in slow-medium exchange regime in the chemical shift time scale. Moreover, $\delta_{N\varepsilon 2}$ and $\delta_{N\delta 1}$ turned out to be much more similar. Thus, the presence

of the Fuc residue likely modifies the dynamics around His52, while its chemical equilibrium is kept. This evidence can be reconciled with the fact that His52 and the L4 loop cannot provide the proper accommodation of the Fuc moiety, as also supported by the ITC and MD results.

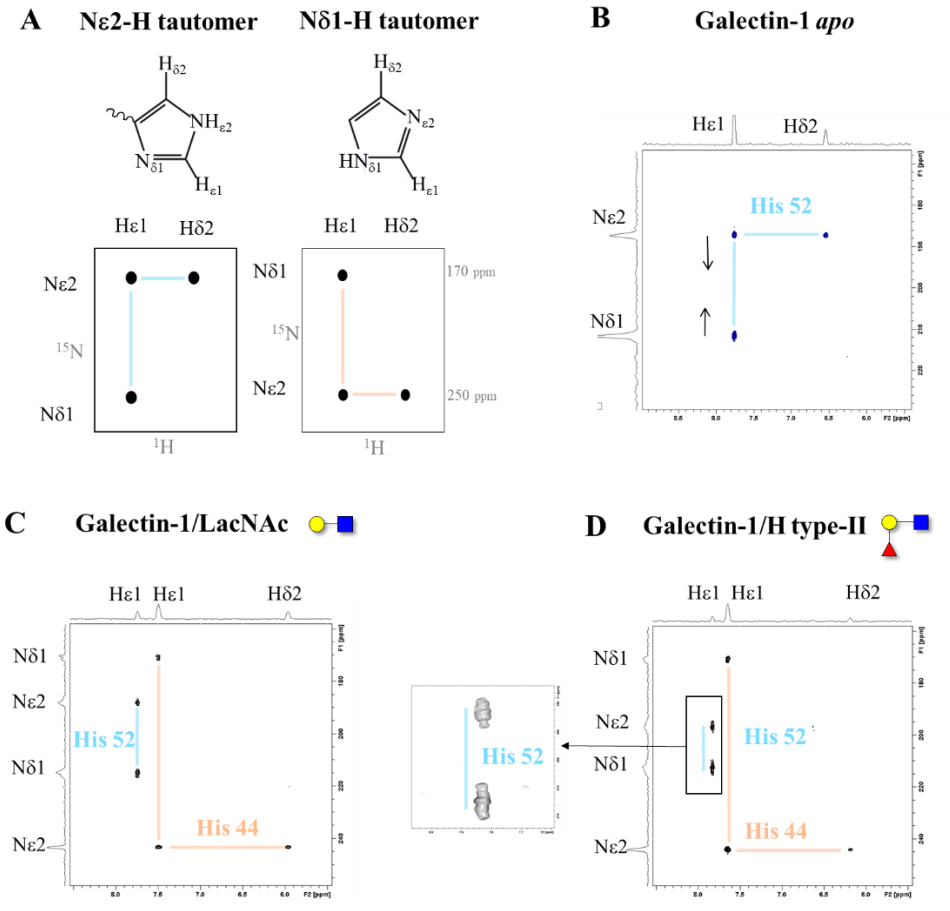


Figure 3.9. A) Representation of the tautomeric forms of imidazole ring of histidine, together with a schematic view of the expected cross-peak pattern in the ^{1}H-^{15}N long range HSQC spectrum for each tautomer. ^{1}H-^{15}N long range HSQC spectrum of galectin-1 alone (B), with 10 equivalents of ligand LacNAc (C) and with 10 equivalents of H type-II

3.4.4 CLEANEX experiments: protein dynamics upon binding

Previous studies have proposed the presence of fast motions (in the ps timescale) of galectin-1 upon binding to lactose through standard relaxation-based R1 and R2 NMR experiments. [29] However, the CSPs detected far beyond the binding site strongly suggest the presence of conformational fluctuations in a much slower timescale. [30,41]

Phase-modulated CLEAN chemical exchange NMR (CLEANEX-PM) is one of the key experiments for studying these features, since allows the detection of NH residues presenting exchange rates with water in a time range of 5-500 ms. [42,43] Thus, these experiments were recorded for galectin-1 alone and in the presence of **LacNAc** and **B type-II**. High protein:ligand ratios were used to ensure the complete saturation of the lectin, thus assessing that the measured NH-H_2O exchange rates correspond to the bound forms.

The measured **global average exchange** rate for the CLEANEX experiment with galectin-1 alone was k_{ex}=23 s^{-1}, whereas in the presence of LacNAc the value decreased at 10 s^{-1}, pointing put a global reduction of the exchange rate with the highest affinity ligand with respect to the *apo* form. In contrast, the global k_{ex} of the spectra recorded with the tetrasaccharide was 20 s^{-1}, suggesting that the binding to weaker ligands produce minor global changes.

The CLEANEX experiment for **galectin-1 *apo*** showed 13 amide NH cross-peaks (out of the 135 total ones) at different mixing times. These residues belong to amino acids located at the dimer interface and in the L3, L4, L6, L7, and L9 loops, which correspond to solvent exposed regions of the protein (Figure 3.10). Interestingly, also some residues directly involved in the binding as well as amino acids located far away from the recognition site were detected, rendering them suitable probes to monitor changes on the protein structure (Figure 3.10).

Remarkably, the residues of the **L4 loop** were differently affected by **LacNAc** and **B-type-II**. Therefore, it is tempting to speculate that different dynamic behaviour are taking place for this loop in the presence of non-fucosylated or fucosylated ligands, as also shown by the long-range ^{1}H-^{15}N HSQC experiments. The different conformation adopted by His52 when the Fuc-containing ligand is accommodated in the binding site (Figure 3.9) can explain the lower protection in average from water.

The exchange rates of several residues located **far away from the binding site** (Ala1-Cys3 at the dimer interface, Ser38 at the L3 loop, Ala94 at the L7 loop, and Asn113

and Glu115 at the F9 loop) were significantly reduced upon addition of LacNAc (Figure 3.10), while the presence of B-type II produced less pronounced effects.

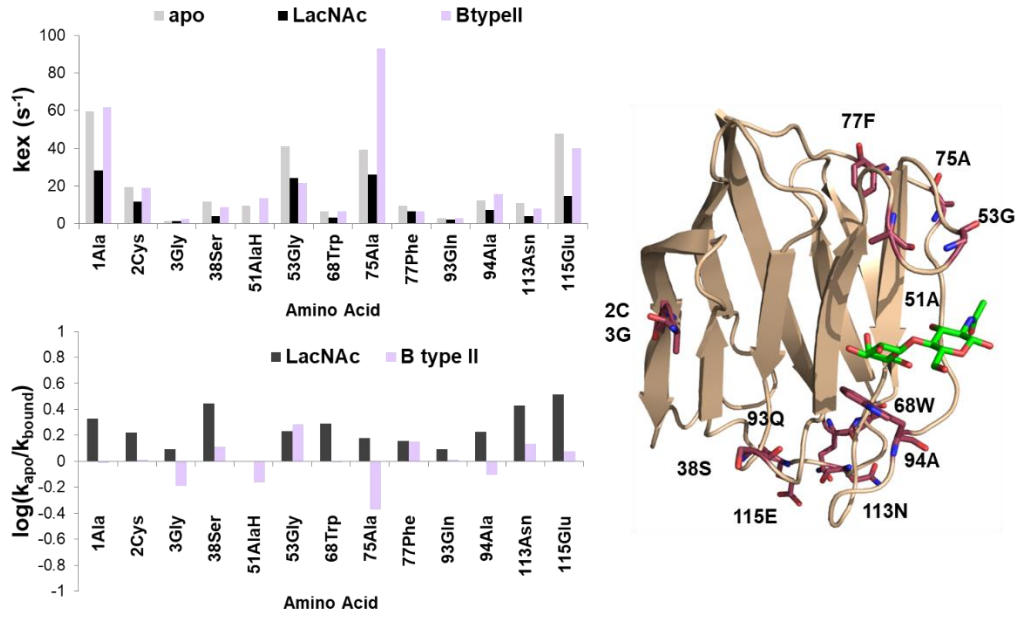


Figure 3.10. *Left*: NH-water exchange rates, k_{ex} (s^{-1}), obtained from CLEANEX-PM experiments for galectin-1 in the *apo* and bound forms upon addition of **LacNAc** and **B-typeII**. *Right*: X-ray structure of galectin-1 associated to LacNAc (PDB ID: 1W6P). Amino acids with detected exchangeable NH amide protons are highlighted.

All together, these evidences confirm that the entire structure of the lectin is perturbed upon ligand binding, especially in the presence of the stronger binder.

3.4.5 Relaxation Dispersion analysis: conformational fluctuations of the protein

The fast librations in the ps-ns timescale previously described for galecin-1 and lactose are likely capturing thermal motions more than the recognition dynamics demonstrated herein. [29] In order to fully decipher these dynamic features of galectin-1, ^{15}N CPMG relaxation dispersion NMR experiments were employed. [44] With this technique, protein residues (backbone amides) with dynamics in the μs-ms timescale can be detected. This time scale is elusive to the standard relaxation experiments, which only capture dynamic events in faster timescales, in the ns-ps range, far from those required to produce large amplitude motions related to allostery.

Relaxation dispersion experiments were then performed for galectin-1 alone and in the presence of **LacNAc** and the **B type-II**; moreover, as a control, also the galectin-3 CRD monomer was investigated in its *apo* and LacNAc-bound form.
After the data analysis, a total of 34 residues displaying µs-ms dynamics upon **LacNAc** binding to galectin-1 were found. Furthermore, a good number of those residues displayed a high degree of consistency in both calculated parameters (k_{ex} and p_B values) and, therefore, a collective fitting was performed. A final cluster of 13 residues showed a concerted dynamics at a k_{ex} of 380 s^{-1} with an excited state showing populations (p_B) of ca. 1.5%. The 13 residues (Leu4, Ser7, Leu9, Arg18, Asp54, Ala55, Val76, Asp92, Ala121, Ala122, Asp123, Phe126 and Phe133) are clustered at the interface between the two monomers of the protein (Figure 3.13, A). This evidence strongly suggest that LacNAc binding to galectin-1 triggers an allosteric response in the dimerization interface, increasing the local dynamics.

The same experiment performed with the other complexes (**galectin-1/B type-II, galectin-3,** and **galectin-3/LacNAc**) only showed a limited number of residues (between 6 and 13) with remarkable fluctuations in the micro-to-milli second time scale. Moreover, the attempts to statistically cluster those residues into collective motions failed in all cases. Then, it is likely that these residues simply show residual thermal motion.

The positive entropy found in this system is related to the presence of motions, which are likely associated to functional dynamics. Moreover, the consistent dynamics in the µs timescale, exclusively associated to LacNAc, can be considered as a direct hint of the presence of allosteric transmission induced upon ligand binding.

3.5 MD simulations: allosteric communication analysis

The presence of allosteric communication between the two domains of galectin-1, triggered upon LacNAc binding, was also analysed through µs molecular dynamics simulations (µs-MD) to further assist the experimental NMR findings. Both the *apo* protein and the bound form in presence of **LacNAc** and **B type-II** were investigated.

As a control, also the monomer, galectin-3 CRD, was studied with the same experimental approach.

From the computed MD data, optimal and suboptimal pathways for dynamic correlation between residues of the binding site and any other amino acid in the protein were traced using the Weighted Implementation of Suboptimal Paths (WISP) algorithm (see the *Chem Eur J* publication for additional information). [1] The length of the pathways were mapped onto the lectin structure to obtain a visual representation of the internal **correlated motions** occurring during the MD run (Figure 3.11).

For galectin-1, the motions propagated from selected amino acids of the binding site all along the monomer and the residues with the highest frequency in the pathways were distributed along the β-strands. In contrast, for galectin-3 such correlated motions were only located nearby the binding site (Figure 3.11).

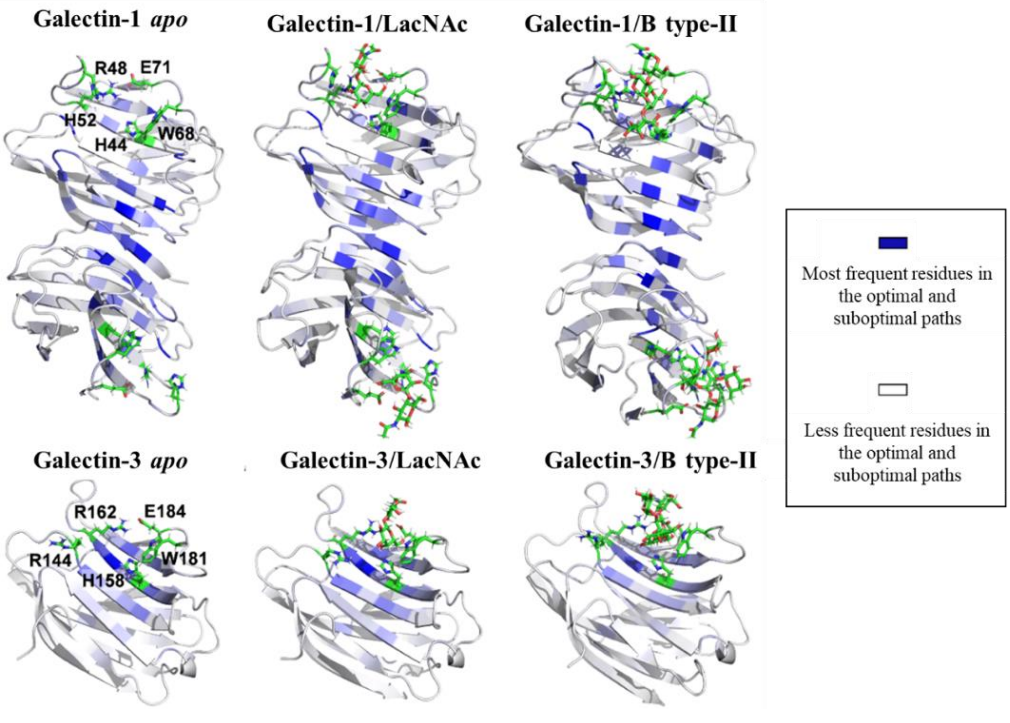


Figure 3.11. Dynamic correlation network representations calculated from μs-MD simulations. For each complex, the residues coloured in blue are those appearing most frequently in all the calculated optimal and suboptimal allosteric pathways based on correlated motions between selected pairs of residues (shown as green sticks) and the rest of the protein.

The optimal and suboptimal allosteric pathways can be represented to easily identify the location of the dynamic information inside the lectin (Figure 3.12). From this representation is even more evident how, for galectin-1, the innervation of clustered residues propagates throughout the whole monomer. In contrast, the CRD of galectin-3 only displays a pathway of amino acids affected by the motions within or close to the canonical binding site.

Fittingly, for galectin-1/**LacNAc**, most of the residues identified as relevantly involved in **allosteric communication** by this MD-based analysis are the same that emerged in the relaxation dispersion experiments as those suffering concertated dynamics (Figure 3.13). This concordance reinforces the existence of long-range dynamic correlations between the binding site and distant amino acids throughout the β-sheets upon LacNAc binding. Overall, the demonstrated perturbations strongly suggest the presence of allosteric communication of the lectin. The presence of allosteric communication is also connected to the negative cooperativity found in the ITC analysis: once the first glycan molecule is bound and the allosteric information is transmitted, the perturbation makes that the second ligand is bound with smaller affinity.

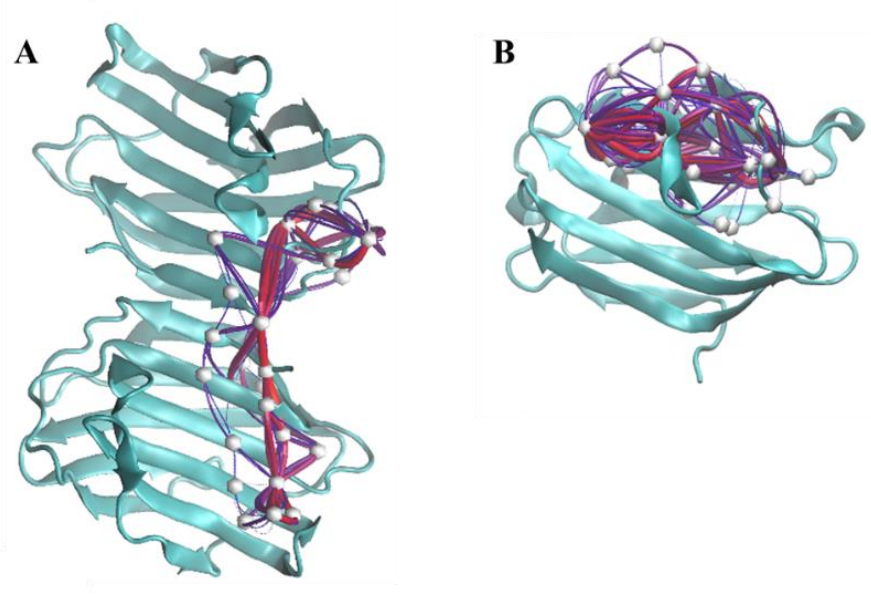


Figure 3.12. Dynamic correlation network representations calculated from the μs-MD simulations. Example of 100 optimal and suboptimal allosteric pathways based on the correlated motions between selected pairs of residues calculated for *apo* galectin-1 (A) and galectin-3 (B).

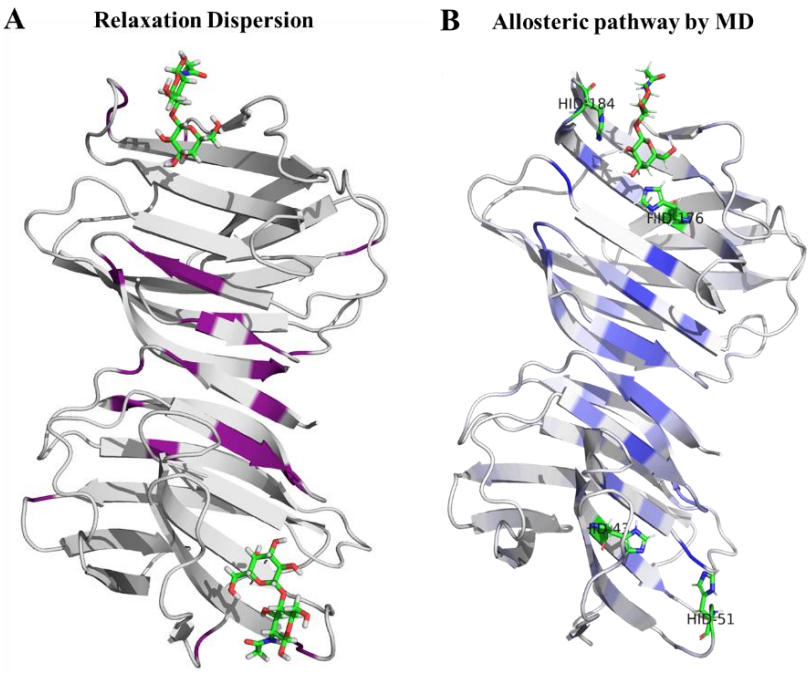


Figure 3.13. Long-range concerted dynamics in galectin-1. B) Residues most frequently involved in the optimal and suboptimal allosteric pathways found from the μs-MD analysis (color-coded in a blue gradient). The pathway is calculated from selected residues (shown as green sticks) in both binding sites. A) Residues (in purple) showing concerted dynamics at 380 s^{-1} as determined by transversal relaxation dispersion (RD) NMR experiments.

3.6 Conclusions

The interaction of human galectin-1 with the histo blood group antigens and their fragments has been deeply analysed and characterized with a multidisciplinary panel of experimental and computational methodologies.

In a first step, the ITC data highlight how the binding, for all ligands under study, is **favoured** by **entropy**, in strong contrast with the observations reported for galectin-3 and for most lectin-sugar interaction events. [2,33] **LacNAc** is the **best** ligand for galectin-1 among those tested. The saccharides of increased complexity (Galili, H-type-II and B-typeII) displayed worst affinities. From the ITC data fitting, sequential binding with **negative cooperativity** was proposed for ligands **1-4**.

The STD-NMR results unravelled the key **binding epitope**: **β-Gal** generates the main contacts, but additional contributions from the α-Gal and α-Fuc moieties also emerged.

Nevertheless, the generated MD-based 3D complexes did not predict any stable contact of galectin-1 with those sugar residues.

The CSP NMR analysis agreed with the proposed 3D complexes showing that the extra residues (α-Fuc and α-Gal) in **2-4** did not establish stabilizing contacts with galectin-1, as also deduced by ITC.

Moreover, the CSP results also revealed **perturbations** for **residues** located **far beyond the binding site** (along the β-sheets and at the dimer interface). Curiously, the entity of those perturbations correlated with ligand affinity, being more pronounced for the best affinity ligand (LacNAc). Water-exchange CLEANEX-PM experiments confirmed these data, pointing out that residues all over the **β-sheets** and **at the dimer interface** of the protein suffered **changes** upon binding in their **presentation to water**.

^{15}N CPMG relaxation dispersion analysis revealed dynamic motions propagating throughout the lectin in the micro-to-millisecond timescale. In particular, NMR experimental evidences demonstrated the presence of **concerted motions** for 13 residues upon LacNAc binding, while no concerted clustered residues were detected in the other cases (galectin-1 with lower affinity binders or galectin-3 with LacNAc).

Molecular dynamics simulations also predicted the existence of an **allosteric correlation** between the residues of the binding site and distant amino acids along the β-sheets upon LacNAc binding. In fact, once the first glycan molecule is bound, the communication is transmitted until reaching the dimer interface, presumably justifying why the second binding event displayed smaller affinity.

The final aim of protein-ligand binding studies is revealing the structural details of the interaction and relating them with the biological function of a given system. However, this is not always an easy task because many processes are usually responsible for triggering the observed final response. In fact, the precise biological roles of human galectins remain elusive.

The results presented herein show how sugar recognition by galectins is an extremely complex process that depends on many factors.

Despite the similarities, a fine-tuned recognition preference has been detected by comparing the chimera-type galectin-3 with the prototype galectin-1. For instance, it has been reported that galectin-1 displays noticeable preference to bind terminal

LacNAc structures in complex N-glycans, whereas galectin-3 preferentially recognizes internal LacNAc moieties. [25,26] Those different binding specificities have also been confirmed in the natural cellular environment. [45] Our results indicate that the intrinsic conformational flexibility strongly differs between galectin-1 and galectin-3 and allow hypothesizing that this feature could be behind the observed difference in specificity. It is tempting to speculate that, since terminal LacNAc structures on glycoconjugates are solvent exposed and accessible, the binding of a lectin displaying a pronounced intrinsic conformational flexibility may reduce the entropic cost of the association. In contrast, the binding to a globular structure would be in disadvantage.

The exclusive allostery detected for galectin-1 definitely provides insights into a functional role of the dimeric architecture of the lectin. The monomeric CRD of galectin-3 did not display this behaviour, highlighting how the motion features are drastically divergent in the two galectins.

Finally, our findings shed light on structural and thermodynamic binding features of the recognition event that could be further exploited as clues for the rational design of compounds capable of selectively binding galectin-1.

3.7 References

1. Bertuzzi, S.; Gimeno, A.; Núñez-Franco, R.; Bernardo-Seisdedos, G.; Delgado, S.; Jiménez-Osés, G.; Millet, O.; Jiménez-Barbero, J.; Ardá, A. Unravelling the Time Scale of Conformational Plasticity and Allostery in Glycan Recognition by Human Galectin-1. *Chem. - A Eur. J.* **2020**, *26*, 15643–15653, doi:10.1002/CHEM.202003212.
2. Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M.A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F.J.; et al. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie Int. Ed.* **2019**, *58*, 7268–7272, doi:10.1002/anie.201900723.
3. Quintana, J.I.; Delgado, S.; Núñez-Franco, R.; Cañada, F.J.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Ardá, A. Galectin-4 N-Terminal Domain: Binding Preferences Toward A and B Antigens With Different Peripheral Core Presentations. *Front. Chem.* **2021**, *9*, 193, doi:10.3389/FCHEM.2021.664097/BIBTEX.
4. Gómez-Redondo, M.; Delgado, S.; Núñez-Franco, R.; Jiménez-Osés, G.; Ardá, A.; Jiménez-Barbero, J.; Gimeno, A. The Two Domains of Human Galectin-8 Bind Sialyl- and Fucose-Containing Oligosaccharides in an Independent Manner. A 3D View by Using NMR. *RSC Chem. Biol.* **2021**, *2*, 932–941, doi:10.1039/D1CB00051A.
5. Bertuzzi, S.; Quintana, J.I.; Ardá, A.; Gimeno, A.; Jiménez-Barbero, J. Targeting Galectins With Glycomimetics. *Front. Chem.* 2020, *8*, 593.
6. Camby, I.; Le Mercier, M.; Lefranc, F.; Kiss, R. Galectin-1: A Small Protein with Major Functions. *Glycobiology* **2006**, *16*, 137R-157R, doi:10.1093/GLYCOB/CWL025.
7. Gauthier, L.; Rossi, B.; Roux, F.; Termine, E.; Schiff, C. Galectin-1 Is a Stromal Cell Ligand of the Pre-B Cell Receptor (BCR) Implicated in Synapse Formation between Pre-B and Stromal Cells and in Pre-BCR Triggering. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 13014–13019, doi:10.1073/PNAS.202323999/SUPPL_FILE/3239FIG7.PDF.
8. Mourcin, F.; Breton, C.; Tellier, J.; Narang, P.; Chasson, L.; Jorquera, A.; Coles, M.; Schiff, C.; Mancini, S.J.C. Galectin-1–Expressing Stromal Cells Constitute a Specific Niche for Pre-BII Cell Development in Mouse Bone Marrow. *Blood* **2011**, *117*, 6552–6561, doi:10.1182/BLOOD-2010-12-323113.
9. Cedeno-Laurent, F.; Dimitroff, C.J. Galectin-1 Research in T Cell Immunity: Past, Present and Future. *Clin. Immunol.* **2012**, *142*, 107–116, doi:10.1016/J.CLIM.2011.09.011.
10. 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, Vol. 17, Page 1566* **2016**, *17*, 1566, doi:10.3390/IJMS17091566.
11. Nambiar, D.K.; Aguilera, T.; Cao, H.; Kwok, S.; Kong, C.; Bloomstein, J.; Wang, Z.; Rangan, V.S.; Jiang, D.; Von Eyben, R.; et al. Galectin-1–Driven T Cell Exclusion in the Tumor Endothelium Promotes Immunotherapy Resistance. *J. Clin. Invest.* **2019**, *129*, 5553–5567, doi:10.1172/JCI129025.
12. Benci, J.L.; Xu, B.; Qiu, Y.; Wu, T.J.; Dada, H.; Twyman-Saint Victor, C.; Cucolo, L.; Lee, D.S.M.; Pauken, K.E.; Huang, A.C.; et al. Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. *Cell* **2016**, *167*, 1540-1554.e12, doi:10.1016/J.CELL.2016.11.022.
13. Jeschke, U.; Karsten, U.; Wiest, I.; Schulze, S.; Kuhn, C.; Friese, K.; Walzel, H. Binding of Galectin-1 (Gal-1) to the Thomsen-Friedenreich (TF) Antigen on Trophoblast Cells and Inhibition of Proliferation of Trophoblast Tumor Cells in Vitro by Gal-1 or an Anti-TF Antibody. *Histochem. Cell Biol.* **2006**, *126*, 437–444, doi:10.1007/S00418-006-0178-1.
14. Hsieh, S.H.; Ying, N.W.; Wu, M.H.; Chiang, W.F.; Hsu, C.L.; Wong, T.Y.; Jin, Y.T.;

Hong, T.M.; Chen, Y.L. Galectin-1, a Novel Ligand of Neuropilin-1, Activates VEGFR-2 Signaling and Modulates the Migration of Vascular Endothelial Cells. *Oncogene* **2008**, *27*, 3746–3753, doi:10.1038/SJ.ONC.1211029.

15. Croci, D.O.; Cerliani, J.P.; Dalotto-Moreno, T.; Méndez-Huergo, S.P.; Mascanfroni, I.D.; Dergan-Dylon, S.; Toscano, M.A.; Caramelo, J.J.; García-Vallejo, J.J.; Ouyang, J.; et al. Glycosylation-Dependent Lectin-Receptor Interactions Preserve Angiogenesis in Anti-VEGF Refractory Tumors. *Cell* **2014**, *156*, 744–758, doi:10.1016/J.CELL.2014.01.043.
16. 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*, 307–319, doi:10.1016/J.CTRV.2013.07.007.
17. Blanchard, H.; Bum-Erdene, K.; Bohari, M.H.; Yu, X. Galectin-1 Inhibitors and Their Potential Therapeutic Applications: A Patent Review. *http://dx.doi.org/10.1517/13543776.2016.1163338* **2016**, *26*, 537–554, doi:10.1517/13543776.2016.1163338.
18. Ramírez Hernández, E.; Sánchez-Maldonado, C.; Mayoral Chávez, M.A.; Hernández-Zimbrón, L.F.; Patricio Martínez, A.; Zenteno, E.; Limón Pérez de León, I.D. The Therapeutic Potential of Galectin-1 and Galectin-3 in the Treatment of Neurodegenerative Diseases. *https://doi.org/10.1080/14737175.2020.1750955* **2020**, *20*, 439–448, doi:10.1080/14737175.2020.1750955.
19. Cho, M.; Cummings, R.D. Galectin-1, a Beta-Galactoside-Binding Lectin in Chinese Hamster Ovary Cells. I. Physical and Chemical Characterization. *J. Biol. Chem.* **1995**, *270*, 5198–5206, doi:10.1074/JBC.270.10.5198.
20. Stowell, S.R.; Cho, M.; Feasley, C.L.; Arthur, C.M.; Song, X.; Colucci, J.K.; Karmakar, S.; Mehta, P.; Dias-Baruffi, M.; McEver, R.P.; et al. Ligand Reduces Galectin-1 Sensitivity to Oxidative Inactivation by Enhancing Dimer Formation. *J. Biol. Chem.* **2009**, *284*, 4989–4999, doi:10.1074/JBC.M808925200.
21. Dias-Baruffi, M.; Zhu, H.; Cho, M.; Karmakar, S.; McEver, R.P.; Cummings, R.D. Dimeric Galectin-1 Induces Surface Exposure of Phosphatidylserine and Phagocytic Recognition of Leukocytes without Inducing Apoptosis. *J. Biol. Chem.* **2003**, *278*, 41282–41293, doi:10.1074/JBC.M306624200.
22. Miura, T.; Takahashi, M.; Horie, H.; Kurushima, H.; Tsuchimoto, D.; Sakumi, K.; Nakabeppu, Y. Galectin-1β, a Natural Monomeric Form of Galectin-1 Lacking Its Six Amino-Terminal Residues Promotes Axonal Regeneration but Not Cell Death. *Cell Death Differ. 2004 1110* **2004**, *11*, 1076–1083, doi:10.1038/sj.cdd.4401462.
23. Earl, L.A.; Bi, S.; Baum, L.G. Galectin Multimerization and Lattice Formation Are Regulated by Linker Region Structure. *Glycobiology* **2011**, *21*, 6–12, doi:10.1093/GLYCOB/CWQ144.
24. Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W.E.G.; et al. Oligosaccharide Specificity of Galectins: A Search by Frontal Affinity Chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572*, 232–254, doi:10.1016/S0304-4165(02)00311-2.
25. Kamili, N.A.; Arthur, C.M.; Gerner-Smidt, C.; Tafesse, E.; Blenda, A.; Dias-Baruffi, M.; Stowell, S.R. Key Regulators of Galectin-Glycan Interactions. *Proteomics* **2016**, *16*, 3111–3125, doi:10.1002/PMIC.201600116.
26. Moure, M.J.; Gimeno, A.; Delgado, S.; Diercks, T.; Boons, G.J.; Jiménez-Barbero, J.; Ardá, A. Selective 13C-Labels on Repeating Glycan Oligomers to Reveal Protein Binding Epitopes through NMR: Polylactosamine Binding to Galectins. *Angew. Chemie Int. Ed.* **2021**, *60*, 18777–18782, doi:10.1002/ANIE.202106056.
27. López-Lucendo, M.F.; Solís, D.; André, S.; Hirabayashi, J.; Kasai, K.I.; 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*, 957–970, doi:10.1016/J.JMB.2004.08.078.

28. Asensio, J.L.; Ardá, A.; Cañada, F.J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946–954, doi:10.1021/AR300024D/ASSET/IMAGES/MEDIUM/AR-2012-00024D_0006.GIF.
29. Nesmelova, I. V; Ermakova, E.; Daragan, V.A.; Pang, M.; Menéndez, M.; Lagartera, L.; Solís, D.; Baum, L.G.; Mayo, K.H. Lactose Binding to Galectin-1 Modulates Structural Dynamics, Increases Conformational Entropy, and Occurs with Apparent Negative Cooperativity. *J. Mol. Biol* **2010**, *397*, 1209–1230, doi:10.1016/j.jmb.2010.02.033.
30. Chien, C.T.H.; Ho, M.R.; Lin, C.H.; Hsu, S.T.D. Lactose Binding Induces Opposing Dynamics Changes in Human Galectins Revealed by NMR-Based Hydrogen–Deuterium Exchange. *Mol. 2017, Vol. 22, Page 1357* **2017**, *22*, 1357, doi:10.3390/MOLECULES22081357.
31. Bains, G.; Freire, E. Calorimetric Determination of Cooperative Interactions in High Affinity Binding Processes. *Anal. Biochem.* **1991**, *192*, 203–206, doi:10.1016/0003-2697(91)90207-A.
32. Shimura, K.; Arata, Y.; Uchiyama, N.; Hirabayashi, J.; Kasai, K.-I. Determination of the Affinity Constants of Recombinant Human Galectin-1 and-3 for Simple Saccharides by Capillary Affinophoresis. *J. Chromatogr. B* **2002**, 199–210.
33. Fox, J.M.; Zhao, M.; Fink, M.J.; Kang, K.; Whitesides, G.M. The Molecular Origin of Enthalpy/Entropy Compensation in Biomolecular Recognition. *https://doi.org/10.1146/annurev-biophys-070816-033743* **2018**, *47*, 223–250, doi:10.1146/ANNUREV-BIOPHYS-070816-033743.
34. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788, doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.
35. Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, doi:10.1021/ja0100120.
36. Viegas, A.; Manso, J.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88*, 990–994, doi:10.1021/ed101169t.
37. Ardá, A.; Jiménez-Barbero, J. The Recognition of Glycans by Protein Receptors. Insights from NMR Spectroscopy. *Chem. Commun.* **2018**, *54*, 4761–4769, doi:10.1039/C8CC01444B.
38. Bonzi, J.; Bornet, O.; Betzi, S.; Kasper, B.T.; Mahal, L.K.; Mancini, S.J.; Schiff, C.; Sebban-Kreuzer, C.; Guerlesquin, F.; Elantak, L. Pre-B Cell Receptor Binding to Galectin-1 Modifies Galectin-1/Carbohydrate Affinity to Modulate Specific Galectin-1/Glycan Lattice Interactions. *Nat. Commun. 2015 61* **2015**, *6*, 1–12, doi:10.1038/ncomms7194.
39. Van Dijk, A.A.; Scheek, R.M.; Dijkstra, K.; Wolters, G.K.; Robillard, G.T. Characterization of the Protonation and Hydrogen Bonding State of the Histidine Residues in IIAmtl, a Domain of the Phosphoenolpyruvate-Dependent Mannitol-Specific Transport Protein. *Biochemistry* **1992**, *31*, 9063–9072, doi:10.1021/BI00152A050/ASSET/BI00152A050.FP.PNG_V03.
40. Pelton, J.G.; Torchia, D.A.; Meadow, N.D.; Roseman, S. Tautomeric States of the Active-Site Histidines of Phosphorylated and Unphosphorylated IIIGlc, a Signal-Transducing Protein from Escherichia Coli, Using Two-Dimensional Heteronuclear NMR Techniques. *Protein Sci.* **1993**, *2*, 543–558, doi:10.1002/PRO.5560020406.
41. Chao, F.A.; Byrd, R.A. Protein Dynamics Revealed by NMR Relaxation Methods. *Emerg. Top. Life Sci.* **2020**, *2*, 93–105, doi:10.1042/ETLS20170139.
42. Hwang, T.L.; Van Zijl, P.C.M.; Mori, S. Accurate Quantitation of Water-Amide Proton

Exchange Rates Using the Phase-Modulated CLEAN Chemical EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *J. Biomol. NMR 1998 112* **1998**, *11*, 221–226, doi:10.1023/A:1008276004875.

43. Hwang, T.L.; Mori, S.; Shaka, A.J.; Van Zijl, P.C.M. Application of Phase-Modulated CLEAN Chemical EXchange Spectroscopy (CLEANEX-PM) to Detect Water - Protein Proton Exchange and Intermolecular NOEs. *J. Am. Chem. Soc.* **1997**, *119*, 6203–6204, doi:10.1021/JA970160J/ASSET/JA970160J.FP.PNG_V03.
44. Singh, A.; Purslow, J.A.; Venditti, V. 15N CPMG Relaxation Dispersion for the Investigation of Protein Conformational Dynamics on the Ms-ms Timescale. *J. Vis. Exp.* **2021**, *2021*, doi:10.3791/62395.
45. Nielsen, M.I.; Stegmayr, J.; Grant, O.C.; Yang, Z.; Nilsson, U.J.; Boos, I.; Carlsson, M.C.; Woods, R.J.; Unverzagt, C.; Leffler, H.; et al. Galectin Binding to Cells and Glycoproteins with Genetically Modified Glycosylation Reveals Galectin-Glycan Specificities in a Natural Context. *J. Biol. Chem.* **2018**, *293*, 20249–20262, doi:10.1074/JBC.RA118.004636.

CHAPTER 4

THE INTERACTION OF GALECTIN-1 AND GALECTIN-3 WITH LACNAC-DECORATED GLYCOPOLYMERS

The contribution of cross-linking effects

4.1 Introduction

In this chapter, the recognition features of multi-LacNAc containing *N*-(2-hydroxypropyl) methacrylamide (HMPA) copolymers by galectin-1 and by the CRD of galectin-3 have been disentangled. For comparison purposes, the LacNAc disaccharide and individual LacNAc-containing components of the polymer were also analysed.

Compounds **2-6** (Figure 4.2) were synthetized by our collaborators from the Institute of Microbiology (Czech Academy of Science, Prague) and their synthesis was previously published. [1] A multidisciplinary approach comprising NMR methodologies, DLS, and cryo-EM measurements has been adopted and the results of this project were successfully published in 2021. [2]

4.1.1 General context

Galectin-1 and galectin-3 are the most widely studied members of the galectin family of lectins. They exhibit relevant biological roles, and, in particular, their implications in cancer proliferation as well as in inflammation processes have been documented. [3–9] Consequently, they have become attractive targets for therapeutic purposes. [10–14]

The sequence identity between the CRD of these two galectins is moderately high and the key residues of the binding site that orchestrate the main contacts with β-galactosides are conserved (Figure 4.1). [15] However, their **dissimilar architecture** strongly determines the reported differences in their binding specificity. [16–20] Certainly, galectin-1 is a homodimer formed by two identical domains, whereas galectin-3, the only chimera-type galectin, is capable of oligomerizing through its non-lectin N-terminal domain. Although the oligomerization of galectin-3 has generally been accepted, the molecular mechanism behind this process and its detailed biological consequences remain enigmatic and far to be completely understood. [21–23] Nevertheless, the monovalent carbohydrate recognition domain (CRD) of the protein conserves the lectin activity and can be used as a good model to understand the binding specificity of galectin-3 and, more in general, of a monomeric galectin. (Figure 4.1).

Besides canonical binders such as lactose, LacNAc, and LacdiNAc, galectin-1 and galectin-3 CRD have shown to differently interact with other oligosaccharides, such as

the blood-group antigens and poly-lactosamine oligosaccharides. [18–20] However, the binding preferences are not extremely diverse and this evidence reflects the difficulty to design and synthesize a selective inhibitor acting on one of them only.

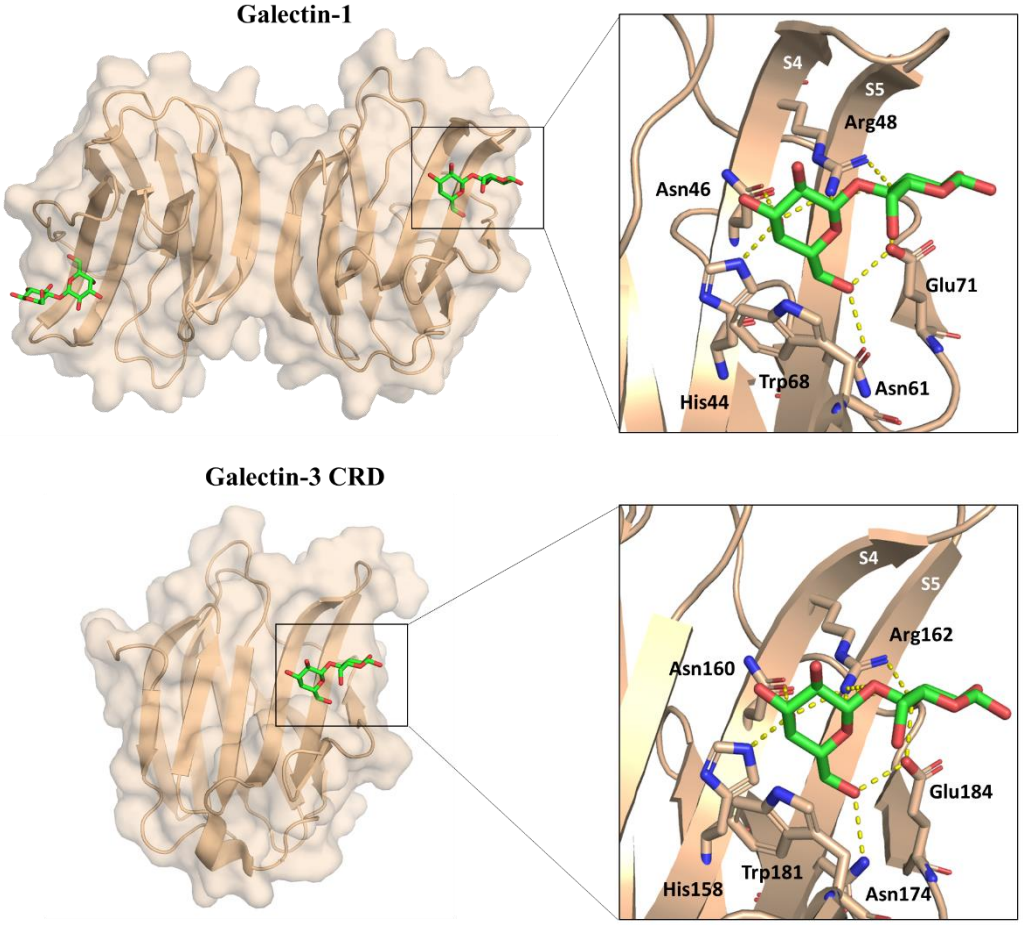


Figure 4.1. Galectin-1 homodimer in complex with lactose (PDB ID: 1GZW) and galectin-3 CRD in complex with lactose (PDB ID: 2NN8). Both binding site are zoomed in. The key residues for the interactions are highlighted.

Since single carbohydrates-lectin recognition events are characterized by low affinities (μM-mM range), the engagement of simultaneous interactions involving the receptor and/or the ligand is a fundamental point. This phenomenon, exploited by Nature to overcome the otherwise weak association, is known as multivalency and can be employed to develop inhibitors for lectins , as discussed in the Introduction Chapter. In fact, scaffolds presenting multiple sugar units have been proven to provide improved affinities for lectins and, among others, for galectin-1 and galectin-3. [15,24–29]

For this reason, different types of multivalent glycoconjugates have been developed and grouped depending on the nature of the scaffold: glycoclusters, glycodendrimers, glyconanoparticles, neoglycoproteins, and glycopolymers. [30–34].

N-(2-hydroxypropyl) methacrylamide **(HMPA) copolymers** are a well-known synthetic water-soluble carriers used for the design of glycopolymers. In fact, the HMPA scaffold displays good biocompatibility and water solubility, as well as prolonged circulation time, adequate delivery to biological targets, and lack of toxicity or immunogenicity *in vivo*. [35] Different glycopolymers with the HMPA carrier have already been proposed as binders for galectins, concluding that their architecture, molecular weight, presentation, and spacer structure strongly modulate the lectin recognition event. [1,29,36–38]

In a recent work performed by our collaborators of the Institute of Microbiology (Czech Academy of Science, Prague) a large panel of LacNAc-decorated HMPA glycopolymers and LacNAc-decorated multivalent scaffolds was synthetized varying the LacNAc density as well as its presentation (mono, bi- or trivalent). [1] This pool of molecules was tested for its ability to inhibit the binding of galectin-1 and galectin-3 to asialofetuin (ASF), concluding that galectin-1 preferred a dense presentation of individually distributed LacNAc epitopes, while galectin-3 favoured a clustered LacNAc presentation. [1]

In particular, HMPA glycopolymers carrying an individual presentation of the sugar epitope reached an avidity in the nanomolar range for galectin-1 and, remarkably, an exceptional preference for galectin-1 over galectin-3 (over 100-fold). [1]

Considering that, as discussed, HMPA glycopolymers are suitable for *in vivo* applications and given the outstanding avidity obtained for galectin-1, these molecules are promising for therapeutic applications.

In the work presented in this chapter, we selected five compounds among the reported library of multivalent ligands. Our aim was to eventually bring some structural insights on the preference for the dimeric over a monomeric galectin architecture (galectin-1 and the CRD of galectin-3, respectively) (Figure 4.2). During the discussion of this

chapter with the naming ‘galectin-3’, we refer to the CRD of galectin-3 that we have employed, unless otherwise specified.

Apart from the HMPA glycopolymers **4**-**6**, whose physicochemical characteristics are summarized in Table 4.1, also simple LacNAc (**1**) and the branched parent glycosylated ligands (**2**-**3**) were analysed. We essentially selected two small branched ligands with bivalent (**2**) and trivalent (**3**) presentation and three glycopolymers with a comparable LacNac content (19-22%) but with different LacNAc presentation: LacNAc individually distributed (**4**), LacNAc on bivalent branching (**5**) and LacNAc on trivalent branching (**6**) (Figure 4.2).

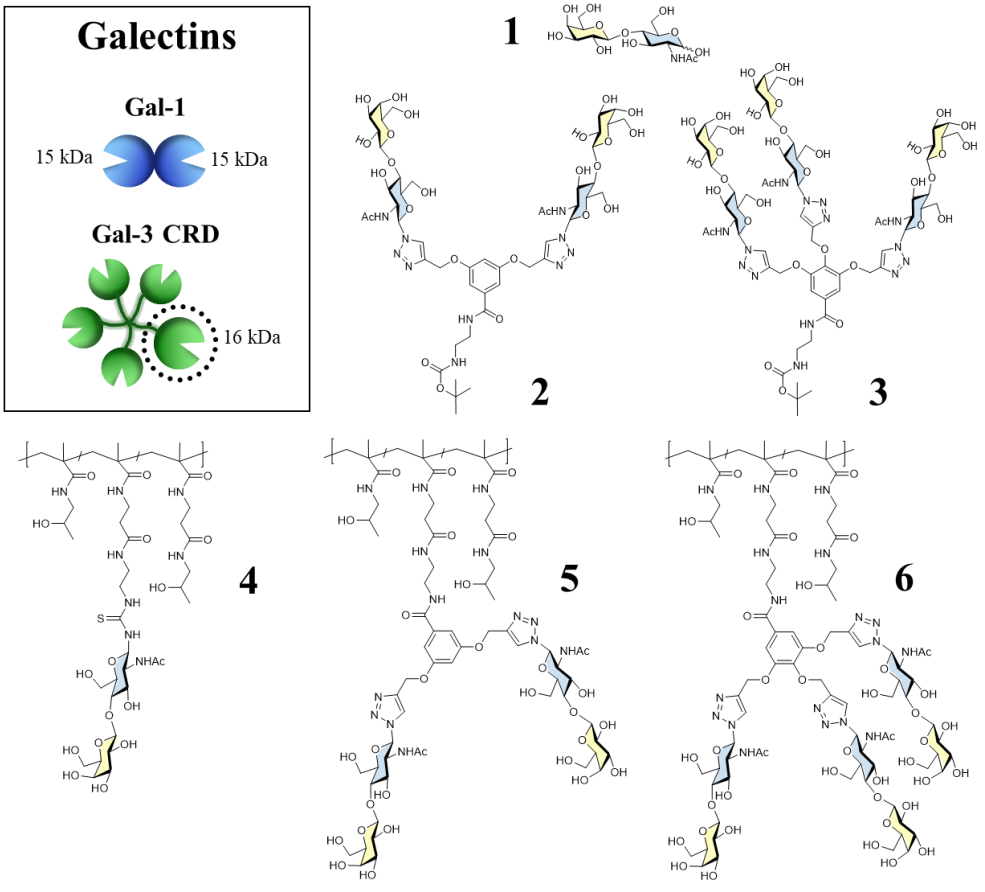


Figure 4.2. Representation of the galectins employed in the study and structures of the ligands with relative numeration (**1**-**6)** selected from the panel presented in the previous publication by Tavares et al. [1]

Ligand	(a) LacNAc presentation	(b) LacNAc content [mol. %]	(c) M_n [g mol^{-1}]	(d) M_w [g mol^{-1}]
4	individual	19.3	25,200	33,600
5	bivalent branching	20.9	30,200	37,900
6	trivalent branching	22.0	32,400	38,400

Table 4.1 Physicochemical characteristics (type of presentation, LacNAc content, and molecular weight) of the polymers **4**-**6** described by Tavares et al. [1] (a) values were calculated by integration of the NMR signals. The number-average molecular weight (M_n) (c) and weight-average molecular weight (M_w) (d) were obtained employing SEC.

The ability of the selected ligands **1**-**6** to inhibit the binding of galectin-1 and -3 to immobilized ASF glycoprotein in a competitive ELISA-assay is reported in Table 4.2. and expressed in terms of IC_{50} [1]

	Ligand **1**	Ligand **2**	Ligand **3**	Ligand **4**	Ligand **5**	Ligand **6**
Galectin-1	**78** ± 23	**19** ± 5	**9** ± 3	0.**086** ± 0.05	**0.41** ± 0.08	**0.082** ± 0.02
rp	1	4.1	8.6	906	190	851.2
rp/*n*	1	2	2.8	906	95	317
Galectin-3	**44** ± 8	**12** ± 2	**4.4** ± 1.6	**11** ± 1	**4.6** ± 1.3	**1.7** ± 0.4
rp	1	3.6	10	4	9.5	25.8
rp/*n*	1	1.8	3.3	4	4.8	8.6

Table 4.2 Inhibitory activity from ELISA assays. IC_{50} (expressed in μM) values previously obtained from competitive ELISA assay. The relative potency (rp) of a certain multivalent ligand is calculated as $IC_{50\text{-LacNAc}}$ / $IC_{50\text{–multivalent ligand}}$. rp/*n* value refers to the relative potency per active unit of LacNAc. [1]

For ligands **2**-**3**, no differences among galectins were detected, being **3** the best inhibitor for both. Comparing the data of **2** and **3** with LacNAc, it can be deduced that the bivalent and trivalent presentation of the multivalent molecules promoted cooperative effects that caused affinity enhancement.

Ligands **4**-**6** displayed an inhibitory potency in the nanomolar range. However, when the glycopolymers' IC_{50} are compared, a clear difference emerged: galectin-1 is always more sensible to the inhibition in all cases. The best binder for galectin-1, considering the relative potency per active unit, is the LacNAc-polymer featuring the individual sugar presentation (**4**), while galectin-3 marginally prefers the glycopolymer with triple epitope presentation (**6**).

4.2 NMR studies

To unravel the intricacies of the recognition phenomenon taking place between the galectins in study and ligands **1**-**6**, we resorted to NMR methodologies. Due to the intrinsic differences between ligands **1**-**3** and the polymers (**4**-**6**), *ad hoc* approaches where used for each system. In particular, we employed experiments from the viewpoint of the ligand (STD-NMR) only with the small building blocks (**1**-**3**), while we used methodologies from the viewpoint of the protein (^{1}H-^{15}N HSQC) with the polymers (**4**-**6**), as well as with **1**-**3**.

4.2.1 Analysis of the building blocks

4.2.1.1 The ligand's perspective: Saturation transfer difference NMR experiments (STD-NMR)

The preliminary structural experimental data for the interaction of galectin-1 and the CRD of galectin-3 with **1**-**3** was obtained through ^{1}H STD-NMR. [39–41] The basis of this technique is commented in the Introduction, but, briefly, from the analysis of this kind of experiment, the binding epitope of a ligand towards a receptor can be disclosed. The excess of each ligand to be added into the NMR tube containing the unlabelled lectin was optimized in each case to obtain the best signal/noise. The glycopolymers **4**-**6** were excluded from this analysis because their large molecular weight exceeds the limits of the STD experiment.

For every combination of ligand/lectin tested, a set of experiments with three different aliphatic irradiations (δ 0.00 ppm, δ -0.5 ppm, δ -1 ppm) was performed. As an example, the STD spectra together with the off resonance spectra of galectin-1 with ligands **2** and **3** are reported in Figure 4.3, panel A. All the STD experiments are gathered in the Supporting Information.

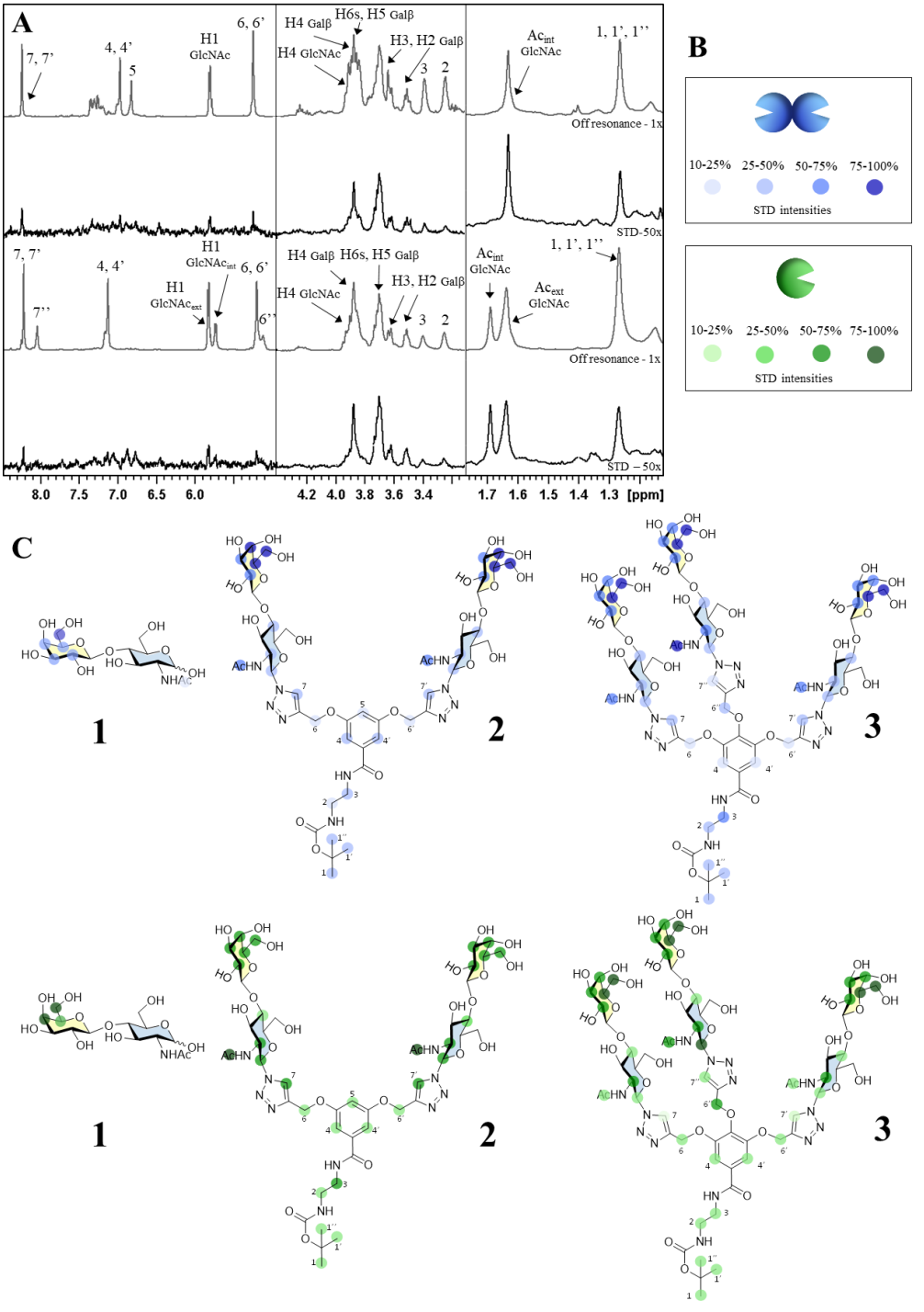


Figure 4.3. STD-NMR results. (A) Off-resonance spectra and STD NMR spectra obtained for galectin-1 with **2** (top) and **3** (down). A protein:ligand ratio of 1:30 (being galectin-1 concentrated at 100 μM and galectin-3 at 50 μM) was employed, with 2s of saturation time at 298 K; the irradiation of the protein was set at δ -0.5 ppm . (B) Colour legend for the epitope mapping. (C) Epitope mapping derived from the STD-NMR experiments for galectin 1 (in blue) and galectin-3 (in green) with ligands **1**-**3**.

All the STD-NMR experiments performed showed clear STD signals, confirming the existence of molecular interaction. The STD-AF values were calculated from the spectra and the final STD % data obtained were mapped into the ligands' structure, as shown in Figure 4.3, C.

In general, the STD NMR spectra for galectin-1 displayed better quality (in terms of signal intensity) than those with galectin-3. Nevertheless, this inequality did not relate to change in affinity (Table 4.2), but, probably, to the different molecular weights of the two lectins. Indeed, the galectin-1 dimers of 30 kDa is a better vehicle for the saturation transfer when compared to the smaller CRD of galectin-3 (16 kDa).

From the comparison of the experiments performed for the same lectin/ligand combination but with different protein's irradiation frequencies, the STD-AF (Amplification Factor) displayed a slight systematic variation, higher for galectin-3 than for galectin-1 (see Supporting Information). This evidence was caused by a change in the quality of the protein irradiation. However, for both galectins, the calculated relative intensities (STD%) were constant as irradiation varied, translating the robustness of the binding epitope that was extrapolated from the analysis.

Focusing on the epitopes, the **canonical binding mode** described for the interaction of LacNAc with galectin-1 and galectin-3 was obtained for **1** with full reproducibility: the β-Gal ring provided the main contacts with the protein surface. [18,19]

Similarly, for **2** and **3**, significant STD intensities arose from H2, H3, H4, H5 and H6 of the β-Gal unit, being H4, H5 and H6 the most intense ones. This recognition mode is completely identical to that obtained for the LacNAc disaccharide and, in general, for β-galactosides interacting with galectins. [15]

Additional STD signals were provided by the GlcNAc moiety, being the acetyl signal the most intense. Moreover, as shown in the spectra and in the epitope maps of Figure 4.3, some weak STD intensities were also identified for the protons belonging to the scaffold unit. However, we may speculate that this is caused by transient contacts of the ligand with the protein surface.

Overall, the STD NMR data revealed that the multiple presentation of the LacNAc units did not modify the canonical recognition events with any of the galectins tested.

4.2.1.2 The lectin's perspective: Chemical shift perturbation analysis (CSP-NMR)

The lectin's perspective of the interactions under study was scrutinized by NMR employing ^{15}N labelled lectins and ^{1}H-^{15}N HSQC experiments. The chemical shift perturbations (CSP) and the intensity variation of the amide cross-peaks were followed upon ligand addition. [42] As discussed in the Introduction Chapter, this technique allows to obtain structural data of the recognition event at atomic level and strongly contributes to provide an all-round picture of the binding process. Molecules **1-3** were titrated into a sample containing ^{15}N labelled galectin-1 or galectin-3 and the chemical shift modifications upon ligand binding were measured and plotted in histogram charts (CSP plots, Figure 4.4). [43]

The results obtained for ligand **1** were **identical to the CSP plots** already described for the interaction of **LacNAc** and/or lactose with galectin-1 and galectin-3, being the most perturbed regions the S5 and S6 strands (black CSP plots, Figure 4.4). [15,18]
The same experiments performed with **2** and **3** displayed a different behaviour: **a huge intensity loss** of the ^{1}H-^{15}N cross-peaks was detected upon ligand addition, wiping out the interpretation of the CSP plots. Such phenomenon can be justified by i) statistical rebinding events of ligands with bivalent or trivalent epitope presentation (**2-3**) and ii) by the free-bound chemical exchange process. In both cases, a strong increase in the transverse relaxation rate of the nuclei could occur, leading to the decrease of their cross-peaks intensities. In order to facilitate the comparison, the CSP plots for the complexes galectin-1/**2**, galectin-1/**3**, galectin-3/**2**, galectin-3/**3** were only obtained using similar number of equivalents (between 5 and 7.5) of LacNAc epitopes per active site (Figure 4.4).

Hence, being the CSP interpretation rather elusive, the variations of the signals intensities for every amino acid were measured and plotted in terms of I_i-$I_f/\Delta I_{max}$ (Intensity-loss plots, Figure 4.4). The obtained histogram charts provided a visual representation of those signals that completely disappeared during the titration as well as the **relative intensity loss** of those that still provided CSP. Interestingly, the residues more perturbed in the CSP plot with the disaccharide **1** were those hat completely disappeared in the presence of the multivalent ligands **2-3** (highest value of intensity

loss). This evidence could be translated into structural information: the same amino acids involved in the binding to LacNAc were also recognizing **2** and **3**.

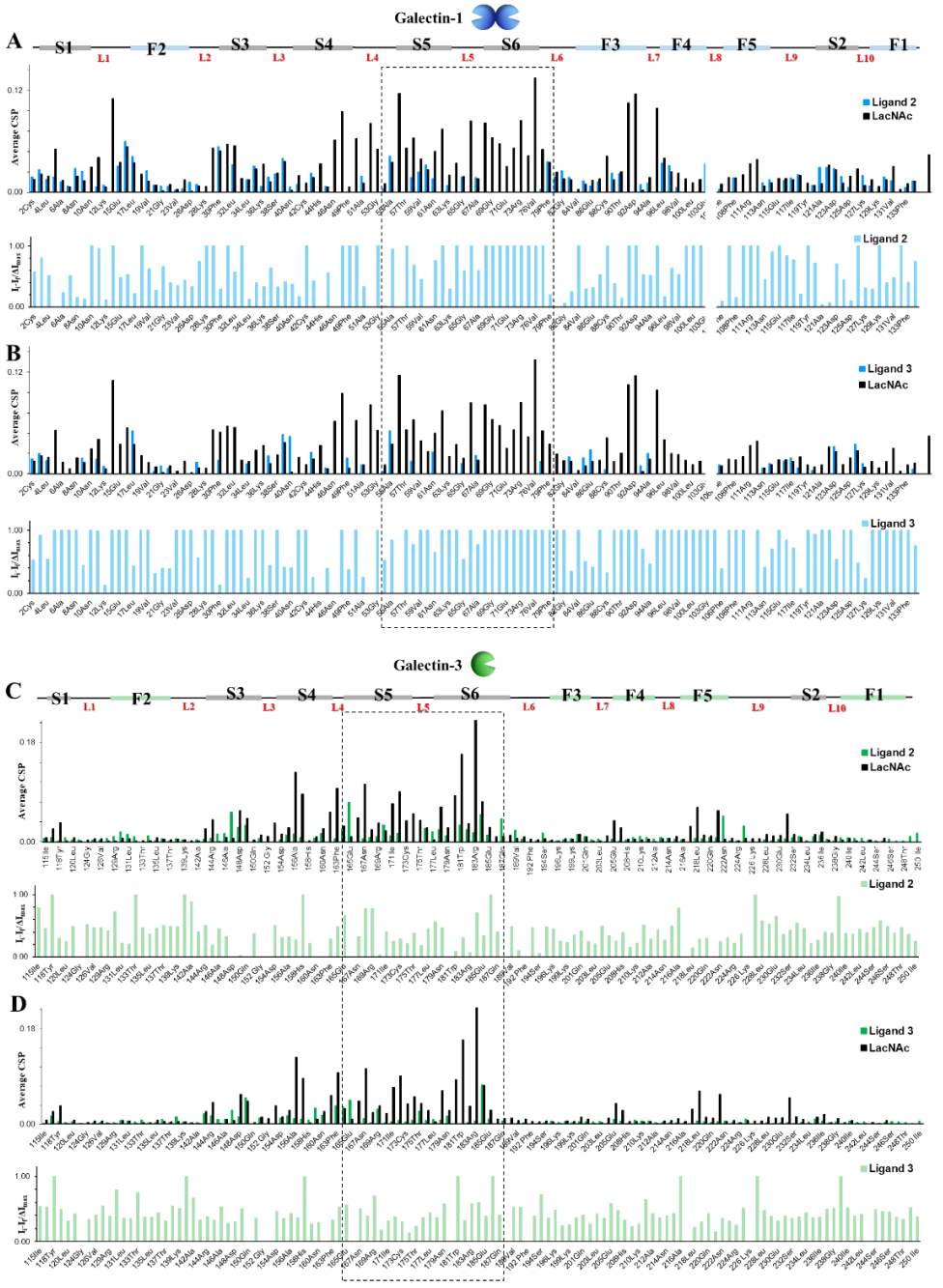


Figure 4.4. CSP and intensity plots obtained for the complexes: galectin-1/LacNAc (in black, panels A and B), galectin-1/**2** (in light blue, panel A), galectin-1/**3** (in light blue, panel B), galectin-3/**2** (in green, panel C), galectin-3/**3** (in green, panel D) and galectin-3/LacNAc (in black, panels C and D). The ^{15}N labelled galectins were employed at 100 μM (for galectin-1) and 50 μM (for galectin-3) and the amount of ligand to be added was calculated accordingly.

In addition, a general trend emerged when comparing the total intensity of the HSQC spectra with the different ligands and galectins: the intensity loss was more pronounced for galectin-1 over galectin-3 and resulted to be always higher for the trivalent molecule (**3**) over the bivalent (**2**) (Figure 4.5). This tendency can be explained by the transient formation of large supramolecular complexes that provide fast relaxation and the consequent line broadening of the signals. The more pronounced intensity loss with **3** over **2** could find an explanation in the triple epitope presentation of **3** that generate more favourable statistical rebinding effects. Accordingly, **3** was shown to bind with higher affinity both galectin-3 (4.4 μM) and galectin-1 (9 μM) when compared to the bivalent ligand **2** (12 μM and 19 μM respectively) (Table 4.2).

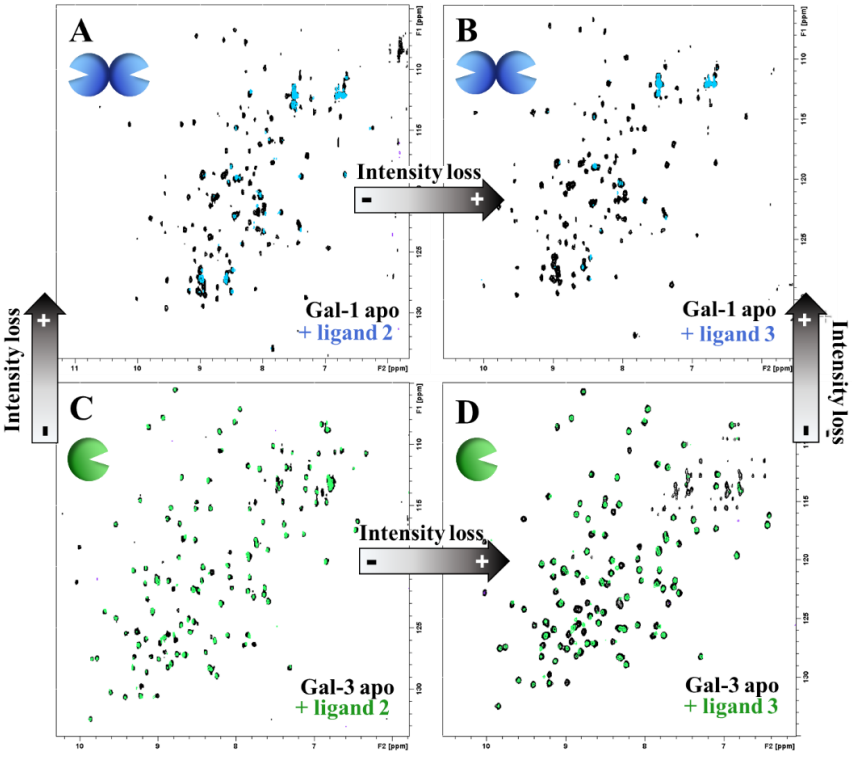


Figure 4.5. Comparison of the ^{1}H-^{15}N HSQC spectra. Above: (A) stacked spectra of galectin-1 *apo* (black) and after the addition of 5 equivalents (10 active equivalents) of **2** (light blue); (B) stacked spectra of galectin-1 *apo* (black) and upon addition of 2.5 equivalents (7.5 active equivalents) of **3** (lightblue). Below: (C) stacked specta of galectin-3 *apo* (black) and in the presence of 5 equivalents (10 active equivalents) of **2** (green); (D) stacked spectra of galectin-3 *apo* (black) and upon addition of 2.5 equivalents (7.5 active equivalents) of **3** (green, D). The ^{15}N labelled galectins were employed at the concentrations of 100 μM (for galectin-1) and 50 μM (for galectin-3) and the amount of ligand to be added was calculated accordingly. Black-grey arrows indicate the trend of the intensity loss between the spectra.

The considerably **higher intensity loss** displayed for **galectin-1** over galectin-3, suggests that a different type of interaction is taking place. As plausible explanation, supramolecular complexes can be generated in the case of galectin-1 not only because of the multiple epitope presentation, but also because of the dimeric architecture of this lectin. Thus, cross-linking effects are possible, guiding to the formation of larger aggregates, absent in the case of galectin-3. A visual representation to explain our NMR experimental results is illustrated in Figure 4.6.

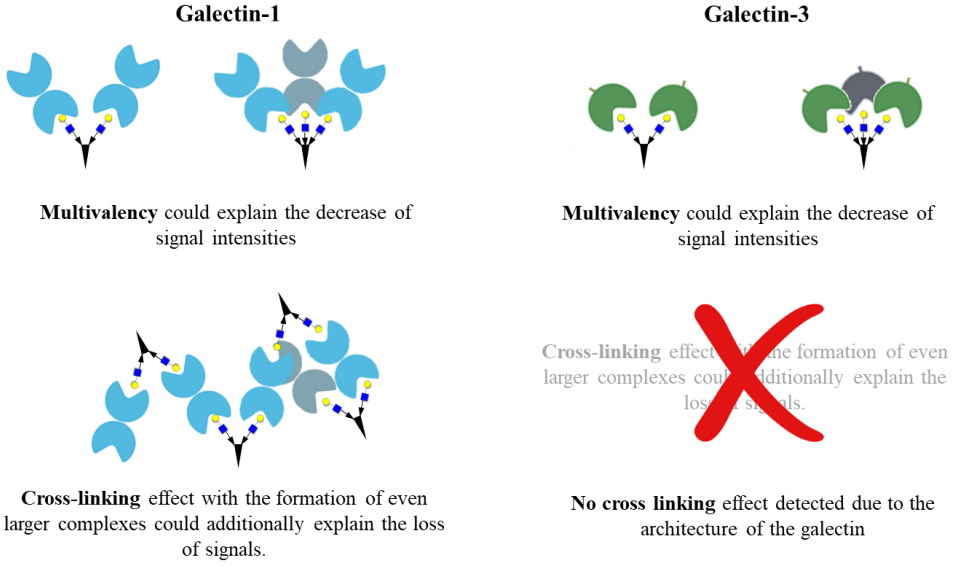


Figure 4.6. Illustration of the putative effects triggered by the interaction of multimeric ligands 2-3 with galectin-1 (left) and galectin-3 CRD (right).

4.2.2 Analysis of the complexes with the glycopolymers: competition experiments

In a first attempt to investigate glycopolymers **4-6** through CSP analysis, ^{15}N labelled galectin-1 and galectin-3 were employed. Nevertheless, the ^{1}H-^{15}N spectra acquired after the addition of small amounts (lectin:ligand ratio of 5:1) of **4-6** to the NMR tube containing the protein appeared almost completely empty. The lectin:ligand ratio of 5:1 was rationalized, considering the epitope repetitions in the glycopolymer, in terms of LacNAc epitopes for galectin's site of ligand. In that way, for galectin-1, 1.5 equivalents of epitopes were estimated for lectin's binding site in each case while for galectin-3, 4.5 equivalents of the polymeric compounds were added separately. The loss of the cross-peaks of the backbone amides was in all cases rather drastic and no information concerning the interaction from the protein's point of view could be

deduced. This evidence may be explained by the formation of large supramolecular structures, undetectable by NMR. Therefore, we decided to change the usual strategy to design an *ad hoc* setup for the experiment (Figure 4.7). Indeed, we titrated increasing amounts of a small **competitor compound** were titrated into the sample containing the lectin and the glycopolymer. LacNAc (**1**), whose interaction with galectin-1 and galectin-3 is well known, was chosen for this purpose. [18,19] Since **1,** as well as **4**-**6**, interact with the lectins through the canonical binding site, the presence of large excess of LacNAc in solution shifted the equilibrium from the lectin:glycopolymer complexes to generate lectin:LacNAc complexes. As a result, the cross-peaks were gradually visible, with the magnitude of this recovery depending on the relative affinity of the lectin-glycopolymer complex.

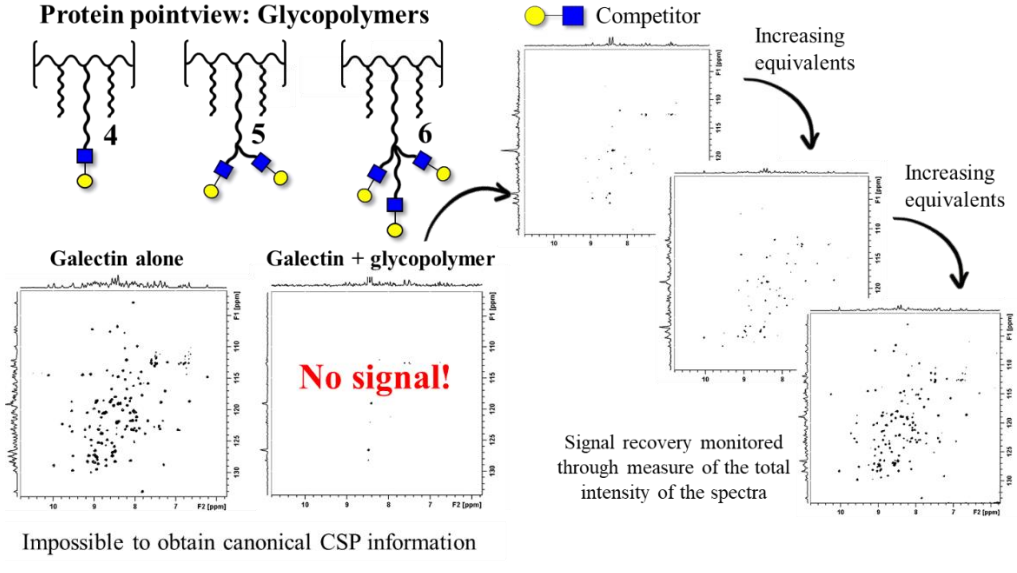


Figure 4.7. Strategy adopted to monitor the interaction between galectin-1 and galectin-3 with glycopolymers **4**, **5** and **6** from the protein viewpoint.

The total intensities of the HSQC spectra recorded during the titration were evaluated and expressed in percentage (%) respect to the intensity of the spectrum registered for the protein alone (being 100%). Fittingly, the lectin cross-peaks became visible again upon addition of increasing amounts of **1**. The highest grade of the recovery of the signals was reached in all cases in the last point of the titration (with the highest equivalents of LacNAc), when the equilibrium was strongly shifted toward the formation of the lectin:LacNAc complexes.

For **galectin-1** (Figure 4.8 A), the complete signal recovery was never achieved. A maximum recovery of 66% was measured for ligand **6** using 20 equivalents of the competitor. The amount of glycopolymers added to the lectin was calculated in terms of LacNAc epitopes per galectin site of ligand and was set at 1.5 equivalents in the three cases. Comparing the results obtained for ligands **4**-**6**, it was apparent that they display different tendencies. The individual LacNAc-distributed polymer **4** caused, upon its addition to the lectin, a dramatic loss of signal (only 4% of the signals were still visible). During the titration with the competitor, the 12% of the total intensity was recovered upon addition of 5 equivalents, and the 55% with the final point, using 75 equivalents of LacNAc. (Figure 4.8).

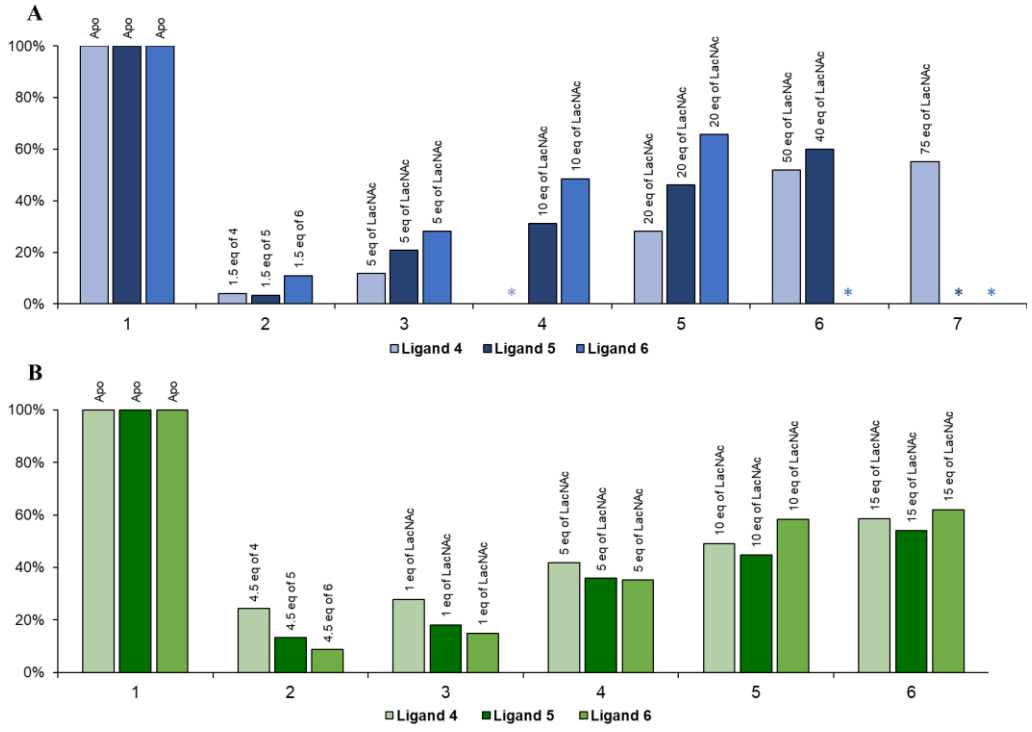


Figure 4.8. ^{1}H-^{15}N HSQC competition experiments with ligands **4**–**6**. The relative total intensity of each acquired spectra is reported and the ligand are represented with different shades of the same color. (A) Histogram chart for the data of galectin-1 *apo* (x-axis, group 1), galectin-1 with 1.5 equivalents of LacNAc epitopes per galectin site of ligands (**4**–**6**, group 2), and following signal recovery (groups 3–7 upon addition of 5, 10, 20, 40, and 75 equivalents of competitor **1**, respectively). (B) Histogram chart for the total relative intensity values calculated for galectin-3 *apo* (group 1), galectin-3 with 4.5 equivalents of LacNAc epitopes per galectin site of ligands (**4**–**6**, group 2), and following signal recovery (groups 3–6 with the addition of 1, 5, 10, and 15 equivalents of competitor **1**, respectively). Symbol * stands for spectra not acquired.

For ligand **5**, a similar initial loss of signals was detected, but the signal recovery during the titration with LacNAc was generally faster (21% with 5 equivalents, and 60% with 40 equivalents). This evidence suggested that no cooperativity effect due to the bivalent presentation of the epitope in compound **5** was occurring and, furthermore, the binding of **5** to galectin-1 was even less efficient than that of **4**.
The addition of the multimeric **6** to galectin-1 induced the loss of the 89% of the signals (11% of the signals could still be detected), a less dramatic response to those obtained for glycopolymers **4** and **5**. Accordingly, the signals' recovery upon LacNAc addition was the fastest one: 28% with 5 equivalents of LacNAc and 66% with only 20 equivalents. Once more, the multiple presentation of the epitope did not induce cooperative effects, but even worsened the recognition process.

A different behaviour was observed with **galectin-3** (Figure 4.8, B). In this case, the amount of glycopolymers to be added to the *apo* samples was calculated at 4.5 equivalents of LacNAc epitopes per galectin site in all cases. First of all, the loss of the intensities of the HSQC cross peaks of galectin-3 in the presence of the glycopolymers **4**-**6** was always less pronounced than for galectin-1. In fact, the 24% of the signal intensities were still present upon addition of **4**, 13% upon addition of **5**, and 9% upon addition of **6**. Moreover, the signal recovery was faster, reaching around the 60% for all the glycopolymers in the presence of only 15 equivalents of the competitor. Thus, although in the samples of galectin-3, more LacNAc epitopes were added in comparison to the samples of galectin-1 (4.5 vs 1.5 equivalents of **4**-**6**), the competitor produced a considerably faster signal recovery with galectin-3. Herein, the equilibrium shifted quicker towards the formation of lectin:LacNAc complexes than for galectin-1, leading to the conclusion that a less effective binding of glycopolymers **4**-**6** to galectin-3 was taking place. In addition, the recovery trend was similar for all the ligands tested, with no evident preference for any of them, meaning that no cooperativity effects driven by the multiple presentation were occurring.

As conclusion of these competition NMR experiments with glycopolymers **4**-**6,** a better affinity of these molecules for galectin-1 over galectin-3 is taking place, as also deduced from the ELISA measurements (Table 4.2). Besides, regardless of the fact that

no clear preference was detected for galectin-3 with any molecule, galectin-1 moderately preferred the glycopolymer that features a monovalent LacNAc presentation. Presumably, no cross linking-effects were occurring in the samples containing galectin-3 because of its intrinsic monomeric architecture. Contrarily, it is tempting to guess that the dimeric nature of galectin-1 is probably favouring the generation of cross-linked macromolecular structures that enhances the affinity.

4.3 Dynamic Light Scattering (DLS) measurements

In order to deepen into the nature of the supramolecular complexes deduced by NMR, DLS measurements were carried out for galectin-1 and galectin-3 in the presence and absence of the glycopolymer **4**. With this technique, the hydrodynamic radius of the species in solution is deduced, and therefore their effective sizes can be approximated. The DLS results for the complexes were compared to those obtained for the biomolecules alone (for both the galectins and the glycopolymer), as presented in Figure 4.9. Specifically, the hydrodynamic radius estimated for galectin-1 alone was 4.54 nm vs 1.56 nm of the CRD of galectin-3. The value obtained for the particle generated by the glycopolymer bearing the single LacNAc presentation was 7.47 nm. Subsequently, the samples with complexes were prepared with a 5:1 galectin:ligand ratio :1 in both cases, corresponding to 1.5 LacNAc epitopes available per galectin-1 binding site and 3 LacNAc epitopes per galectin-3 binding site (according to the NMR ratios). The resulting hydrodynamic radius of the complex formed by glycopolymer **1** and galectin-3 barely exceeded the sum of the two biomolecules alone (8.43 nm). A completely different observation was carried out in the case of the sample containing galectin-1 and **4**. In this case, the particle size enormously increased to 875 nm.

Fittingly, these data are in full agreement with the NMR evidences: much larger supramolecular complexes are formed when galectin-1 interacts with the glycopolymers, probably due to cross-linking effects, while no evidences for this phenomenon are observed for galectin-3.

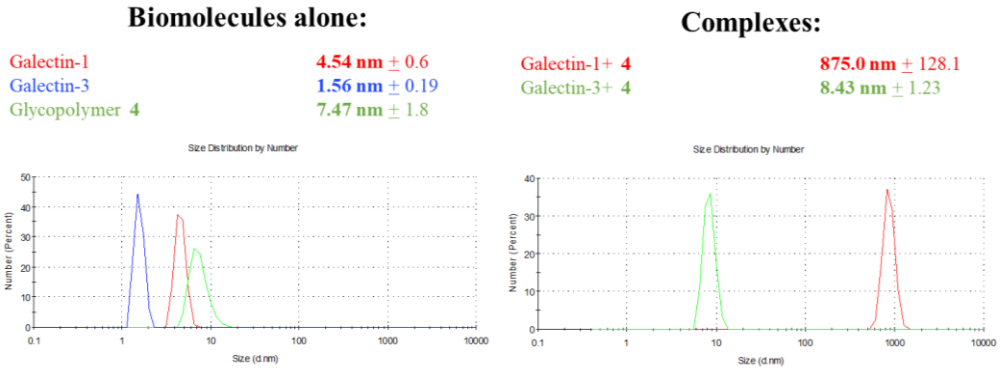


Figure 4.9. Hydrodynamic radius distribution by dynamic light scattering. Left: hydrodynamic radius distribution for the biomolecules alone; the respective concentrations are reported. Right: hydrodynamic radius graphs for ligand **4** complexed with galectin-3 and galectin-1; the respective concentrations are reported.

4.4 Cryo-EM experiments

Finally, cryo-EM experiments were performed for the biomolecules alone as a control (galectin-1, galectin-3 and glycopolymer **4**) as well as for their complexes.

The lectin:ligand ratio was set at 5:1 and the concentrations were adjusted and optimized to the ranges of the technique (0.5 mg/mL for the protein and 0.17 mg/mL for the ligand). In order to avoid the presence of any type of contaminants or aggregates, a size exclusion chromatography was first performed to the samples before the cryo-EM experiments, as described in Chapter 9.

The acquired images were visually inspected to perform a quality control on the type and morphology of the samples (Figure 4.10). The biomolecule alone, as control, displayed a homogeneous distribution, without any aggregates detected. Instead, the formation of networks were visible in the images with the complexes. Comparing the two mixtures (galectin-1/**4** and galectin-3/**4**), two remarkably different phenomena could be distinguished: the supramolecular complexes formed by the interaction of galectin-1 and glycopolymer **4** were much more pronounced and perceptible than those observed with the same ligand in the presence of galectin-3. The presence of cross-linking effects is now evident.

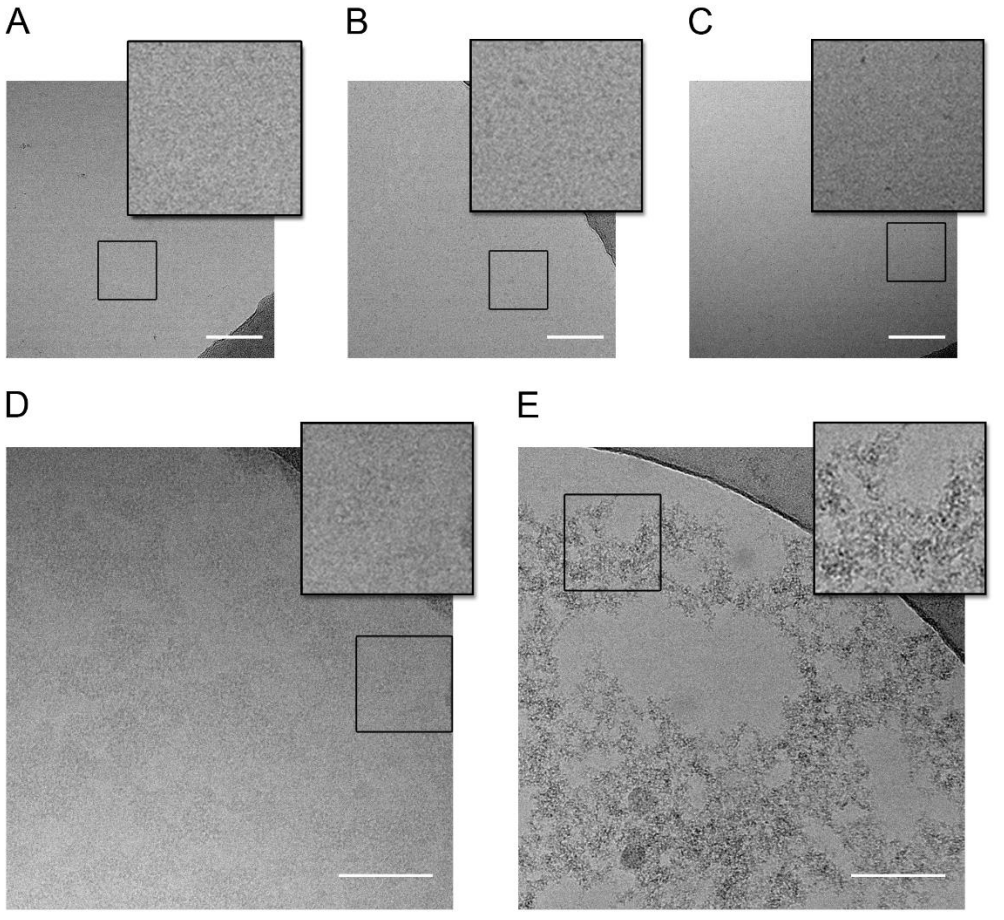


Figure 4.10. Cryo-EM images. Top figure: controls of galectin-1 (0.5 mg/mL, A), glycopolymer **4** (0.17 mg/mL, B) and galectin-3 (0.5 mg/mL, C). Below: images of the complexes obtain with galectin-3 CRD and glycopolymer **4** (D) and galectin-1 and glycopolymer **4** (E). The scale bar is set at 100 nm.

4.5 Conclusions

The interaction of a variety of glycopolymers with two human galectins of biomedical interest (galectin-1 and the CRD of galectin-3) has been deeply analysed. A detailed picture of the recognition event at different degrees of resolution have been provided: from the atomic to the supramolecular perspective. To achieve this goal, a combination of technique has been synergically employed: NMR, DLS, and cryo-EM.

First, starting from the ligands' perspective, the STD-NMR experiments showed that the constituting units of the glycopolymers (**2**-**3**) are recognized by both galectins in a canonical manner through the LacNAc disaccharide, mainly driven by favourable

interactions with the β-Gal moiety. A canonical recognition mode was also defined when focusing on the protein point of view with titrations monitored through ^{1}H-^{15}N HSQC and consequent CSP and intensity analysis. Noticeably, a high reduction of the HSQC cross peaks was observed in all the cases upon ligand addition, strongly suggesting the presence of transient macromolecular complexes. This tendency was more significant with the trivalent LacNAc presenting molecule (**3**) over the bivalent one (**2**). Furthermore, the event was much more pronounced for galectin-1 than for galectin-3.

^{1}H-^{15}N HSQC NMR experiments were also performed for galectin-1 and galectin-3 in the presence of the glycopolymers (**4**-**6**). Herein, a drastic loss of the intensities of the cross-peaks made the analysis of the CSP impossible, suggesting the formation of supramolecular entities much larger than those formed by the parent ligands **2** and **3**. Hence, to gain information on the binding event from these samples, the signal recovery induced by adding an excess of a known competitor (**1**), was followed. A clear discrimination between galectins was detected in this way, with the signal recovery being always faster for galectin-3. We proposed that such an event is linked to a faster shift of the equilibrium from the lectin:glycopolymer to lectin:LacNAc complex, a clear indication of a worse affinity of galectin-3 towards the glycopolimers when compared to galectin-1.

Focusing on the data for the different glycopolymers, negligible changes in the recovery rate were detected for galectin-3, whereas for galectin-1, **4** exhibited considerably lower recovery rates than **5**-**6**, which relates to a better affinity. A first explanation of these results was attributed to the **generation of cross-linked entities** formed by **galectin-1**, apart from the effect of the intrinsic multiple presentation of the glycopolymers. This cross-linking intermolecular effect was not present in the case of the galectin-3, due to its monomeric architecture.

These evidences were completely in line with previous work from our collaborators, based on competitive ELISA assays, which proposed a general better avidity (in the nanomolar range) for the glycopolymers interacting with galectin-1 over galectin-3 (low micromolar range). [1] In particular, glycopolymer **4** emerged as the best candidate considering its relative potency per active unit (rp/*n*). The bivalent or trivalent presentation of LacNAc on the **5** and **6** branches did not trigger any affinity

enhancement. Contrarily, galectin-3 here preferred the clustered LacNAc epitopes, shedding light into how the ligand density and orientation can exquisitely modulate the association event.

The systems were further investigated through DLS and cryo-EM techniques. Our hypothesis of cross-linking effects promoted by galectin-1 was unambiguously confirmed with both methodologies. In fact, the DLS measurements showed how the hydrodynamic radius of the particles generated in the solution containing galectin-1 and glycopolymer **4** suffered a huge increase in comparison with the relative values of the macromolecules alone. The same kind of observation could not be estimated for the mixture of galectin-3 and **4**, where a particle size slightly larger than the sum of the two entities alone was measured.

The images acquired through cryo-EM further confirmed our hypothesis: supramolecular networks were visible both for galectin-3:**4** and galectin-1:**4** mixtures because of the formation of supramolecular structures driven by the multivalency of the glycopolymers. Nevertheless, for galectin-1 those networks were more evident, fitting with the existence of cross-linking intermolecular effects orchestrated by the lectin dimer.

Overall, this is an exquisite example of how a **combined experimental approach** provides complementary outcomes essential to define the global picture of a multivalent interaction.

The multivalent effect of these molecules triggers affinity enhancement when compared to their monovalent counterparts. However, the multiple presentations on the glycopolymer's branches did not entail a clear affinity increase, especially for galectin-1. We hypothesize that the reason is steric hindrance: the simultaneous recognition of more than one dimeric lectin at a single branch presenting di- or trivalent epitope is hindered in a crowded microenvironment or, at least, disfavoured, also in entropy. Thus, a single sugar presentation for the branches of the glycopolymers is sufficient to reach great avidity in the case of galectin-1.

The NMR data for galectin-3 are less biased towards a clear preference and the affinities are worst; therefore, we can confirm that those particular glycopolymers bring a **clear discrimination between the galectins**. This selectivity based on the

structure of a multivalent molecule is exceptional and pave the way to the development of compounds of biomedical interest, also considering the compatible characteristics of HMPA polymers with therapeutic applications.

Moreover, the figures obtained by Cryo-EM of the mixtures visually testify, for the first time, the different networks formed by the monomeric and dimeric galectins with the glycoclusters and confirm the presence of **cross-linked intermolecular entities** mediated by the prototype galectin.

4.6 References

1. Tavares, M.R.; Bláhová, M.; Sedláková, L.; Elling, L.; Pelantová, H.; Konefał, R.; Etrych, T.; Křen, V.; Bojarová, P.; Chytil, P. High-Affinity N-(2-Hydroxypropyl)methacrylamide Copolymers with Tailored N-Acetyllactosamine Presentation Discriminate between Galectins. *Biomacromolecules* **2020**, *21*, 641–652, doi:10.1021/ACS.BIOMAC.9B01370/SUPPL_FILE/BM9B01370_SI_001.PDF.
2. Bertuzzi, S.; Gimeno, A.; Martinez-castillo, A.; Lete, M.G.; Delgado, S.; Airoldi, C.; Tavares, M.R.; Bláhová, M.; Chytil, P.; Křen, V.; et al. Cross-linking effects dictate the preference of galectins to bind lacnac-decorated hpma copolymers. *Int. J. Mol. Sci.* **2021**, *22*, 22, doi:10.3390/ijms22116000.
3. Ebrahim, A.H.; Alalawi, Z.; Mirandola, L.; Rakhshanda, R.; Dahlbeck, S.; Nguyen, D.; Jenkins, M.; Grizzi, F.; Cobos, E.; Figueroa, J.A.; et al. Galectins in cancer: Carcinogenesis, diagnosis and therapy. *Ann. Transl. Med.* **2014**, *2*, doi:10.3978/j.issn.2305-5839.2014.09.12.
4. Bartolazzi, A. Galectins in Cancer and Translational Medicine: From Bench to Bedside. *Int. J. Mol. Sci.* **2018**, *19*, doi:10.3390/IJMS19102934.
5. Videla-Richardson, G.A.; Morris-Hanon, O.; Torres, N.I.; Esquivel, M.I.; Vera, M.B.; Ripari, L.B.; Croci, D.O.; Sevlever, G.E.; Rabinovich, G.A. Galectins as Emerging Glyco-Checkpoints and Therapeutic Targets in Glioblastoma. *Int. J. Mol. Sci.* **2021**, *23*, doi:10.3390/IJMS23010316.
6. Jeschke, U.; Karsten, U.; Wiest, I.; Schulze, S.; Kuhn, C.; Friese, K.; Walzel, H. Binding of galectin-1 (gal-1) to the Thomsen-Friedenreich (TF) antigen on trophoblast cells and inhibition of proliferation of trophoblast tumor cells in vitro by gal-1 or an anti-TF antibody. *Histochem. Cell Biol.* **2006**, *126*, 437–444, doi:10.1007/S00418-006-0178-1.
7. 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, Vol. 17, Page 1566* **2016**, *17*, 1566, doi:10.3390/IJMS17091566.
8. Nangia-Makker, P.; Balan, V.; Raz, A. Regulation of tumor progression by extracellular galectin-3. *Cancer Microenviron.* 2008, *1*, 43–51.
9. Capone, E.; Iacobelli, S.; Sala, G. Role of galectin 3 binding protein in cancer progression: a potential novel therapeutic target. *J Transl Med* **2021**, *19*, 405, doi:10.1186/s12967-021-03085-w.
10. Blanchard, H.; Yu, X.; Collins, P.M.; Bum-Erdene, K. Galectin-3 inhibitors: A patent review (2008-present). *Expert Opin. Ther. Pat.* **2014**, doi:10.1517/13543776.2014.947961.
11. 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.
12. Girard, A.; Magnani, J.L. Clinical Trials and Applications of Galectin Antagonists. *Trends Glycosci. Glycotechnol.* **2018**, doi:10.4052/tigg.1744.1se.
13. Dings, R.P.M.; Miller, M.C.; Griffin, R.J.; Mayo, K.H. Galectins as molecular targets for therapeutic intervention. *Int. J. Mol. Sci.* 2018.
14. Laaf, D.; Bojarová, P.; Elling, L.; Křen, V. Galectin–Carbohydrate Interactions in Biomedicine and Biotechnology. *Trends Biotechnol.* **2019**, *37*, 402–415, doi:10.1016/J.TIBTECH.2018.10.001.
15. Bertuzzi, S.; Quintana, J.I.; Ardá, A.; Gimeno, A.; Jiménez-Barbero, J. Targeting Galectins With Glycomimetics. *Front. Chem.* 2020, *8*, 593.
16. Hirabayashi, J.; Hashidate, T.; Arata, Y.; Nishi, N.; Nakamura, T.; Hirashima, M.; Urashima, T.; Oka, T.; Futai, M.; Muller, W.E.G.; et al. Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572*, 232–254, doi:10.1016/S0304-4165(02)00311-2.

17. 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*, 10109–10123, doi:10.1074/jbc.M709545200.
18. Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M.A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F.J.; et al. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie Int. Ed.* **2019**, *58*, 7268–7272, doi:10.1002/anie.201900723.
19. Bertuzzi, S.; Gimeno, A.; Núñez-Franco, R.; Bernardo-Seisdedos, G.; Delgado, S.; Jiménez-Osés, G.; Millet, O.; Jiménez-Barbero, J.; Ardá, A. Unravelling the Time Scale of Conformational Plasticity and Allostery in Glycan Recognition by Human Galectin-1. *Chem. - A Eur. J.* **2020**, *26*, 15643–15653, doi:10.1002/CHEM.202003212.
20. Moure, M.J.; Gimeno, A.; Delgado, S.; Diercks, T.; Boons, G.J.; Jiménez-Barbero, J.; Ardá, A. Selective 13C-Labels on Repeating Glycan Oligomers to Reveal Protein Binding Epitopes through NMR: Polylactosamine Binding to Galectins. *Angew. Chemie Int. Ed.* **2021**, *60*, 18777–18782, doi:10.1002/ANIE.202106056.
21. 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*, 10841–10847, doi:10.1074/jbc.M312834200.
22. Fermino, M.L.; Polli, C.D.; Toledo, K.A.; Liu, F.-T.; Hsu, D.K. LPS-Induced Galectin-3 Oligomerization Results in Enhancement of Neutrophil Activation. *PLoS One* **2011**, *6*, 26004, doi:10.1371/journal.pone.0026004.
23. Zhao, Z.; Xu, X.; Cheng, H.; Miller, M.C.; He, Z.; Gu, H.; Zhang, Z.; Raz, A.; Mayo, K.H.; Tai, G.; et al. Galectin-3 N-terminal tail prolines modulate cell activity and glycan-mediated oligomerization/phase separation. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, doi:10.1073/PNAS.2021074118/-/DCSUPPLEMENTAL.
24. Bernardi, A.; Jiménez-Barbero, J.; Casnati, A.; De Castro, C.; Darbre, T.; Fieschi, F.; Finne, J.; Funken, H.; Jaeger, K.E.; Lahmann, M.; et al. Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev.* **2013**, *42*, 4709–4727, doi:10.1039/C2CS35408J.
25. Cecioni, S.; Imberty, A.; Sébastien Vidal, S. Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. **2014**, doi:10.1021/cr500303t.
26. Porkolab, V.; Pifferi, C.; Sutkeviciute, I.; Ordanini, S.; Taouai, M.; Thépaut, M.; Vivès, C.; Benazza, M.; Bernardi, A.; Renaudet, O.; et al. Development of C-type lectin-oriented surfaces for high avidity glycoconjugates: towards mimicking multivalent interactions on the cell surface. *Org. Biomol. Chem.* **2020**, *18*, 4763–4772, doi:10.1039/D0OB00781A.
27. Su, L.; Feng, Y.; Wei, K.; Xu, X.; Liu, R.; Chen, G. Carbohydrate-Based Macromolecular Biomaterials. *Chem. Rev.* **2021**, *121*, 10950–11029, doi:10.1021/ACS.CHEMREV.0C01338/ASSET/IMAGES/MEDIUM/CR0C01338_0045.GIF.
28. Goti, G.; Colombo, C.; Achilli, S.; Vivès, C.; Thépaut, M.; Luczkowiak, J.; Labiod, N.; Delgado, R.; Fieschi, F.; Bernardi, A. Precision Glycodendrimers for DC-SIGN Targeting. *European J. Org. Chem.* **2022**, *2022*, e202200113, doi:10.1002/EJOC.202200113.
29. Vrbata, D.; Filipová, M.; Tavares, M.R.; Červený, J.; Vlachová, M.; Šírová, M.; Pelantová, H.; Petrásková, L.; Bumba, L.; Konefał, R.; et al. Glycopolymers Decorated with 3-O-Substituted Thiodigalactosides as Potent Multivalent Inhibitors of Galectin-3. *J. Med. Chem.* **2022**, *65*, 3866–3878, doi:10.1021/ACS.JMEDCHEM.1C01625/SUPPL_FILE/JM1C01625_SI_002.CSV.

30. Sakamoto, J.I.; Koyama, T.; Miyamoto, D.; Yingsakmongkon, S.; Hidari, K.I.P.J.; Jampangern, W.; Suzuki, T.; Suzuki, Y.; Esumi, Y.; Hatano, K.; et al. Thiosialoside clusters using carbosilane dendrimer core scaffolds as a new class of influenza neuraminidase inhibitors. *Bioorganic Med. Chem. Lett.* **2007**, *17*, 717–721, doi:10.1016/j.bmcl.2006.10.085.
31. Soomro, Z.H.; Cecioni, S.; Blanchard, H.; Praly, J.P.; Imberty, A.; Vidal, S.; Matthews, S.E. CuAAC synthesis of resorcin[4]arene-based glycoclusters as multivalent ligands of lectins. *Org. Biomol. Chem.* **2011**, *9*, 6587–6597, doi:10.1039/c1ob05676j.
32. Zhang, H.; Laaf, D.; Elling, L.; Pieters, R.J. Thiodigalactoside-Bovine Serum Albumin Conjugates as High-Potency Inhibitors of Galectin-3: An Outstanding Example of Multivalent Presentation of Small Molecule Inhibitors. *Bioconjug. Chem.* **2018**, *29*, 1266–1275, doi:10.1021/acs.bioconjchem.8b00047.
33. Zhou, C.; Reesink, H.L.; Putnam, D.A. Selective and Tunable Galectin Binding of Glycopolymers Synthesized by a Generalizable Conjugation Method. *Biomacromolecules* **2019**, *20*, 3704–3712, doi:10.1021/ACS.BIOMAC.9B00759/SUPPL_FILE/BM9B00759_SI_001.PDF.
34. Heine, V.; Hovorková, M.; Vlachová, M.; Filipová, M.; Bumba, L.; Janoušková, O.; Hubálek, M.; Cvačka, J.; Petrásková, L.; Pelantová, H.; et al. Immunoprotective neo-glycoproteins: Chemoenzymatic synthesis of multivalent glycomimetics for inhibition of cancer-related galectin-3. *Eur. J. Med. Chem.* **2021**, *220*, 113500, doi:10.1016/j.ejmech.2021.113500.
35. Chytil, P.; Koziolová, E.; Etrych, T.; Ulbrich, K. HPMA Copolymer–Drug Conjugates with Controlled Tumor-Specific Drug Release. *Macromol. Biosci.* **2018**, *18*, 1700209, doi:10.1002/MABI.201700209.
36. Bojarová, P.; Chytil, P.; Mikulová, B.; Bumba, L.; Konefał, R.; Pelantová, H.; Krejzová, J.; Slámová, K.; Petrásková, L.; Kotrchová, L.; et al. Glycan-decorated HPMA copolymers as high-affinity lectin ligands. *Polym. Chem.* **2017**, *8*, 2647, doi:10.1039/c7py00271h.
37. Bojarová, P.; Tavares, M.R.; Laaf, D.; Bumba, L.; Petrásková, L.; Konefał, R.; Bláhová, M.; Pelantová, H.; Elling, L.; Etrych, T.; et al. Biocompatible glyconanomaterials based on HPMA-copolymer for specific targeting of galectin-3. *J. Nanobiotechnology* **2018**, *16*, 1–16, doi:10.1186/S12951-018-0399-1/FIGURES/2.
38. Filipová, M.; Bojarová, P.; Rodrigues Tavares, M.; Bumba, L.; Elling, L.; Chytil, P.; Gunár, K.; Křen, V.; Etrych, T.; Janoušková, O. Glycopolymers for Efficient Inhibition of Galectin-3: In Vitro Proof of Efficacy Using Suppression of T Lymphocyte Apoptosis and Tumor Cell Migration. *Biomacromolecules* **2020**, *21*, 3122–3133,doi:10.1021/ACS.BIOMAC.0C00515/SUPPL_FILE/BM0C00515_SI_001.PDF.
39. Mayer, M.; Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788, doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.
40. Mayer, M.; Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **2001**, doi:10.1021/ja0100120.
41. Viegas, A.; Manso, J.; Nobrega, F.L.; Cabrita, E.J. Saturation-transfer difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J. Chem. Educ.* **2011**, *88*, 990–994, doi:10.1021/ed101169t.
42. Ardá, A.; Jiménez-Barbero, J. The recognition of glycans by protein receptors. Insights from NMR spectroscopy. *Chem. Commun.* **2018**, *54*, 4761–4769, doi:10.1039/C8CC01444B.
43. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16, doi:10.1016/J.PNMRS.2013.02.001.

CHAPTER 5

THE RECOGNITION OF LSECTIN BY ASYMMETRIC N-GLYCANS IN SOLUTION *VS* ON MICROARRAYS

The branch specificity is dictated by the immobilization of asymmetric N-glycans on surfaces

5.1 Introduction

The detailed study of the molecular recognition of complex asymmetric N-glycans by LSECtin has been conducted. The results, presented in this Chapter, have been satisfactorily published in 2022. [1]

Our collaborators from CIC biomaGUNE had previously proposed an outstanding ligand selectivity for one individual positional isomer over the other, when interrogating with LSECtin an array composed of asymmetric N-glycans. [2] Driven by the aim of adding detailed structural insights on the observed preferences, we have investigated the same system employing NMR methodologies assisted by MD simulations. Specifically, we selected a panel of sugars including the disaccharide GlcNAcβ1-2Man, the non-elongated N-glycan (**G0**) and two asymmetric N-glycans (**LDN3** and **LDN6**) to deeply characterize their interactions with the CRD of LSECtin (Figure 5.1 & Figure 5.3). The lectin was provided by the group of Dr. Franck Fieschi (IBS, University of Grenoble), whereas the ligands were synthetized in the laboratory of Dr. Niels C. Reichardt (CIC biomaGUNE, San Sebastian, Spain). The MD simulations presented herein were carried out and analysed by the Computational Chemistry Lab (CIC bioGUNE). Additional information as well as the details regarding the MD simulations can be found in the publication. [1]

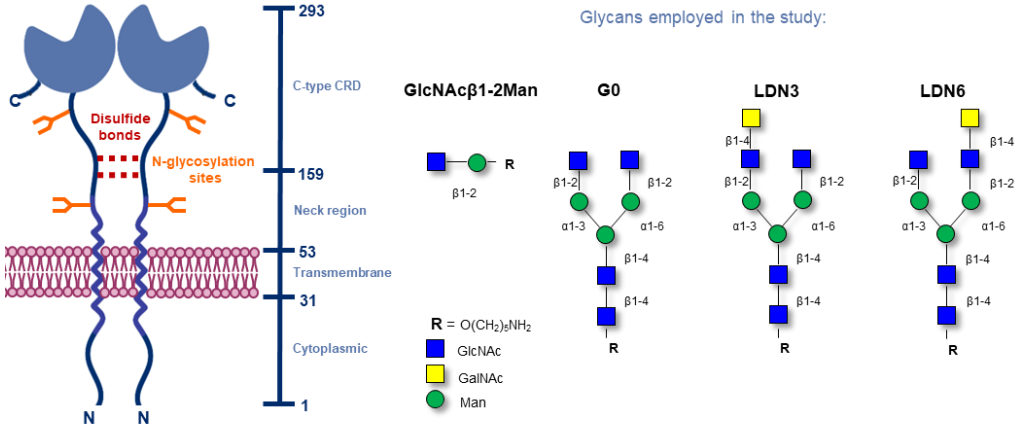


Figure 5.1. On the left: model of the full receptor dimer of LSECtin (only the CRD has been employed in our analysis); on the right: SNFG representation of the studied ligands (the disaccharide GlcNAcβ1-2Man, the non-elongated **G0,** and the asymmetric **LDN3** and **LDN6**).

5.1.1 General context

Glycan-protein interactions are now universally recognized as key regulatory events for the onset of various diseases. [3] Consequently, the comprehensive elucidation of the involved recognition processes is fundamental to design molecules able to modulate interaction and eventually, control the pathological response. [4]

The low affinity that characterizes the basic monovalent sugar-lectin binding event is overcome by Nature through the engagement of multivalent interactions (for both the receptor and the ligand), which remarkably enhance the affinity. [5] The strategy of multimeric binding has been successfully exploited for the rational design of molecules able to target lectins; [6] this aspect has been discussed in the general Introduction and, additionally, a practical example has been presented in Chapter 4 of this thesis.

However, the binding event does not only depend on the composition and stoichiometry of the lectin/ligand complex, but also on the spatial organisation of both actors and, in particular, of the glycan counterpart, intrinsically characterized by an extreme flexibility. [7–11] Indeed, besides multivalency, also glycan presentation plays a crucial role in the recognition process. Aspects such as the orientation, the dynamics, the flexibility, the accessibility, and the density of the oligosaccharides have been proven to be decisive at the time of interacting with the protein partner. [12] Specifically, Grant et al. have predicted, by using MD simulations, how modifications in the glycan presentation on an array surface can influence the final sugar binding to the lectin. They guessed how changes in the length, in the organization, and in the orientation of the linker that anchors the glycan to the solid support, could drastically change the final 3D presentation of the saccharide, compromising the recognition. [12]

In the last two decades, printed glycan microarrays have emerged as a powerful technique to study lectin-sugar interaction in a high throughput manner. [13–16] As commented in the general Introduction, they are essential tools in Glycosciences and allow a rapid screening of hundreds of glycan structures against a protein target. Glycan array-based studies have shown, in multiple occasions, how the lectin binding toward N-glycans can be highly influenced by the branch position of the recognized main epitope. [17–19] However, this methodology has also some limitations because diverse factors can influence the final output: orientation, packing, and valency of the

glycan. In fact, the solid support, as well as the linker, are not precisely mimicking the biochemical characteristics of an *in-vivo* microenvironment but, on the contrary, they are an additional external biophysical components that may somehow affect the final measurements. [12,20] The physical properties of the linker and of the solid support can determine how the glycan is presented and, therefore, can tune its availability for receptor binding. This aspect has to be considered and a good way to corroborate the array screening is to subsequently employ complementary techniques to validate the results, to provide a global picture of the interaction, and to avoid inaccuracies.

For that reason, we investigated through NMR and MD simulations the interaction of the N-glycan positional isomers described above with LSECtin. [2]

LSEctin is a human transmembrane receptor involved in innate immunity. A comprehensive overview of the information available on this lectin has been given in the General Introduction. [21] Briefly, it is a member of the C-type lectin family and it recognizes sugars through its extracellular carbohydrate recognition domain (CRD) in a Ca^{2+}-dependent manner (Figure 5.1). [22,23] Its 3D structure is not available, but, at least, homology models of the soluble portion (CRD) can be obtained considering that it is closely related (high degree of sequence identity) with DC-SIGN and DC-SIGNR. [24] The biological function of this lectin has not exhaustively been studied; however, it is known that it regulates cell adhesion, cell signalling, and glycoprotein clearance through recognition of both self and non-self glycoconjugates. [25] It has also been proposed that LSECtin is an adhesion factor for viruses (Ebola, SARS, and Lassa, among others), while its level has been found to be upregulated in tumour-associated macrophages in breast cancer. [26–32] More recently, LSECtin has emerged as promotor of the adhesion, proliferation, and migration of gastric cancer cells. [33]

Regarding its binding preferences, apart from simple monosaccharides, such as Man, Fuc, Glc and GlcNAc, it preferentially binds to the terminal GlcNAcβ1-2Man epitope presented on N-glycans, with an affinity considerably higher than DC-SIGN and DC-SIGNR. [27,32] Nevertheless, besides the biomedical importance of this lectin, its binding partners and the fine details of its interaction with sugars remained poorly understood for years. [34] To fill this gap, our colleagues from CIC biomaGUNE employed the CRD of LSECtin to test its binding ability against a synthetic N-glycan

array of positional isomers. [2] The choice of building the array playing on asymmetry derives from the fact that most natural N-glycans structures present non-identical antennae, generating isomers and structural variations. [35] They expanded an existing version of glycan array (already proportionating a wide variety of structures) with a series of complex biantennary N-glycans (Figure 5.2) presenting Galβ1-4GlcNAc (LacNAc), GalNAcβ1-4GlcNAc (LDN) and GalNAcβ1-4[Fucα1-3]GlcNAc (LDNF) epitopes on either one or both branches. [2,36,37]

Cy3-labelled LSECtin was incubated on the glycan array slide and, after several washes, the fluorescence was measured and analysed. The results highlighted a clear preference of the lectin for asymmetric N-glycans decorated at the 6-branch over the positional isomers substituted at the 3-branch (Figure 5.2).

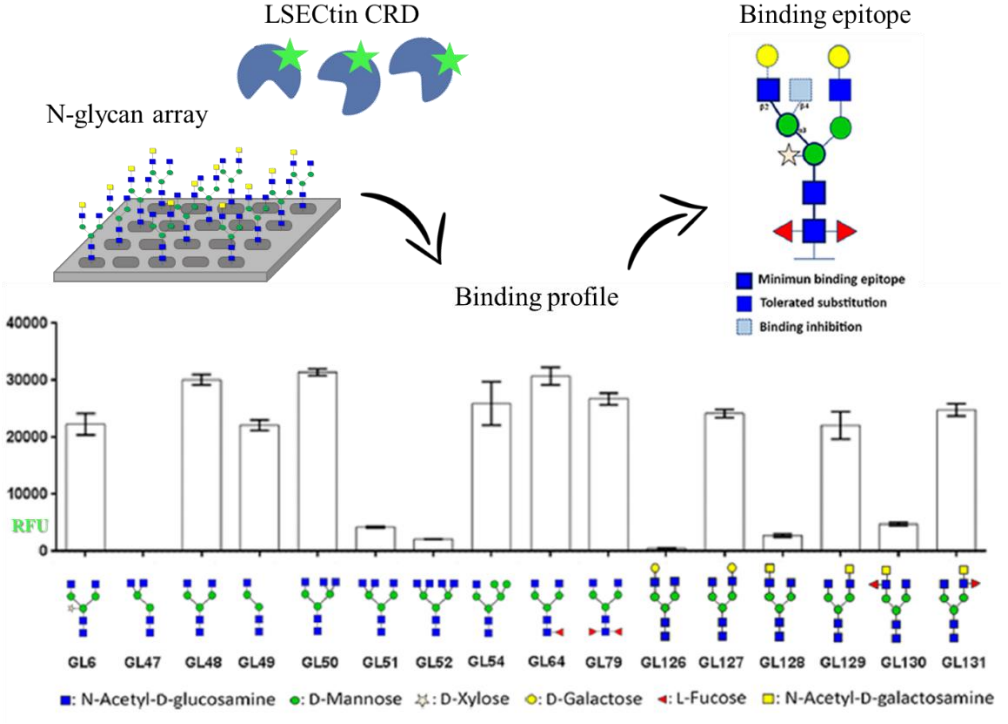


Figure 5.2. Binding profile and final binding epitope obtained from the microarray analysis with the Cy3 labelled CRD of LSECtin. Array data express in relative fluorescence units (RFU) and taken from Echeverria et al., 2018. [2]

Indeed, complex and hybrid type N-glycans with the terminal GlcNAcβ1-2Man epitope at the 3-branch were recognized by LSECtin, independently of the glycan composition at the 6-branch. In particular, complex type N-glycans with

GalNAc/Galβ1-4 capping at the 6-arm were bound by LSECtin, while the same decoration at the 3-arm abrogated the association event. This structure-dependent binding is a clear example of how subtle differences in the glycan structure may dictate significant changes in the recognition process. Moreover, the strong glycan specificity displayed by LSECtin for one individual positional isomer underlines the biological relevance of certain glycan motifs and suggests that they can be considered as key starting elements for the selective targeting of LSECtin.

Considering these outcomes, we focused our attention on a restricted panel of glycans, which were submitted to structural and binding analysis through NMR methodologies assisted by MD simulations. In particular, the selection included the minimum binding epitope (the disaccharide GlcNAcβ1-2Man), the non-elongated N-Glycan (**G0**), and two positional N-glycans isomers capped with a GalNAc residue either at the 3- (**LDN4**) or at the 6- arm (**LDN6**). The structures are reported in Figure 5.3.

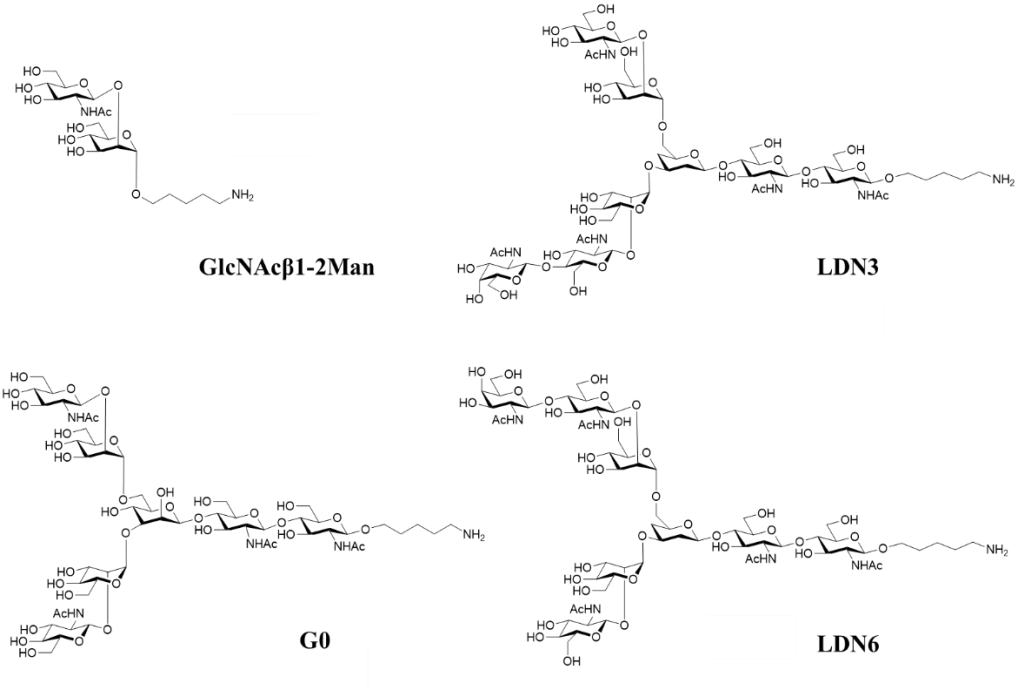


Figure 5.3. Representation of the glycans used in the NMR study, along with the nomenclature.

5.2 NMR studies in solution

A variety of NMR methodologies were employed to achieve the comprehensive elucidation of the motifs that provide the recognition specificities of the systems under study. [38] To this purpose, only the monomeric and soluble CRD (residues 162-293, 16 kDa) of LSECtin was used.

5.2.1 The interaction of LSECtin with the basic disaccharide: GlcNAcβ1-2Man

In a first step to characterize the binding event, the proposed minimum binding epitope recognized by LSECtin (the GlcNAcβ1-2Man disaccharide) was analysed.

A K_D of 3.5 μM has been previously estimated for the disaccharide-lectin association employing solid phase binding competition assays. Interestingly, this motif has emerged as the preferred one in the *termini* of N-glycans structures. [27] An affinity constant in the low micromolar range is considerably high for sugar-lectin interactions and especially for C-type lectins, which generally display affinities for their natural ligands in the high micromolar or millimolar ranges.

Considering the relatively high affinity, the analysis was started by **performing ^{1}H-NMR-based titrations:** the effect of the addition of increasing amounts of the ligand to a sample containing the lectin in its *apo* form (concentrated at 120 μM in deuterated buffer) was monitored. At each point of the titration (1:0, 1:1, 1:2, 1:5, 1:10, and 1:20 of protein:ligand ratios), a ^{1}H-NMR experiment was recorded. The processed spectra were subsequently analysed to detect, if any, chemical shift perturbations or linewidth variations in the protein's signals upon ligand binding. From the comparison of the ^{1}H-NMR spectra (Figure 5.4 and the Supporting Information), clear changes in the signals arising from the protein were visible upon addition of 1 or 2 equivalents of the disaccharide. Once exceeded the 2 equivalents, no more perturbations were detected, suggesting that the system reached saturation at low lectin:ligand ratios.

Moreover, in the NMR spectra recorded after the first two additions of the disaccharide (ratios 1:1 and 1:2), independent signals corresponding to the free and bound forms of the lectin were detected, suggesting that the system exhibits slow exchange in the chemical shift time scale; in accordance with the low affinity previously reported. [27]

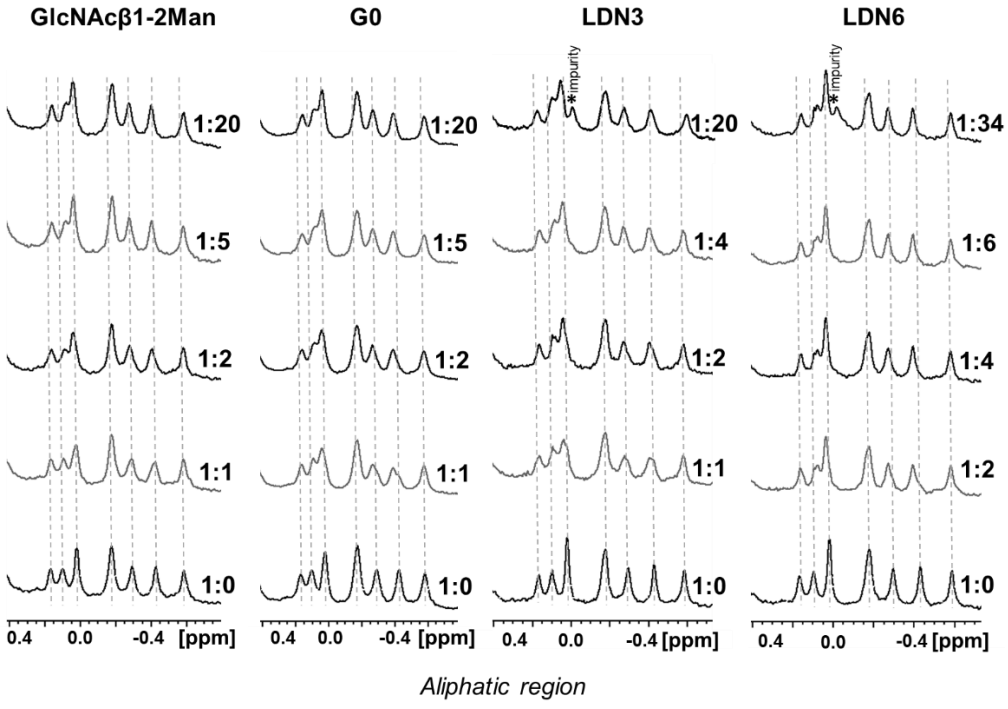


Figure 5.4. Stacked ^{1}H NMR spectra acquired for the titration of (from left to right) the disaccharide GlcNAcβ1-2Man, **G0**, **LDN3**, and **LDN6** to LSECtin. The relative amounts of equivalents are reported. Only the high field region (aliphatic protons) of the spectra is shown.

Consequently, Saturation Transfer Difference NMR **(STD-NMR) experiments** were performed to disclose the binding epitope. [39–41] After screening different experimental conditions, the final temperature was set of 310 K, while the disaccharide was employed with a molar excess of 70 equivalents. The STD-NMR experiment under aliphatic irradiation of the protein showed clear STD signals mostly arising from the GlcNAc ring (Figure 5.5). Indeed, H2, H6, H6' and H4/H5 (overlapped) of the GlcNAc moiety displayed unambiguous and strong STD intensities. The methyl group from the N-Acetyl (NAc) also showed STD, but relatively weaker when compared to the other signals from the same sugar. For the Man moiety, only H3 was detected. These data suggested that the main binding epitope for LSECtin is the GlcNAc ring of the disaccharide, but that also the mannose unit lays close to the protein surface.

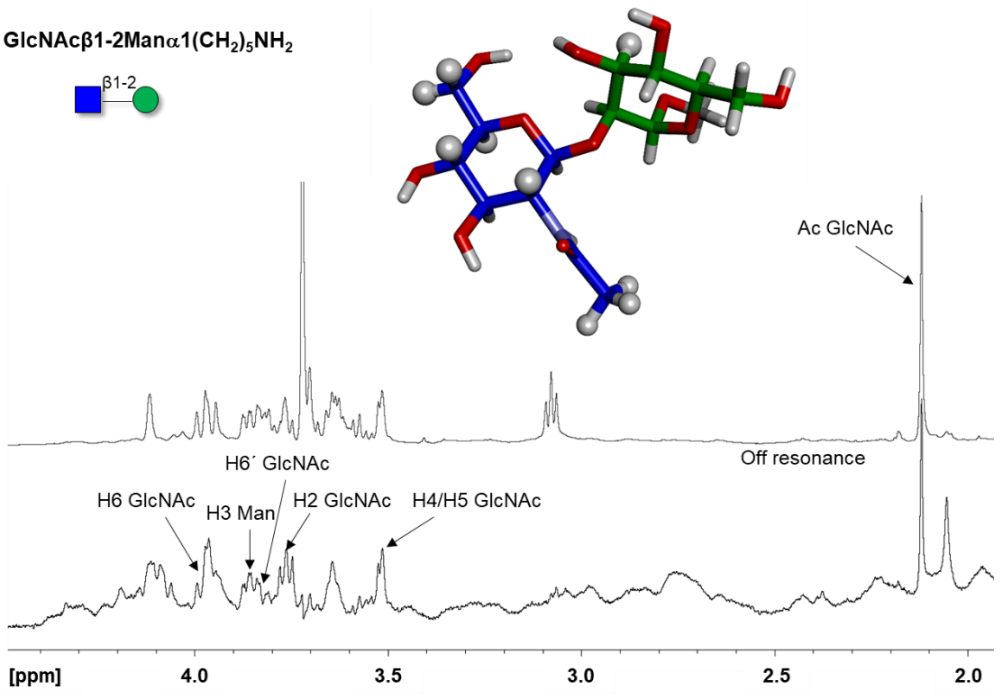


Figure 5.5. STD NMR experiment for the complex formed between LSECtin and the disaccharide GlcNAcβ1-2Man. Off-resonance spectrum (above) and STD spectrum (below) obtained with aliphatic irradiation at δ 0.83 ppm. The annotations of the main ^{1}H signals are reported below. The lectin:ligand molar ratio was 1:70 (LSECtin CRD concentration was 120 μM). The experiment was acquired with 2 seconds of saturation time and 5 seconds of relaxation delay at 310K. On the top: 3D representation of the disaccharide, where the protons that provided clear STD signals are highlighted with grey spheres.

Rotating-frame Overhauser SpectroscopY NMR (**ROESY-NMR**) experiments were then employed to detect chemical exchange cross-peaks of the ligand. In fact, if the chemical exchange is sufficiently slow, through ROESY-NMR, it is possible to detect the chemical shift perturbations suffered by the protons of the ligand in the bound state, thus providing additional information about the interacting ligand epitope. [42]

The 2D ROESY-NMR experiment acquired using a 1:10 LSECtin:disaccharide ratio displayed chemical exchange cross peaks (opposite sign to the ROE-derived peaks) for some protons of the ligand, both from the GlcNAc and the Man moieties (Figure 5.6). This fact further revealed that the process is in the slow exchange regime in the chemical shift time scale, as also suggested by the ^{1}H-NMR-based titration results.

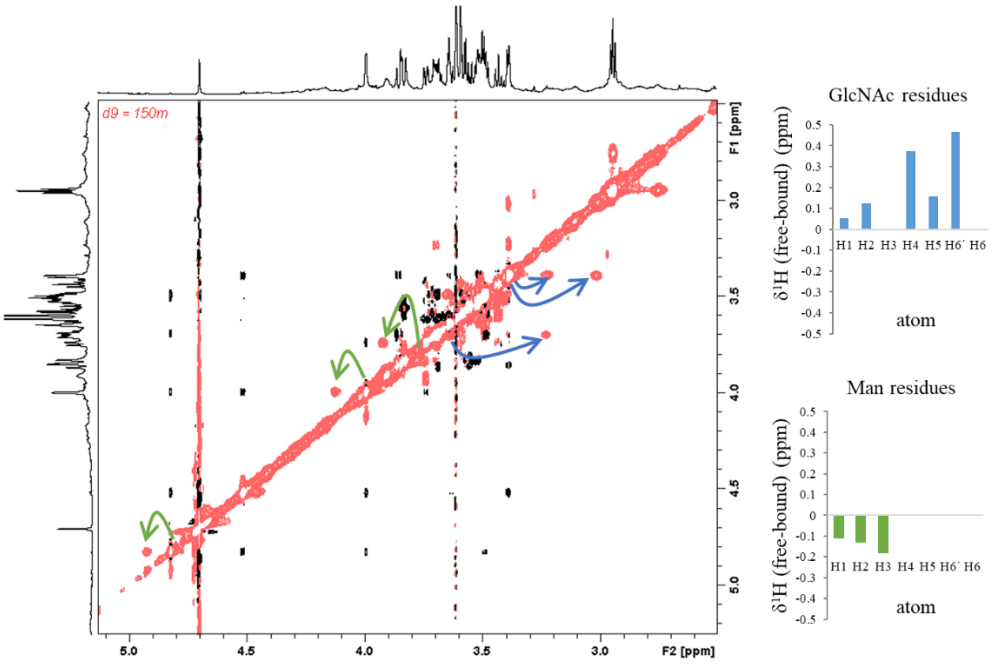


Figure 5.6. ROESY NMR experiment for LSECtin (120 µM) with the disaccharide GlcNAcβ1-2Man. The lectin:ligand molar ratio was 1:10 and the experiment was acquired with 150 ms of mixing time. Chemical exchange-mediated crosspeaks of the ligand protons are pointed with arrows (blue for the GlcNAc contributions and green for the Man protons). On the right, δ^1H differences (in ppm, free-bound) for each proton of the two sugar moieties

The alterations, in terms of chemical shift perturbations, suffered by the protons of a bound ligand are informative of the chemical environment in the bound form. In this system, most protons of the GlcNAc ring experienced a significant upfield shift (especially remarkable for H6´and H4), suggesting that these protons are in front of an aromatic ring of the lectin. [43] Fittingly, both H6' and H4 display axial-like orientation on the β-pyranose ring face. In contrast, Man H1, Man H2, and Man H3 suffered an opposite trend, being all downfield shifted (Figure 5.6). This evidence confirmed that not only the GalNAc but also the Man moiety is in close contact to the lectin surface, as also concluded by the STD NMR analysis. In addition, with these results it is tempting to speculate that a completely different chemical environment surrounds the two sugar (GlcNAc and Man) units when bound to LSECtin.

5.2.2 The interaction of LSECtin with the N-glycans

The recognition modes of LSECtin with the **G0**, **LDN3,** and **LDN6** N-glycans were then scrutinized, using the same NMR strategy described in the previous paragraphs.

The results of the **^{1}H-NMR-based titrations** are presented in Figure 5.4 and in the Supporting Information.

Focusing on the **non-elongated G0** (lectin:ligand molar ratios: 1:0, 1:1, 1:2, 1:5, and 1:20) perturbations in the lectin signals were detected only up to 2 equivalents of added ligand, in an almost identical fashion than those one observed for the disaccharide. This fact suggests similar binding affinities for the two molecules (Figure 5.4). Indeed, an affinity of 2.6 μM has been reported for the interaction of LSECtin with **G0** through solid phase binding competition assays. [27]

The attention was then paid to the **LDN3** and **LDN6 asymmetric N-glycans.** Surprisingly, the ^{1}H-NMR-based titration experiments for both systems shared the same pattern (Figure 5.4), indicating that no binding preferences were present in solution. Using 1:0, 1:1, 1:2, 1:4 and 1:20 of lectin:ligand molar ratios for the titrations with **LDN3,** and 1:0, 1:2, 1:4, 1:6 and 1:34 for those with **LDN6**, the saturation of the system was reached in both cases upon addition of 2-4 equivalents of the respective ligand. Remarkably, both titration profiles were also extremely similar to those obtained for the disaccharide and for **G0**. Gathering all this information, it can be deduced that all the ligands tested are able to bind to LSECtin in a similar manner. No recognition preferences of LSECtin for any asymmetric glycan was detected in solution with these experiments. Furthermore, the binding affinities of this C-type lectin for **G0**, **LDN3** and **LDN6** fall in the same range of affinity. Thus, a strong discrepancy emerged with respect to the results obtained when the glycans were immobilized on the array surface. [2]

Following the approach used with the disaccharide, **ROESY-NMR** experiments were performed also with **non-elongated G0** using a protein:ligand molar ratio of 1:10 (Supporting Information). The analysis of the detected exchange peaks, showed how they all belonged to the terminal Man and GlcNAc residues of both 3- and 6-arm.

This experiment, together with the one performed with the disaccharide, allowed the calculation of the kinetic exchange rates k_{on} and k_{off} (Supporting Information). In fact, once the exchange cross peaks were identified, the corresponding kinetic data were estimated with the EXSY analysis for the complexes LSECtin:disaccharide and

LSECtin:**G0**. [44,45] In particular, two 2D-NOESY (0 and 100 ms) were recorded for each sample and after integration of the cross- and diagonal-peaks, the measured intensities were analysed using the ExsyCalc software to deduce the pseudo-first-order exchange rates. The EXSY analysis yielded an estimation of k_{off} = 0.2 s^{-1} and 1 s^{-1} for **G0** and the disaccharide, respectively. Both k_{on} and k_{off} rates are faster for the disaccharide, evidence that leads to similar affinities, although the binding kinetics are different (Supporting Information).

Additional structural rationale for the binding of the N-glycans (**G0**, **LDN3** and **LDN6**) to LSECtin was achieved performing **STD-NMR experiments** (Figures 5.7, 5.8, 5.9 and 5.10). [39–41] The samples were prepared using 1:34 of lectin:ligand molar ratio and the experiment were performed at 310 K with both aliphatic (δ 0.6 ppm) and aromatic (δ 6.82 ppm) irradiation of the protein. The final STD spectra here presented are a sum (Σ) of the difference spectra obtained with the two irradiations. A long relaxation delay (15s) was set after initial optimization trials, in order to allow the complete relaxation of all the protons of the N-glycans. Indeed, we have experienced that the acetyl groups in these complex molecules relax very slowly and could give rise to false STD NMR signals. Thus, previously estimated the T1 relaxation times for these signals and then placed the relaxation delay value accordingly.
Although the overlapping of many signals of the ligands precluded a full quantitative analysis of the STD data, some isolated signals were the key for the definition of the binding epitope.

For the **non-elongated G0**, the two terminal GlcNAc (GlcNAc-3 and GlcNAc-6, see the nomenclature in Figure 5.7) residues provided the majority of the STDs signals, but some contributions also arose from the directly linked mannose units (Man-3 and Man-6). The general STD profile is comparable to the one obtained for the disaccharide, indicating that both sugars interact with LSECtin through the same epitope (Figure 5.7).

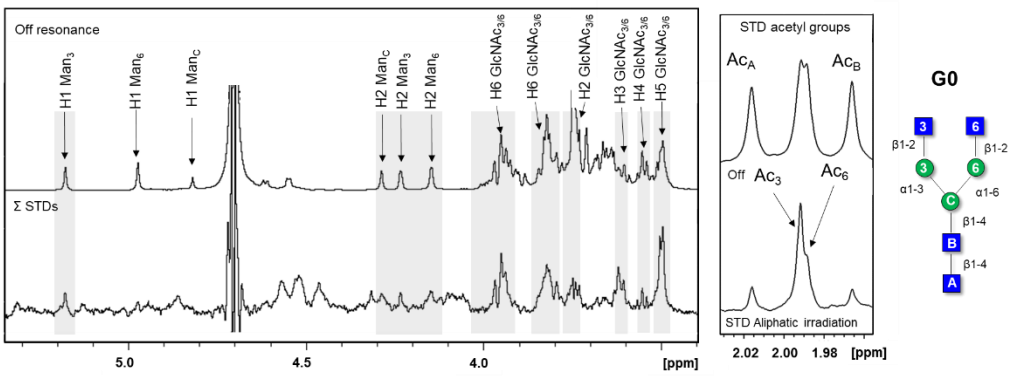


Figure 5.7. STD NMR experiments performed for the complex formed by LSECtin and **G0**. Left: off-resonance spectrum and the sum (Σ) of the STD NMR experiments carried out under aliphatic and aromatic irradiations (δ 0.6 ppm and δ 6.82 ppm, respectively). The main ^{1}H NMR signals are annotated in the off-resonance spectrum and the nomenclature used is reported in the SNGF representation of the glycan on the right. The lectin:ligand molar ratio was 1:34, with a LSECtin CRD concentration of 120 μM. The STD NMR experiments were acquired with 2 s of saturation time, 15 s of relaxation delay, and 2880 scans at 310 K. Right: expansion of the STD NMR spectra showing the acetyl group region of the GlcNAc moieties (only irradiation at the aliphatic region).

However, the protons of the two pyranose rings of the terminal GlcNAc are equivalent to NMR and could not be distinguished and assigned. Only the acetyl groups of these residues (Ac-GlcNAc-3 and -6), which display a minor difference in the ^{1}H chemical shift, have been reliably assigned through NOE correlations between them and the anomeric protons of the respective bound mannose. Both the terminal Ac-GlcNAc displayed STD signals, but the one Ac-GlcNAc-3 was twice as intense as Ac-GlcNAc-6 (Figure 5.7 and 5.9). Fittingly the Ac-GlcNAc of the inner core (A and B) did not display STD signals. Minor STD intensity were detected for the H1 and H2 protons of the Man-3 (located in the 3-arm) but not for the same protons of the Man-6 residue. To summarize the STD response for **G0**, it can be affirmed that both the terminal GlcNAcβ1-2Man epitopes are recognized by LSECtin with a detectable preference for the motif presented at the 3-branch.

Moving to the **asymmetric LDN3** and **LDN6**, the STD-NMR experiments were performed using the same experimental approach described above for **G0**. The global STD profiles were similar, involving both GlcNAc and Man terminal residues (Figure

5.8). The unambiguous differentiation of the key NMR signals belonging either to the 3- or to the 6-branch was fundamental to obtain a clear binding epitope.

For **LDN3**, the terminal GlcNAc H5 at the 6-arm provided STD intensity. Alternatively, for **LDN6,** the visible STD NMR signal belonged to the terminal GlcNAc H5 of the opposite branch (GlcNAc-3). Moreover, Man-6 H1 and H2 protons of **LDN3** displayed STD intensities, while the same proton at Man-3 moiety did not generate any signal. A completely opposite trend was observed for **LDN6**, where only Man-3 H1 and H2 protons were detected.

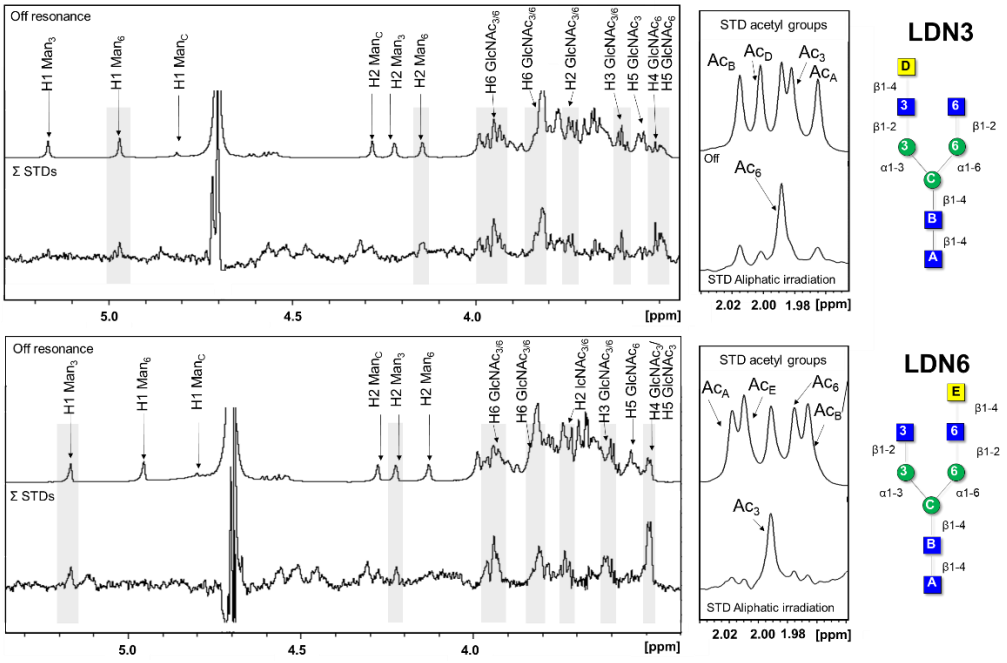


Figure 5.8. STD NMR experiments performed for the complexes formed by LSECtin and the asymmetric N-glycans (top: **LDN3**; bottom: **LND6**). Left: off-resonance spectrum and sum (Σ) of STD NMR experiments carried out under aliphatic and aromatic irradiations (δ 0.6 ppm and δ 6.82 ppm, respectively). The key ^{1}H NMR signals are annotated in the off-resonance spectrum and the nomenclature used is reported with SNGF representations of the glycans (right). The lectin:ligand molar ratio was 1:34, with a LSECtin CRD concentration of 120 μM. The STD NMR experiments were acquired with 2 s of saturation time, 15 s of relaxation delay, and 2880 scans at 310 K. Right: expansion of the STD NMR spectra showing the acetyl group regions of the GlcNAc/GalNAc moieties (only irradiation at the aliphatic region) for both complexes.

The analysis of N-acetyl groups of the ligands strongly contributed to defining the final picture of the interaction. In fact, the signals in the acetyl region of the spectra (δ 1.95-2.0 ppm) generated by GlcNAc-3, GlcNac-6, GlcNAc-A, GlcNAc-B and GalNAc-D/GalNAc-E did not display signal overlap and each contribution was successfully

assigned (Figure 5.9). For **LDN3**, only the acetyl group of GlcNAc-6 appeared in the STD NMR spectrum while contrarily, for **LDN6,** the STD intensities in the acetyl region were generated by GlcNAc-3 (Figure 5.9).

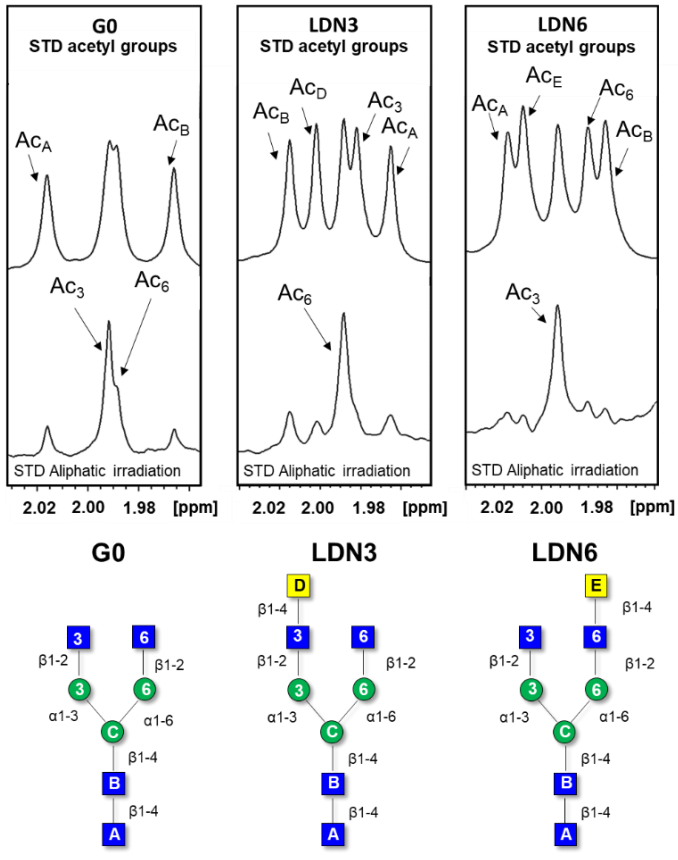


Figure 5.9. Acetyl region of the GlcNAc/GalNAc residues in the off resonance spectra (above) and in the STD NMR spectra under aliphatic irradiation (bottom). From left to right: the complexes of LSECtin with **G0**, **LDN3**, and **LDN6**, respectively. SDNF representation of the glycans with the used nomenclature are presented in the low panel of the figure. The experimental details of the experiments can be found in the captions of Figures 5.7 and 5.8.

Interestingly, the STD-NMR signals for the acetyl groups of **LDN3**, **LDN6**, and **G0** were always stronger under aliphatic irradiation, suggesting that the methyl groups are close to aliphatic residues at the binding site of the protein. This speculation was strengthened by the fact that no direct irradiation of the ligand was taking place during the experiments, since the on-resonance aliphatic frequency was set at δ 0.6 ppm, 1.3 ppm far from the signals of the acetyl groups, which can be considered as a safe irradiation frequency that avoids spurious saturation. Moreover, a direct ligand

irradiation would have affected homogeneously all the signals close to the irradiation frequency, thus providing a general STD enhancement of the whole acetyl region. Contrarily, the spectra displayed STD signals of specific N-acetyl contributions (Ac-3 and Ac-6 for **G0**, Ac-6 for **LDN3** and Ac-3 for **LDN6**), which additionally, fall in between the other acetyl groups (Ac-A and Ac-B). Consequently, we can safely exclude direct irradiation and assess that the acetyl groups of the GlcNAc moieties providing STD signals are close to aliphatic chains of the protein.

From the global STD results, it is evidenced that the terminal GalNAc residue acts a stop light, precluding the binding of LSECtin to that branch. In fact, in solution, the recognition epitope completely switched from the 6-branch for **LDN3** to the 3-branch for **LDN6**. For the non-elongated N-glycan **G0**, the epitope presented in both branches was recognized, with a slightly preference for the binding to the 3-arm (Figure 5.10).

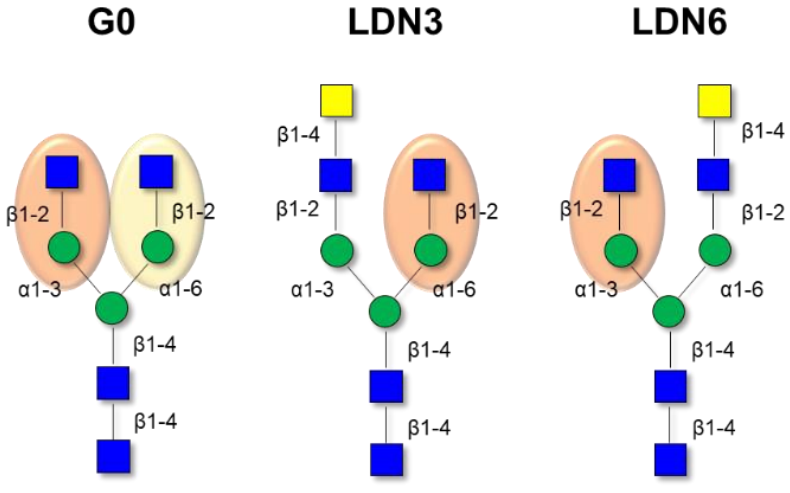


Figure 5.10. Binding epitopes of the N-glycans obtained by STD-NMR for the interaction with LSECtin. The three N-glycans employed in the study are represented, highlighting the key GlcNAcβ1-Man epitope of each. For **G0**, the 1,3-branch (highlighted in pale red) is better recognized than the 1,6-branch (highlighted in pale yellow), according to STD NMR results.

In conclusion, the binding features of the interaction of the N-glycans with LSECtin have been disclosed through NMR experiments in solution. The conclusions strongly differ from the observations using microarrays. [2] This evidence shed light on the importance of glycan presentation when interacting with a protein partner. In fact, **LDN3** is not capable to be recognized by LSECtin under the array conditions. In

contrast, in solution **LDN3** interacts with the lectin with similar affinity as its **LDN6** positional isomer.

A computational analysis work by Fadda and co-workers has proposed that the presence of GalNAc residues, at the terminal end of a **G0** N-glycan, triggers distinct conformational changes. [46] Specifically, they calculated that the N-galactosylation at the 6-arm generated a 25% increase of the population of back-folded conformers (75% in total) over the non-elongated N-glycan (50% in total), whereas this effect did not occur in the simulations conducted when capping took place at the 3-branch. Such findings suggest the existence of different presentations for **LDN6** and **LDN3**; however, they cannot explain why **LDN3** is not bound by LSECtin in microarrays.

5.3 MD simulations

Motivated by finding a rational explanation of the divergent results, MD calculations simulating array conditions were performed.

This task was intricate, since the physical properties of the array components are complex and often there are no available information about their exact characteristics. [15] For instance, the composition of the support and the distribution of the glycan conjugating groups (PEGn-NHS on Nexterion Slide H) cannot be quantitatively described due to their intrinsic heterogeneity. For this reason, it is necessary to include a degree of simplification for simulating such systems. [12]

The Nexterion Slide H22, which was employed in the glycan array experiments described in this Chapter, consists of a glass slide coated with a hydrophilic polymer. [2] The glycans were attached to this solid support through long, polar, and flexible linkers composed of ca. 45 ethylene glycol PEG units (PEG2K-NHS) and a 5-amino-pentyl linker at the reducing-end that provided irreversible binding through the formation of an amide bond with the NHS group. [2]

For the MD simulation, the Nexterion Slide H22 **surface** was modelled as a highly polar cellulose I-beta phase squared slab of 150 x 150 x 20 Å. The regular almost perpendicular arrangement of the hydroxymethyl groups from the Glc moieties composing the surface provided easily accessible anchoring points for the linkers.

The use of full-sized **linkers** would create a model computationally not affordable for an all-atom simulation and for that reason the linker length was reduced, while preserving its chemical nature (PEG). A range of linkers' sizes with lengths up to 40 Å was explored, for a total of five bCell-PEGn-G0 models generated: bCell-PEG0-G0, bCell-PEG2-G0, bCell-PEG6-G0, bCell-PEG8-G0, bCell-PEG10-G0.

The system was built with a single linker-**G0** N-glycan attached at the centre of the slab, and molecular dynamics (MD) simulations of 200 ns were performed to analyse the degree of accessibility of the GlcNAcβ1-2Man epitope at both arms 3 and 6.
The analysis was focused on the **bCell-PEG8-G0** model, for which 28 independent replicas of 60 ns MD simulations were run for a total of 1.7 accumulated μs.
The choice of performing replicas of short simulations instead of long calculations was dictated by:

- the large size of the solvated system which would require very long calculation times;
- the observed tendency of the glycan to quickly and irreversibly interact with the polar surface, particularly when short linkers were used. Short simulation times helped to simultaneously exploring a much wider conformational space.

As a representative example, for the simulation involving the 8-unit PEG linker (bCell-PEG8-G0 model), it could be observed that while the **3-arm** remained **fully exposed** to the solvent, the **6-arm** interacted with the polar surface in such a way that the GlcNAcβ1-Man epitope is partially **buried** (Figure 5.11). This tendency is expected to translate in a reduced availability for receptor binding. Indeed, there is an intrinsic unequal degree of flexibility of the two epitopes: the additional ω torsion angle at the α(1 → 6) linkage provides a superior mobility at the corresponding epitope over that at the α(1 → 3) linkage.

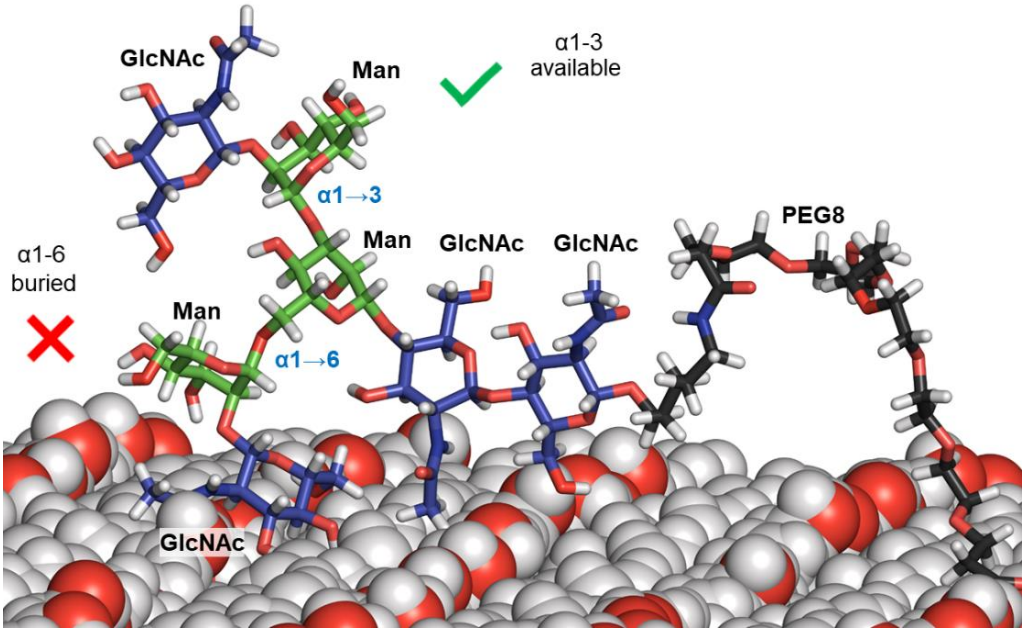


Figure 5.11. A selected molecular dynamics snapshot of the bCell-PEG8-G0 model showing the 6-arm of the GlcNAcβ1-2Man epitope partially buried and interacting with the support's surface. The same epitope located at the 3-arm remains fully accessible to the solvent.

Thus, the GlcNAc moiety at the α(1 → 6) arm, which explores a wider conformational space, holds a higher possibility of interacting with the support's surface. This aspect has been also proved by Solvent-Accessible Surface Area (SASA) analysis, that provided a statistically relevant difference between the propensity of each terminal GlcNAc unit to interact with the polar surface. Indeed, smaller SASA values, indicating a lower exposition, were calculated for the GlcNAc at the terminus of the 6-branch along the accumulated trajectory. On the other hand, the GlcNAc residue located at the 3-arm displayed higher SASA values, indicating that this moiety populates more exposed conformations during the simulation (Figure 5.12).

To conclude, the MD simulations predict that the established surface-glycan interactions makes the epitope located at the 6-arm to be partially buried with respect to that presented at the 3-branch, which remains fully accessible for interactions

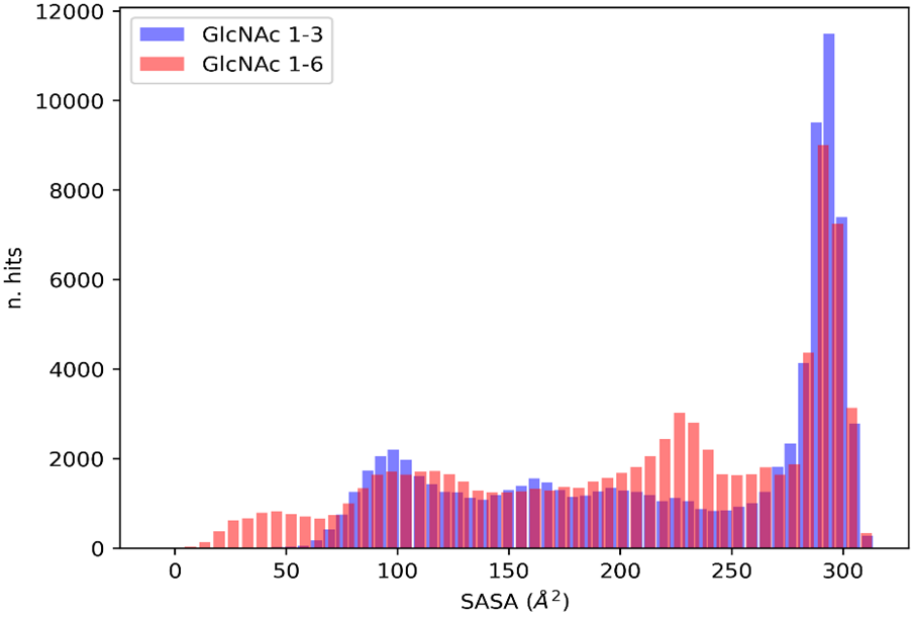


Figure 5.12. Distribution of SASA values for the terminal GlcNAc units at the α(1→3) (blue) and α(1→6) (red) arms of the bCell-PEG8-G0 model accumulated over 28 independent 60 ns MD simulations. SASA values are distributed asymmetrically among the two carbohydrates, with the less accessible α(1→6) arm showing higher population at low (0-50 Å) and medium (175-260) SASA values (buried) and the more accessible α(1→3) branch showing higher population at high SASA values (275-300 Å) (exposed).

5.4 Conclusions

The fine details of the recognition process taking place in solution between the C-type lectin LSECtin and a panel of N-glycans, including the non-elongated **G0** and two conformational isomers (**LDN3** and **LDN6**), has been unravelled by NMR.

Firstly, through STD-NMR experiments, it has been confirmed that the **minimum binding epitope** recognized by LSECtin is the GlcNAcβ1-Man disaccharide, presented at the terminal ends of N-glycans. Moreover, the atomic details of the interactions have been delineated, highlighting how the recognition prevalently involves the GlcNAc ring but also demonstrating that some protons of the Man moiety are also in close contact to the protein surface.

It has been described that, under array conditions, the GalNAc substitution in the 3-arm of biantennary N-glycans (**LDN3**) completely abrogated the binding to LSECtin, whereas the same substitution at the 6-arm (**LDN6**) was tolerated. [2] However, herein,

through ^{1}H-NMR-based titration and STD-NMR protocols, a **completely different outcome** has emerged. Strikingly, our analysis has demonstrated that, in solution, both asymmetric N-glycans are recognized by LSECtin in a similar fashion and with good affinities, also comparable to those obtained for the disaccharide and for the non-elongated N-glycans. On the other hand, in the absence of the GalNAc residue, the lectin recognizes both arms, with certain preference for the epitope on the α(1→3) arm. Since no preference for any of the two asymmetric N-glycans was detected in solution, this study has also then focused on understanding the reason behind such a marked difference observed with respect to the array data. Our initial hypothesises was that, on the array, the recognition was hindered due to intrinsic difference in the flexibility of the two arms, but more data were needed to confirm the theory.

MD calculations mimicking the array conditions (linker, surface and sugar) were then performed with the **G0** N-glycan, demonstrating that the α(1 → 6) arm of the sugar was interacting with the polar surface for the majority of the time of the trajectory. This observation could be considered as a direct hint of the fact that the epitope presented at the 6-arm was less prone to be accessible to the protein when the glycan was immobilized on a hydrophilic surface. The situation changes in solution, where in the presence of isotropic motion, both branches are nearly equally accessible and ready to interact with LSECtin.

The global results of this work have general consequence for the molecular recognition field, since they are the proof of how different outcomes can be obtained using diverse experimental approaches. Care should be taken when extracting conclusions from experiments conducted under specific conditions, because the molecular recognition details may differ from solution state to surfaces.

Under **immobilized settings**, the presentation of the interacting sugar epitope is essential, and therefore, the length and chemical nature of the linkers used to attach the ligands to surfaces and the composition of the solid support itself could also influence the final outcome and the interpretation of the obtained results.

On the contrary, when the analysis is taken **in solution**, the actors enjoy motional freedom that do not have in a natural environment.

Therefore, the adequate understanding of how rationalize the outcomes of each technique is of paramount importance.

Is it tempting to speculate which is the system closer to those existing in nature. We know that glycans are flexible structures usually exposed on cell surfaces as part of glycoconjugates forming the glycocalyx and, in this sense, the arrays at least provide a support mimicking the reality. [47] However, the architecture of cell glycocalyx is highly complex and **glycan presentation** is difficult to fit in a single, simple and flat surface presentation model. With our study we indeed proved that the immobilization on array surface can lead to false negative results because of changes in the presentation of the epitopes.

In conclusion, this compelling example points out the tremendous difficulty of translating *in vitro* results to the *in vivo* environment. Since a full understanding of sugar-lectin recognition events *in vivo* remains still elusive, a hot challenge in the field is to find **complementary systems** to study these interactions in a microenvironment closer to the real one, without losing the level of precision. [20] The use of liposomes, cell-based assays, or glycogenomic profiling are only some examples emerged in the last years and the continuous refinement of such techniques enables the study of glycan-protein interactions in a more natural context. An overview of promising results employing some of these emerging methodologies are discussed in Chapter 6.

5.5 References

1. Bertuzzi, S.; Peccati, F.; Serna, S.; Artschwager, R.; Notova, S.; Thépaut, M.; Jiménez-Osés, G.; Fieschi, F.; Reichardt, N.C.; Jiménez-Barbero, J.; et al. Immobilization of Biantennary N-Glycans Leads to Branch Specific Epitope Recognition by LSECtin. *ACS Cent. Sci.* **2022**, *8*, 1415–1423, doi:https://doi-org.unimib.idm.oclc.org/10.1021/acscentsci.2c00719.
2. Echeverria, B.; Serna, S.; Achilli, S.; Vivès, C.; Pham, J.; Thépaut, M.; Hokke, C.H.; Fieschi, F.; Reichardt, N.-C. Chemoenzymatic Synthesis of N-Glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-Type Lectin Receptors. *ACS Chem. Biol.* **2018**, *13*, 2269–2279, doi:10.1021/ACSCHEMBIO.8B00431.
3. Cummings, R.D. "Stuck on Sugars - How Carbohydrates Regulate Cell Adhesion, Recognition, and Signaling." *Glycoconj. J.* **2019**, *36*, 241–257, doi:10.1007/S10719-019-09876-0.
4. Meiers, J.; Siebs, E.; Zahorska, E.; Titz, A. Lectin Antagonists in Infection, Immunity, and Inflammation. *Curr. Opin. Chem. Biol.* **2019**, *53*, 51–67, doi:10.1016/J.CBPA.2019.07.005.
5. Cecioni, S.; Imberty, A.; Sébastien Vidal, S. Glycomimetics versus Multivalent Glycoconjugates for the Design of High Affinity Lectin Ligands. **2014**, doi:10.1021/cr500303t.
6. Kim, Y.; Young Hyun, J.; Shin, I. Multivalent Glycans for Biological and Biomedical Applications. *Chem. Soc. Rev* **2021**, *50*, 10567, doi:10.1039/d0cs01606c.
7. Porkolab, V.; Pifferi, C.; Sutkeviciute, I.; Ordanini, S.; Taouai, M.; Thépaut, M.; Vivès, C.; Benazza, M.; Bernardi, A.; Renaudet, O.; et al. Development of C-Type Lectin-Oriented Surfaces for High Avidity Glycoconjugates: Towards Mimicking Multivalent Interactions on the Cell Surface. *Org. Biomol. Chem.* **2020**, *18*, 4763–4772, doi:10.1039/D0OB00781A.
8. Ordanini, S.; Varga, N.; Porkolab, V.; Thépaut, M.; Belvisi, L.; Bertaglia, A.; Palmioli, A.; Berzi, A.; Trabattoni, D.; Clerici, M.; et al. Designing Nanomolar Antagonists of DC-SIGN-Mediated HIV Infection: Ligand Presentation Using Molecular Rods. *Chem. Commun.* **2015**, *51*, 3816–3819, doi:10.1039/C4CC09709B.
9. Won, S.; Richards, S.J.; Walker, M.; Gibson, M.I. Externally Controllable Glycan Presentation on Nanoparticle Surfaces to Modulate Lectin Recognition. *Nanoscale Horizons* **2017**, *2*, 106–109, doi:10.1039/C6NH00202A.
10. Wormald, M.R.; Dwek, R.A. Glycoproteins: Glycan Presentation and Protein-Fold Stability. *Structure* **1999**, *7*, R155–R160, doi:10.1016/S0969-2126(99)80095-1.
11. Srivastava, A.D.; Unione, L.; Bunyatov, M.; Gagarinov, I.A.; Delgado, S.; Abrescia, N.G.A.; Ardá, A.; Boons, G. Chemoenzymatic Synthesis of Complex N -Glycans of the Parasite S. Mansoni to Examine the Importance of Epitope Presentation on DC-SIGN Recognition . *Angew. Chemie* **2021**, *133*, 19436–19445, doi:10.1002/ANGE.202105647.
12. Grant, O.C.; Smith, H.M.K.; Firsova, D.; Fadda, E.; Woods, R.J. Presentation, Presentation, Presentation! Molecular-Level Insight into Linker Effects on Glycan Array Screening Data. *Glycobiology* **2014**, *24*, 17–25, doi:10.1093/GLYCOB/CWT083.
13. Blixt, O.; Collins, B.E.; van den Nieuwenhof, I.M.; Crocker, P.R.; Paulson, J.C. Sialoside Specificity of the Siglec Family Assessed Using Novel Multivalent Probes. *J. Biol. Chem.* **2003**, *278*, 31007–31019, doi:10.1074/jbc.m304331200.
14. Alvarez, R.A.; Blixt, O. Identification of Ligand Specificities for Glycan-Binding Proteins Using Glycan Arrays. *Methods Enzymol.* **2006**, *415*, 292–310, doi:10.1016/S0076-6879(06)15018-1.

15. Gao, C.; Wei, M.; McKitrick, T.R.; McQuillan, A.M.; Heimburg-Molinaro, J.; Cummings, R.D. Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins. *Front. Chem.* **2019**, *7*, 833, doi:10.3389/FCHEM.2019.00833/BIBTEX.
16. Song, X.; Heimburg-Molinaro, J.; Cummings, R.D.; Smith, D.F. Chemistry of Natural Glycan Microarrays. *Curr. Opin. Chem. Biol.* **2014**, *18*, 70–77, doi:10.1016/J.CBPA.2014.01.001.
17. Li, L.; Guan, W.; Zhang, G.; Wu, Z.; Yu, H.; Chen, X.; Wang, P.G. Microarray Analyses of Closely Related Glycoforms Reveal Different Accessibilities of Glycan Determinants on N-Glycan Branches. *Glycobiology* **2020**, *30*, 334–345, doi:10.1093/GLYCOB/CWZ100.
18. Klamer, Z.; Staal, B.; Prudden, A.R.; Liu, L.; Smith, D.F.; Boons, G.J.; Haab, B. Mining High-Complexity Motifs in Glycans: A New Language to Uncover the Fine Specificities of Lectins and Glycosidases. *Anal. Chem.* **2017**, *89*, 12342–12350, doi:10.1021/ACS.ANALCHEM.7B04293/SUPPL_FILE/AC7B04293_SI_001.PDF.
19. Wang, Z.; Chinoy, Z.S.; Ambre, S.G.; Peng, W.; McBride, R.; De Vries, R.P.; Glushka, J.; Paulson, J.C.; Boons, G.J. A General Strategy for the Chemoenzymatic Synthesis of Asymmetrically Branched N-Glycans. *Science (80-.).* **2013**, *341*, 379–383, doi:10.1126/SCIENCE.1236231/SUPPL_FILE/WANG.SM.PDF.
20. Wisnovsky, S.; Bertozzi, C.R. Reading the Glyco-Code: New Approaches to Studying Protein–Carbohydrate Interactions. *Curr. Opin. Struct. Biol.* **2022**, *75*, doi:10.1016/J.SBI.2022.102395.
21. Liu, W.; Tang, L.; Zhang, G.; Wei, H.; Cui, Y.; Guo, L.; Gou, Z.; Chen, X.; Jiang, D.; Zhu, Y.; et al. Characterization of a Novel C-Type Lectin-like Gene, LSECtin: Demonstration of Carbohydrate Binding and Expression in Sinusoidal Endothelial Cells of Liver and Lymph Node. *J. Biol. Chem.* **2004**, *279*, 18748–18758, doi:10.1074/jbc.M311227200.
22. Dominguez-Soto, A.; Aragoneses-Fenoll, L.; Martin-Gayo, E.; Martinez-Prats, L.; Colmenares, M.; Naranjo-Gomez, M.; Borras, F.E.; Munoz, P.; Zubiaur, M.; Toribio, M.L.; et al. The DC-SIGN–Related Lectin LSECtin Mediates Antigen Capture and Pathogen Binding by Human Myeloid Cells. *Blood* **2007**, *109*, 5337–5345, doi:10.1182/BLOOD-2006-09-048058.
23. Gramberg, T.; Soilleux, E.; Fisch, T.; Lalor, P.F.; Hofmann, H.; Wheeldon, S.; Cotterill, A.; Wegele, A.; Winkler, T.; Adams, D.H.; et al. Interactions of LSECtin and DC-SIGN/DC-SIGNR with Viral Ligands: Differential PH Dependence, Internalization and Virion Binding. *Virology* **2008**, *373*, 189–201, doi:10.1016/J.VIROL.2007.11.001.
24. Zhang, F.; Ren, S.; Zuo, Y. DC-SIGN, DC-SIGNR and LSECtin: C-Type Lectins for Infection. *Int. Rev. Immunol.* **2014**, *33*, 54–66, doi:10.3109/08830185.2013.834897.
25. Domínguez-Soto, Á.; Aragoneses-Fenoll, L.; Gómez-Aguado, F.; Corcuera, M.T.; Clária, J.; García-Monzón, C.; Bustos, M.; Corbí, A.L. The Pathogen Receptor Liver and Lymph Node Sinusoidal Endotelial Cell C-Type Lectin Is Expressed in Human Kupffer Cells and Regulated by PU.1. *Hepatology* **2009**, *49*, 287–296, doi:10.1002/HEP.22678.
26. Gramberg, T.; Hofmann, H.; Möller, P.; Lalor, P.F.; Marzi, A.; Geier, M.; Krumbiegel, M.; Winkler, T.; Kirchhoff, F.; Adams, D.H.; et al. LSECtin Interacts with Filovirus Glycoproteins and the Spike Protein of SARS Coronavirus. *Virology* **2005**, *340*, 224–236, doi:10.1016/J.VIROL.2005.06.026.
27. Powlesland, A.S.; Fisch, T.; Taylor, M.E.; Smith, D.F.; Tissot, B.; Dell, A.; Pöhlmann, S.; Drickamer, K. A Novel Mechanism for LSECtin Binding to Ebola Virus Surface Glycoprotein through Truncated Glycans *. *J. Biol. Chem.* **2008**, *283*, 593–602, doi:10.1074/JBC.M706292200.
28. Pipirou, Z.; Powlesland, A.S.; Steffen, I.; Pöhlmann, S.; Taylor, M.E.; Drickamer, K. Mouse LSECtin as a Model for a Human Ebola Virus Receptor. *Glycobiology* **2011**,

21, 806–812, doi:10.1093/GLYCOB/CWR008.
29. Shimojima, M.; Stroher, U.; Ebihara, H.; Feldmann, H.; Kawaoka, Y. Identification of Cell Surface Molecules Involved in Dystroglycan-Independent Lassa Virus Cell Entry. *J. Virol.* **2012**, *86*, 2067–2078, doi:10.1128/jvi.06451-11.
30. Shimojima, M.; Kawaoka, Y. Cell Surface Molecules Involved in Infection Mediated by Lymphocytic Choriomeningitis Virus Glycoprotein. *J. Vet. Med. Sci.* **2012**, *74*, 1363–1366, doi:10.1292/JVMS.12-0176.
31. Hoffmann, D.; Mereiter, S.; Jin Oh, Y.; Monteil, V.; Elder, E.; Zhu, R.; Canena, D.; Hain, L.; Laurent, E.; Grünwald-Gruber, C.; et al. Identification of Lectin Receptors for Conserved SARS-CoV-2 Glycosylation Sites. *EMBO J.* **2021**, *40*, doi:10.15252/EMBJ.2021108375.
32. Liu, D.; Lu, Q.; Wang, X.; Wang, J.; Lu, N.; Jiang, Z.; Hao, X.; Li, J.; Liu, J.; Cao, P.; et al. LSECtin on Tumor-Associated Macrophages Enhances Breast Cancer Stemness via Interaction with Its Receptor BTN3A3. *Cell Res. 2019 295* **2019**, *29*, 365–378, doi:10.1038/s41422-019-0155-6.
33. Zhang, Y.; Feng, Z.; Xu, Y.; Jiang, S.; Zhang, Q.; Zhang, Z.; Wang, K.; Li, X.; Xu, L.; Yuan, M.; et al. Novel Roles of LSECtin in Gastric Cancer Cell Adhesion, Migration, Invasion, and Lymphatic Metastasis. *Cell Death Dis. 2022 137* **2022**, *13*, 1–11, doi:10.1038/s41419-022-05026-x.
34. Valverde, P.; Martínez, J.D.; Cañada, F.J.; Ardá, A.; Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* **2020**, *21*, 2999–3025, doi:10.1002/CBIC.202000238.
35. Cummings, R.D. The Repertoire of Glycan Determinants in the Human Glycome. *Mol. Biosyst.* **2009**, *5*, 1087–1104, doi:10.1039/B907931A.
36. Yang, Y.Y.M.; Li, X.H.; Brzezicka, K.; Reichardt, N.C.; Wilson, R.A.; van Diepen, A.; Hokke, C.H. Specific Anti-Glycan Antibodies Are Sustained during and after Parasite Clearance in Schistosoma Japonicum-Infected Rhesus Macaques. *PLoS Negl. Trop. Dis.* **2017**, *11*, e0005339, doi:10.1371/JOURNAL.PNTD.0005339.
37. Brzezicka, K.; Echeverria, B.; Serna, S.; Van Diepen, A.; Hokke, C.H.; Reichardt, N.C. Synthesis and Microarray-Assisted Binding Studies of Core Xylose and Fucose Containing N-Glycans. *ACS Chem. Biol.* **2015**, *10*, 1290–1302, doi:10.1021/CB501023U/SUPPL_FILE/CB501023U_SI_001.PDF.
38. Ardá, A.; Jiménez-Barbero, J. The Recognition of Glycans by Protein Receptors. Insights from NMR Spectroscopy. *Chem. Commun.* **2018**, *54*, 4761–4769, doi:10.1039/C8CC01444B.
39. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788, doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.
40. Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, doi:10.1021/ja0100120.
41. Viegas, A.; Manso, J.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88*, 990–994, doi:10.1021/ed101169t.
42. Platzer, G.; Mayer, M.; Beier, A.; Brüschweiler, S.; Fuchs, J.E.; Engelhardt, H.; Geist, L.; Bader, G.; Schörghuber, J.; Lichtenecker, R.; et al. PI by NMR: Probing CH–π Interactions in Protein–Ligand Complexes by NMR Spectroscopy. *Angew. Chemie Int. Ed.* **2020**, *59*, 14861–14868, doi:10.1002/ANIE.202003732.
43. Asensio, J.L.; Ardá, A.; Cañada, F.J.; Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Acc. Chem. Res.* **2013**, *46*, 946–954, doi:10.1021/AR300024D/ASSET/IMAGES/MEDIUM/AR-2012-00024D_0006.GIF.
44. Perrin, C.L.; Dwyer, T.J. Application of Two-Dimensional NMR to Kinetics of Chemical Exchange. *Chem. Rev.* **2002**, *90*, 935–967, doi:10.1021/CR00104A002.

45. Latham, M.; Zimmermann, G.; Pardi, A. NMR Chemical Exchange as a Probe for Ligand-Binding Kinetics in a Theophylline-Binding RNA Aptamer. *J. Am. Chem. Soc.* **2009**, *131*, 5052–5053, doi:10.1021/JA900695M.
46. Harbison, A.M.; Brosnan, L.P.; Fenlon, K.; Fadda, E. Sequence-to-Structure Dependence of Isolated IgG Fc Complex Biantennary N-Glycans: A Molecular Dynamics Study. *Glycobiology* **2019**, *29*, 94–103, doi:10.1093/GLYCOB/CWY097.
47. Möckl, L.; Pedram, K.; Roy, A.R.; Krishnan, V.; Gustavsson, A.K.; Dorigo, O.; Bertozzi, C.R.; Moerner, W.E. Quantitative Super-Resolution Microscopy of the Mammalian Glycocalyx. *Dev. Cell* **2019**, *50*, 57-72.e6, doi:10.1016/J.DEVCEL.2019.04.035.

CHAPTER 6

NEW AVENUES FOR THE STUDY OF LECTIN-GLYCAN INTERACTIONS

Towards the observation of *in-cell* and *on-cell* binding events

6.1. Introduction

In the previous Chapters, the importance of aspects such as conformational motions, multivalency and presentation, involved in the recognition of glycans by lectins, has been highlighted. In all these cases, the experiments were conducted under *in vitro* conditions. Therefore, a full understanding of the features that regulate sugar-lectin interactions *in vivo* remains elusive. The standard reductionistic approach employed in chemistry labs has to be complemented with experiments towards the natural-like context, trying to analyse the lectin-glycan interaction events in systems related to the real biological environment. In the light of this, our most promising results obtained adopting *in-cell* and *on-cell* strategies is presented in this chapter (Figure 6.1).

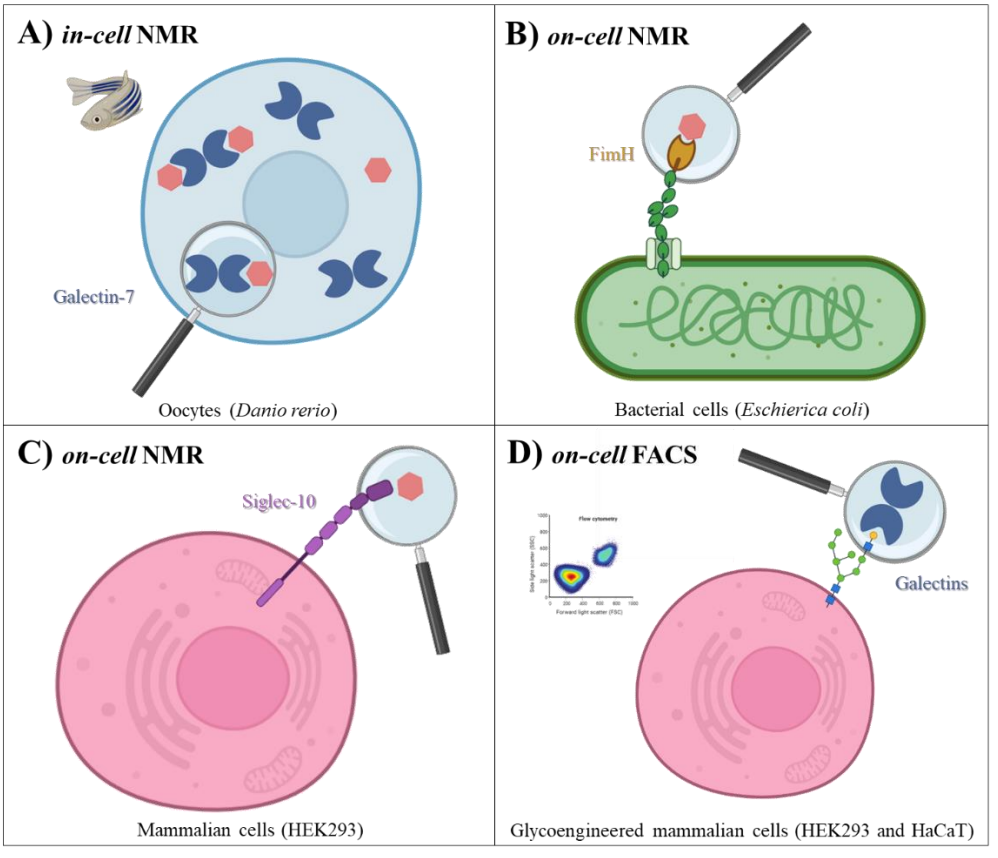


Figure 6.1. Graphic overview of the systems used to detect lectin-sugar binding in cellular contexts. A) *In-cell* NMR study of the interaction of galectin-7 (microinjected into zebrafish oocytes) with **TDG**, chapter 6.2. B) *On-cell* NMR study of the interaction of FimH (present on *E. coli* pili) with glycodendrimers, chapter 6.3. C) *On-cell* NMR study of the interaction of siglec-10 (overexpressed in HEK cells) with **F9** glycomimetic, chapter 6.4. D) *On-cell* FACS analysis of galectins' binding to glycoengineered HEK293 and HaCaT cell lines, chapter 6.5.

Specifically, an NMR approach employing zebrafish (*Danio rerio*) oocytes to detect the binding of galectin-7 to thiodigalactoside (**TDG**) directly *in-cell* is discussed in Chapter 6.2 (Figure 6.1, A).

Then, two similar *on-cell* NMR strategies using bacterial (Chapter 6.3) as well as mammalian (Chapter 6.4) cells to study the sugar recognition accomplished by exposed receptors are presented (Figure 6.1, B and C). In the first case, the protein target was FimH, a mannose-binding protein naturally expressed at the apical end of pili of uropathogenic strain of *E. coli*'s. In this case, its binding with dendrimers has been scrutinized in a cellular environment (Chapter 6.3). On the other hand, HEK293 cells were manipulated in order to overexpress the full-length receptor siglec-10 and its recognition with a selective glycomimetic (**F9**) was analysed *on-cell* (Chapter 6.4).

Finally, motivated by the recent and rapid developments in the field of glycoengineering, some preliminary results of glycogenomic modification on mammalian cells (HEK293 and HaCaT) as well as their employment in FACS-based assays to study the binding of exogenous galectins *on-cell*, are described in Chapter 6.5 (Figure 6.1, D). This last work was performed during a two-months secondment at the Copenhagen Center for Glycomics (University of Copenhagen, Denmark).

6.2 *In-cell* NMR using *Danio rerio* (zebrafish) oocytes

The results presented in this sub-chapter describe the setup and initial results of an *in-cell* NMR method for the detection of galectin-7 and its interactions within zebrafish (*Danio rerio*) oocytes.

This project is still ongoing and it was started by a former postdoc at the lab (Marta Gutierrez-Lete). It is also based on the collaboration with Biobide (*Bionaturis* Group, San Sebastian, Spain), which provided us the oocyte samples and is in charge of the microinjection procedure.

6.2.1 General context

The development and optimization of in solution *in-cell* NMR protocols have allowed to characterize the conformation and dynamics of biological macromolecules inside living cells. [1–5] Various examples, employing different cell types (bacterial, yeast, or mammalian cells and oocytes), as well as different strategies to deliver the protein inside the cell (direct protein expression, microinjection, CCP delivery, electroporation or pore-forming toxins) and different labellings, have been proposed. [6–16]

However, due to different causes, the application of this technique remains far from trivial. In fact, there are some fundamental aspects to be considered that may hamper the final success of the *in-cell* experiment:

- High concentrations of protein need to be used (or overexpressed directly into the target cell). This condition may cause aggregation and/or unspecific binding effects. Apart from that, sometimes is not possible to obtain high concentrated batches from the protein expression and purification protocols. Nevertheless, the use of ultra-high magnetic field spectrometers can partially overcome this weak point. [16]
- Selective labelling of the protein is required (fluorescent tags, NMR spin-active isotopes).
- The tumbling of the target molecule is the key point. In the case of *in-cell* NMR, the intracellular tumbling is very different to that of the receptor purified in solution for two main reasons:
    - The intracellular media is a crowded environment much more complex than the most typical aqueous buffers used to characterize

proteins *in vitro.* The high viscosity affects the rotational diffusion correlation time, leading to an apparent increase of molecular weight.

  - Unspecific interactions with large cellular components, (other proteins, cytoskeleton components, or membranes) slow down the tumbling rate, the transverse relaxation time became rather fast and, as a result, the resonance signals are broadened or even disappear.

- Cell vitality needs to be maintained during the acquisition of the spectra and this point is translated into the necessity of recording the spectra in short acquisition times. Moreover, in many cases, the natural cell sedimentation can prejudge the final results. To solve this issue, the use of bioreactors or the employment of gels like agar or alginate as well as methylcellulose hydrogels have been proposed. [17–20]

With the purpose to compare the data collected for sugar-lectin interactions *in-vitro*, with those obtained using experimental conditions that mimic the dense cell atmosphere, the *in-cell* NMR methodology was employed and adapted to our system. However, it has to be emphasized that this is a challenging and difficult task, since lectins can interact with multiple glycosylated partners inside the cell, as well as participate in protein-protein interaction events. In fact, the majority of the works involving *in-cell* NMR experiments use target proteins with high binding selectivity and specificity (e.g. enzymes), which have likely much smaller chances to interact with the cellular components than a lectin, which is intrinsically promiscuous.

Initial attempts to detect ^{15}N labelled galectins into *E. coli* cells or mammalian (HEK293T adherent and CHO Lec8) cells were performed in our lab. In the first case, the exogenous protein was firstly purified and then inserted into the bacterial cells using electroporation protocols. In the second effort, the protein overexpression was obtained directly in the target mammalian cells through transfection and growth in labelled medium. However, both protocols failed: the pattern of ^{1}H-^{15}N cross peaks of different galectins in HSQC spectra was never detected.
From the analysis of the attempts with *E. coli*, paying attention to the results of the experiments carried out with the cell lysates, we concluded that the level of

internalization of the protein upon electroporation was too low to detect a proper NMR signal.

Though, in the case of the mammalian cells, we never achieved high protein localization in the cytosol upon induction of the overexpression (checked with western blot). For this reason, we concluded that some regulatory pathway intervened to prevent the localization of high concentrations of galectin in the cytoplasm, which are probably toxic to the cells. Furthermore, we could also exclude the presence of unspecific interactions of the galectin with large glycoconjugates present in the complex cellular milieu, since the UDP-Gal transporter is inactive in the employed CHO Lec8 cell lines, what results in the impossibility of adding galactose residues at the GlcNAc-terminating branches of complex N-glycans.

Hence, analysing the outcomes of the first tests, the strategy for the protein-delivery seemed to be the bottleneck of the methodology. Consequently, we decided to change it and move to the microinjection technology, which allows a precise control of the concentration of the sample. In particular, a selected dimeric lectin (galectin-7) has been microinjected into zebrafish (*Danio rerio*) oocytes in the absence and in presence of a non-hydrolyzable glycomimetic ligand: thiodigalactoside (**TDG**).

6.2.2. *Danio rerio* (zebrafish) oocytes as host cells

The zebrafish oocytes have been described as a promising cellular system for studying proteins in a more biologically relevant environment. Indeed, they have already been employed for *in-cell* NMR studies. [21,22]

The use of zebrafish oocytes in this particular context is justified by some favourable properties:

- ✓ The transparency of the oocytes (Figure 6.2). On the contrary, frog (*Xenopus laevis*) oocytes are opaque and limit the ability to determine the localization of the injected sample. Moreover, dead oocytes are bright white and they can be easily identified.
- ✓ The easy manipulation of the oocyte. The safe microinjection and also the transference into NMR tubes are not risky operations for the cell vitality.

- ✓ The fixed development. In fact, the oocytes are not fertilized and do not suffer changes due to the cell cycle and growth in development, minimizing the experimental heterogeneity.

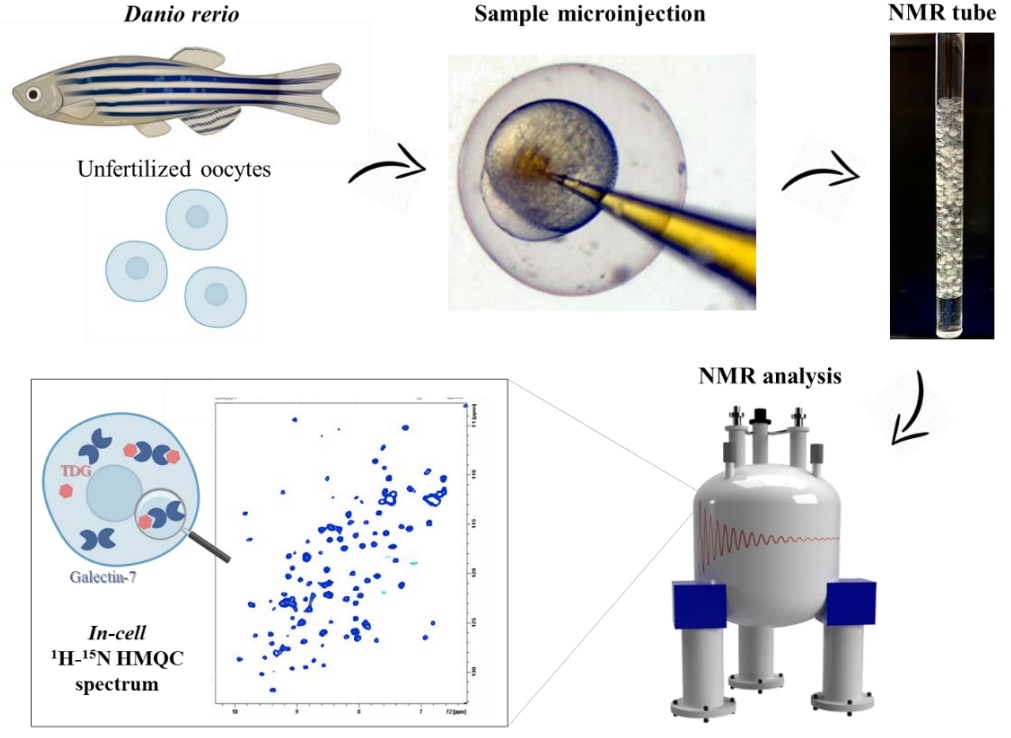


Figure 6.2. *In-cell* NMR procedure adopted to detect galectin-7 into zebrafish oocytes in its *apo* form and in the presence of **TDG**.

6.2.3 *In-cell* ^{1}H-^{15}N HMQC experiments

The samples of galectin-7 *apo* and galectin-7 with **TDG** (a non-hydrolizable lactose analogue, thanks to the presence of S-glycosidic linkage) were prepared, shipped to Biobide, and microinjected into the oocytes, as detailed in Chapter 7.

For each trial (two in total at the moment), two different samples were prepared:

- **The control**: ^{15}N labelled galectin-7 at a concentration of 2-2.5 mM. The protein concentration needed is considerably high. Galectin-7 has been chosen among other member of the galectin family because of its stability even at elevate concentrated conditions.

- **The sample**: ^{15}N labelled galectin-7 at a concentration of 2-2.5 mM in the presence of 60 molar equivalents of thiodigalactoside (**TDG**).

As evidenced in Figure 6.3, clear protein signals are detected in all the *in-cell* HMQC spectra. The protein fingerprint is the typical one obtained for galectin-7 in its folded form. [23] After the analysis, the extracellular media was scrutinized with the same 2D experiment and the resulting spectra displayed no traces of the labelled protein, thus confirming that the signals of the *in-cell* experiments arose form intracellular galectin-7 (Figure 6.3).

Moreover, a general trend can be deduced by simple observation of the spectra. In fact, in the presence of **TDG**, more protein cross peaks are detected.

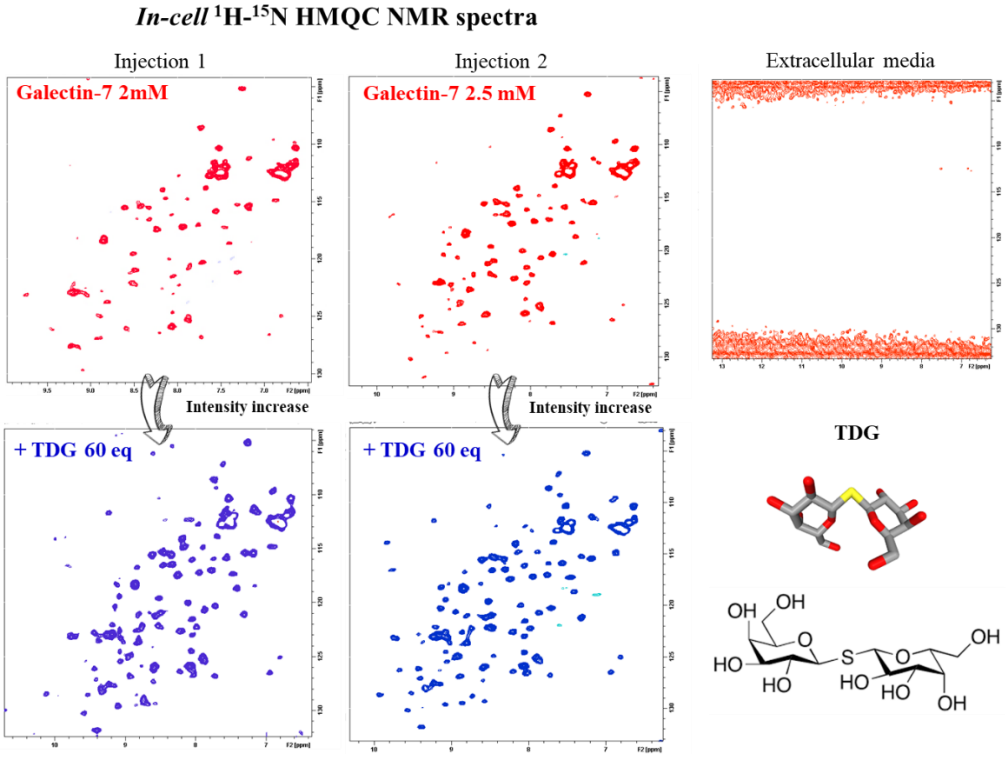


Figure 6.3. On the left: ^{1}H-^{15}N HMQC experiment *in-cell* acquired with oocytes microinjected with galectin-7 *apo* (red) and galectin-7 with **TDG** (blue). On the top right: ^{1}H-^{15}N HMQC experiment acquired for the extracellular media taken from the tube after the *in-cell* experiment. All the experiments were recorded with 64 scans, using the 800 MHz spectrometer and at 301K. On the bottom right: chemical structure and stick representation of the thiodigalactoside (**TDG**) lactose analogue.

A quantitative estimation of the percentage of the missing (or barely detectable) peaks in each spectrum was then performed.

Following the analysis of the **first injection experiments**, in the *in-cell apo* spectra, 20% (23 peaks) of the total protein peaks were missing, while in those spectra acquired at the presence of **TDG,** only 7% (8 peaks) of the peaks remained undetectable.

However, in the **second injection**, under an increased galectin concentration, only 10% of the peaks (12 signals) were absent in the *apo* spectra and almost all of them (only 3 cross peaks were missing) were recovered in the spectra acquired with the ligand.

The structural inspection of the missing peaks demonstrated that, in all cases, they were not localized in a specific site of galectin-7 monomer, but homogenously distributed throughout the β-sheets and the loops of the protein (Figure 6.4).

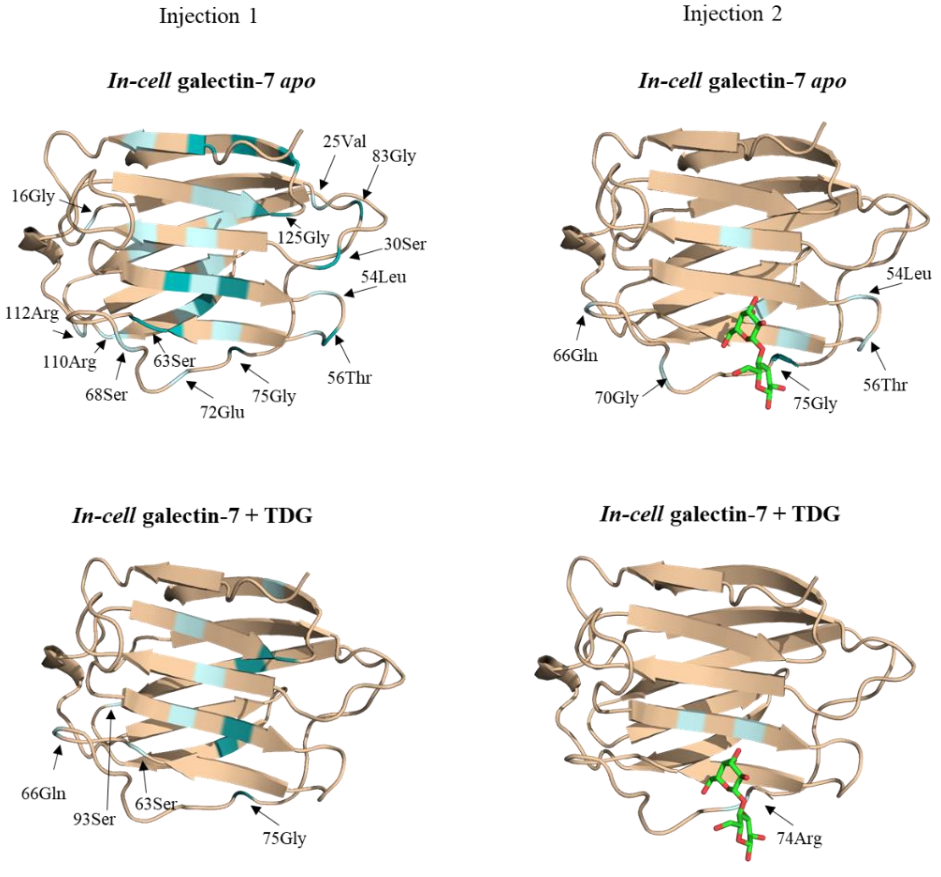


Figure 6.4. The missing peaks (dark blue) or those showing a drastic intensity decrease (light blue) detected by *in-cell* HMQC experiments plotted on the galectin-7 monomer (apo PDB 1BKZ, bound to lactose PDB 4GAL). The peaks at flexible loops are indicated with arrows.

The intensity analysis of the peaks was then carefully performed. To start with, the individual height of each detectable cross peak was analysed. The **height** values were always **higher** for the samples containing **TDG** than for the *apo* samples (confirming the trend deduced by simple visual inspection of the spectra). Furthermore, the individual heights were almost identical when the same type of sample (*apo* or with **TDG**) from different injections was compared. This observation can be considered as proof of the reliability of the set of experiments. The heights of the *on-cell* spectra were also compared with those obtained for galectin-7 *in-vitro* (dissolved in PBS) in its *apo* form (Figure 6.5). The values of the heights were normalized (plot B) by dividing the heights obtained *in-vitro* for a factor of 6.5 (based on the *in-vitro*/*in-cell* factor calculated for the most intense peak detected *in-vitro*). A huge decrease of the signal in the cellular conditions is evident. Interestingly, the peaks that in the *in-vitro* conditions displayed a moderate of low intensity, probably due to intrinsic dynamics in a particular time scale, matched with those lost in the oocyte sample.

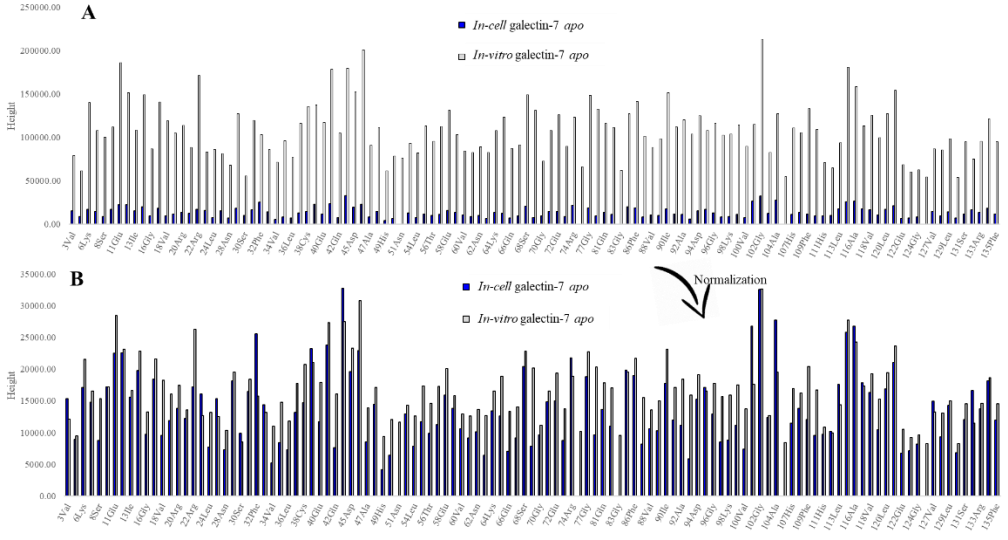


Figure 6.5. A) Height plot of the *on-cell* NMR HMQC spectrum of galectin-7 *apo* (injection 2) compared with that obtained for the same protein under *in-vitro* conditions. B) Same plot normalized. The heights obtained *in-vitro* were divided by a factor of 6.5. The factor (*in-cell*/*in-vitro* height fraction) measured for the most intense *in-vitro* peak.

Finally, from the analysis of the total height of the spectra, it can be observed that the intensity is higher for both experiments of the second injection, very likely due to the higher protein concentration used (Figure 6.6). More remarkably, the trend of intensity

enhancement when passing from the *apo* form to the samples containing **TDG** is conserved along the two different set of injections (Figure 6.6).

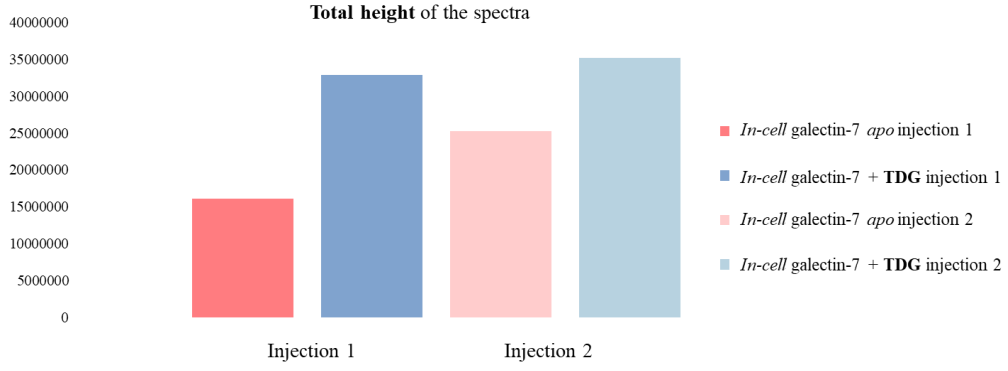


Figure 6.6. Total heights measured in the ^{1}H-^{15}N HMQC spectra recorded with *in-cell* conditions.

We hypothesized that the loss of the lectin cross-peaks is caused by the interaction of galectin-7 with nonspecific high molecular weight-binders inside the cell. The formation of supramolecular structures made some residues of the protein undetectable by NMR. The majority of the lectin cross-peaks was visible again after the addition of **TDG**. This suggested that the exogenous ligand competes with nonspecific glycan-containing binders dissociating the complexes and enabling the recovery of the signals.

As a future plan, more injections will be performed in order to enhance the repertory of data and increase the statistical significance of the outcome. In addition, experiments employing the purified galectins in the presence of crowding agents (such as Ficoll70) are currently ongoing, with the aim to compare these results with those obtained using oocytes. Besides, we will also strive to find a tailored methodology to identify the putative interactors of galectin-7 in the intracellular environment.

Although the experimental portfolio is not completed yet, the first *in-cell* NMR observation of a folded galectin has been herein achieved.

6.3 *On-cell* NMR for the identification of FimH ligands using the uropathogenic *E. coli* (UPEC) strain

The project discussed in this section described the development of an *on-cell* NMR method for the rapid screening of putative inhibitors of the bacterial adhesin FimH.

I was involved in this work during various stays at Milan, within the framework of the co-tutorship of my doctorate within the University of the Basque Country and the University of Milano-Bicocca. In fact, this section was directly supervised by my thesis co-director Professor Cristina Airoldi (https://en.unimib.it/cristina-airoldi, University of Milano-Bicocca, Dipartimento di Biotecnologie e Bioscienze, Italy) and developed in collaboration between her laboratory (bioOrg NMR lab) and the laboratory of Professor Alessandra Polissi (https://www.unimi.it/it/ugov/person/alessandra-polissi, University of Milano, Dipartimento di Scienze Farmacologiche e Biomolecolari, Italy). The results presented herein were successfully published in 2021. [24]

6.3.1 General context

The extensive use of antibiotics has raised a public health problem dictated by the antibiotic resistance of bacterial pathogens. [25] The uropathogenic *E. coli* (UPEC) strain is one of the main responsible factors of urinary tract infections. In this context, there is an increasing need to develop therapies to revert the bacterial resistance phenomenon. [26] UPEC is able to adhere and colonize host-tissues through the orchestrate actions of its adhesin proteins. [27,28] The type 1 pili are extracellular fibers exposed on UPEC's cell surface that present the FimH adhesin at their extremity, which is in charge of the bacterial binding to the bladder epithelium. [29–32] Specifically, FimH is a mannose-binding lectin capable of recognizing mannosylated proteins exposed at the host epithelium and its binding to these molecules facilitates the bacterial colonization. [33] Therefore, the selective targeting of FimH for antiadhesive purposes has been proposed as a possible avenue for developing therapeutics (Figure 6.7). [34–36] The rational design of new antiadhesive molecules able to inactivate the pathologic mechanism initiated by FimH is, therefore, an urgency, since no anti-virulence agents *versus* UPEC are currently under clinical development.

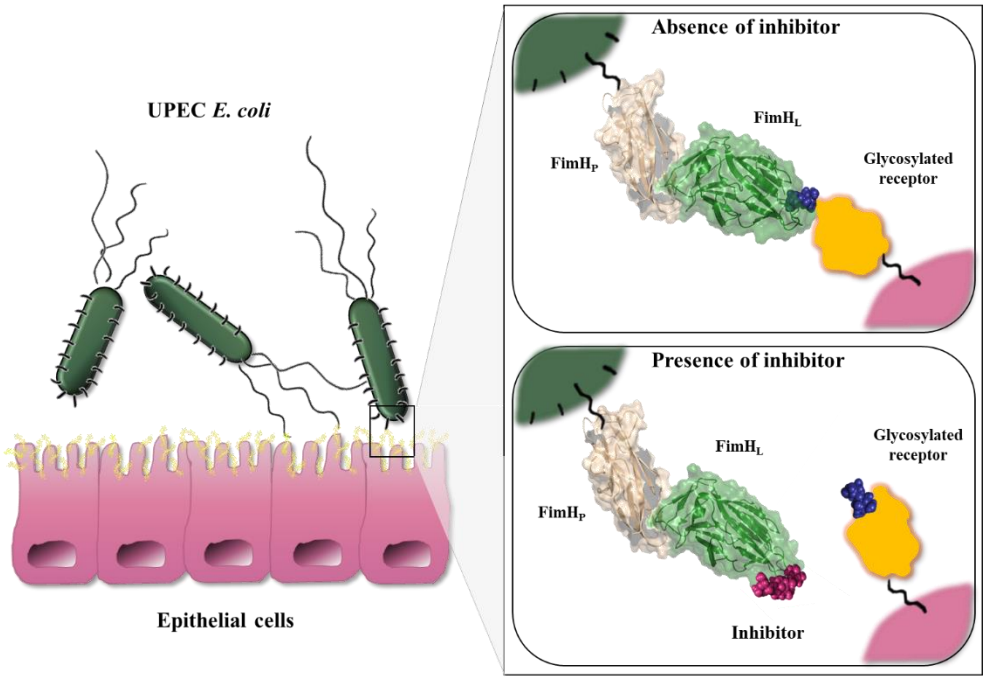


Figure 6.7. Representation of FimH-mediated adhesion of UPEC strain to host mammalian epithelial cells. Zoomed-in: FimH (PDB ID: 6GTZ) recognition of glycosylated receptors in pathological conditions (top) *versus* the inhibition of the adhesion process in the presence of a selective inhibitor (below).

Since the FimH affinity for monosaccharides is low, the multivalent presentation of mannose epitopes can be used to obtain high affinity ligands. The multivalent presentation has been previously exploited in the host laboratory to synthetize potential lectin inhibitors. [37–39] With this background, a variety of FimH potential inhibitors bearing a variable number (4, 6 or 18) of the key D-mannose epitope were generated, using pentaerythritol as core scaffold (Figure 6.8). The experimental details of the synthesis of the final mannosylated glycodendrimers (**Man$_4$**, **Man$_6$**, **Man$_{18}$**) and of the control molecule, which is decorated with galactose units (**Gal$_6$**) are discussed in the corresponding publication. [24]

These ligands were used to set up a NMR-based methodology for the rapid screening of potential FimH binders. The protocol is composed by:

- The preliminary test of putative inhibitors exploiting the ability of bacterial cells exposing FimH to agglutinate yeast cells.

- The *on-cell* STD-NMR experiments using intact and living cells. In this way, the receptor has not to be purified or labelled and the molecular recognition is studied by NMR under physiologically relevant conditions. This methodology has already been employed by us and other groups to detect on-cell recognition events at atomic level [24,39,40]

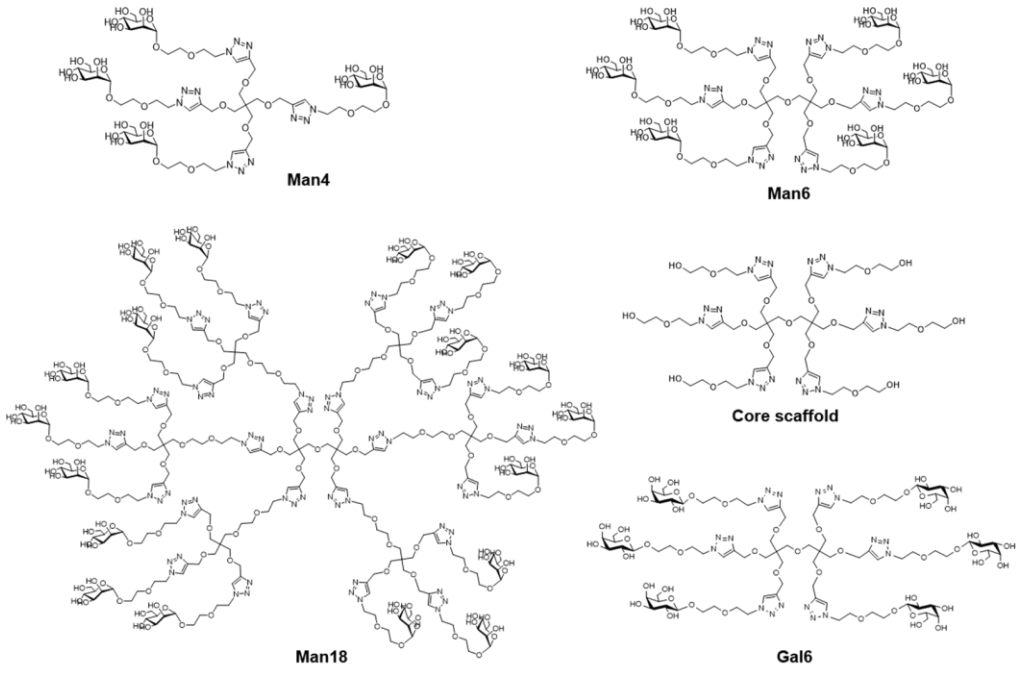


Figure 6.8. Structure of the glycodendrimers (**Man$_4$**, **Man$_6$**, **Man$_{18}$**), the core scaffold, and the control molecule (**Gal$_6$**) used in this study.

6.3.2 Uropathogenic *E. coli* CFT073 *ΔfimH* strain construction and initial screening of FimH inhibitors using yeast agglutination

The *E. coli* CFT073 (UPEC) strain, naturally overexpressing FimH, was selected as target cell line. Apart from the employment of the glycodendrimer control decorated with galactose units (**Gal$_6$**) and of the naked core scaffold, another important cell-negative control was developed. In fact, the uropathogenic *E. coli* CFT073 *ΔfimH* strain was obtained in the laboratory of Prof. Alessandra Polissi upon deletion of the chromosomal *fimH* gene, employing common techniques for bacterial gene engineering (details described in Chapter 7). This strain, lacking FimH adhesin, is not capable of binding the tailored mannosylated glycodendrimers.

Then, to test the ability of *E. coli* CFT073 and *E. coli* CFT073 *ΔfimH* to bind D-mannose residues through FimH, the yeast aggregation test was employed. [41–43] In fact, FimH is able to bind to mannans exposed on the yeast's surface, leading to the formation of cell clumps, called agglutinates, which are visible to the eye. Agglutinated yeast cells form a diffuse mat, whereas non-agglutinated yeast cells sediment and form a clear dot in the bottom of the well.

Hence, CFT073 cultures grown overnight were diluted to reach an OD_{600} value of 1 and suspended in PBS. Afterwards, they were incubated in a 96-well microtiter plate and the suspension was mixed with a fixed amount (10 mg/ml) of commercial *Saccharomiciae cerevisiae* (suspended in PBS). The ability of preventing or favouring the yeast agglutination was visually evaluated after 2 h of incubation, and the results are shown in Figure 6.9 From the visual inspection, it can be stated that, while in the presence of the wt CFT073 the agglutination is evident, the ability to form homogenous aggregation is lost with the *ΔfimH* mutants and clear dots of precipitation are visible at the bottom of the well. This observation can be translated into loss of FimH-yeast interaction.

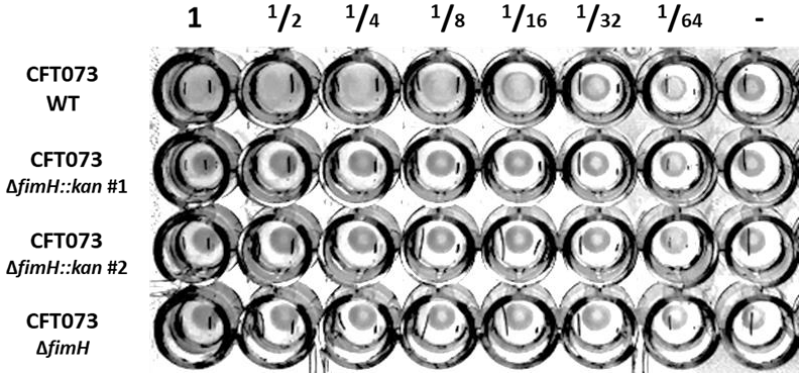


Figure 6.9. Yeast agglutination assay of *E. coli* CFT073 strain and of isogenic *ΔfimH* mutants. Three different *ΔfimH* were used, two of them bearing the kanamycin resistance cassette. The bacterial culture were incubated in serial two-fold dilutions of cultures starting from OD_{600} = 1.

Yeast agglutination assays were also employed for an initial screening of FimH inhibitors. In fact, the agglutination phenomenon can be inhibited by adding mannosylated molecules that compete with the yeast epitopes for the binding of FimH. The higher the affinity of the competitor for FimH, the stronger will be the reversion of the agglutination. The CFT073 strain was incubated with two-fold serial dilutions

of the glycodendrimers, as described above. The ability of the glycodendrimers to inhibit the aggregation of yeast cells mediated by CFT073 type I pili was evaluated visually after 2 h of incubation and the minimal concentration leading to negative aggregation was recorded as MIC. Serial dilutions of the core-scaffold and **Gal$_6$** were used as negative control. The **Man$_{18}$** glycodendrimer displayed the highest anti-aggregation activity, with an MIC value of 63 μM, four-fold and five-fold lower than those for **Man$_6$** (250 μM) and **Man$_4$** (500 μM), respectively. The core scaffold showed a MIC > 500 μM. Moreover, the incubation of CFT073 cells with high concentrations **Gal$_6$** did not display inhibitory activity in yeast agglutination assay, confirming that the interactions of the mannose units of the Man-derived glycodendrimers are specifically responsible for the inhibition of the agglutination process.

Apart from the visual inspection, a microscopy examination was also performed, using samples prepared with the same experimental conditions set for the yeast agglutination assay. **Gal$_6$** was again tested as negative control, showing no inhibitory activity in the yeast agglutination by CFT073 cells. Moreover, the BL21 (DE2) strain was also employed and displayed no agglutination after the incubation with yeast (Figure 6.10). On the contrary, both **Man$_6$** and **Man$_{18}$** clearly inhibited the agglutination process.

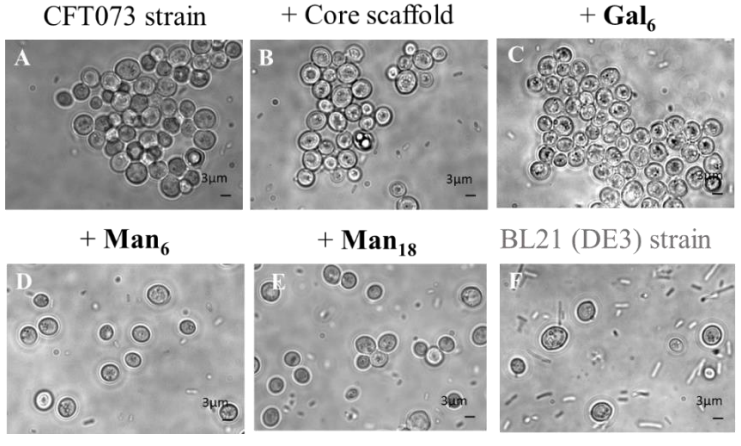


Figure 6.10. Yeast agglutination assay visualized by light microscopy (magnification, x 63, scale bar 3 μm.). A) *E. coli* CFT073 strain induces agglutination of *S. cerevisiae* in the absence (**A**) or presence of 500 μM core-scaffold (B) and 500 μM of **Gal$_6$** (C). Incubation of CFT073 with 500 μM **Man$_6$** (D) or 100 μM of **Man$_{18}$** (E) inhibit yeast agglutination. Incubation of CFT073 with *E. coli* BL21(DE3) strain does not induce agglutination of *S. cerevisiae* cells (**F**).

All together, these experimental results indicate a clear correlation among the number of sugar units exposed on the surface of glycodendrimers and their inhibitory efficacy, suggesting Man_{18} as a hit compound for the development of new antiadhesive molecules targeting FimH.

6.3.3 *On-cell* STD NMR binding studies

Man_6 and **Man_{18}** were then tested as FimH ligands employing *on-cell* STD NMR interaction studies (Figure 6.11).

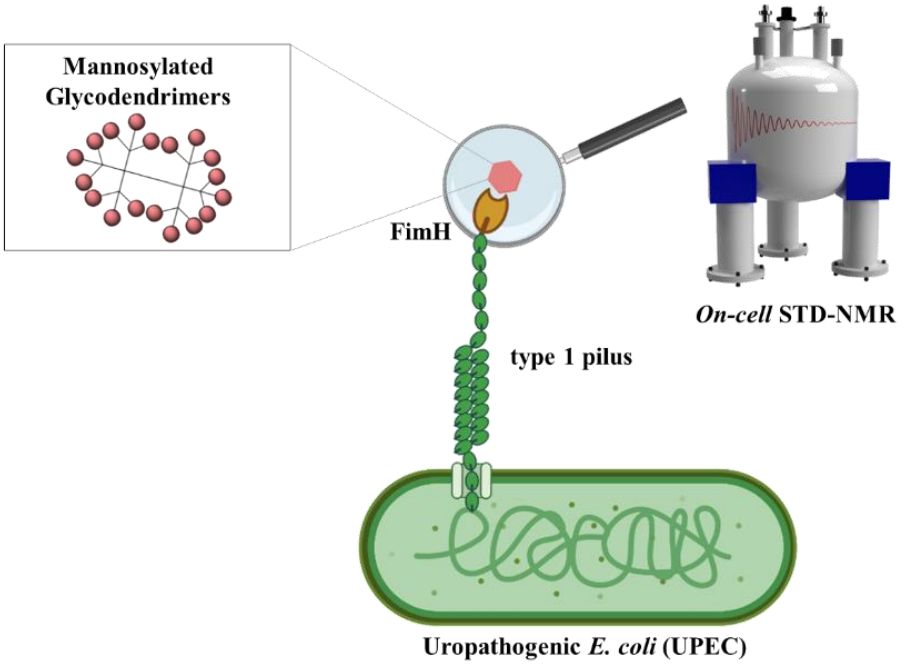


Figure 6.11. Representation of the strategy used to scrutinize FimH binding to hit compounds through *on-cell* NMR.

The experimental conditions were carefully fixed (playing with cells density, ligand concentrations, on-resonance saturation frequency and number of scans) to guarantee the success and reliability of the protocol. The final setup was established as follow:

- 2×10^9 cells were used for each NMR sample. Due to the small dimension of the bacterial cells, a high cell number could be used without cell aggregation and precipitation during acquisition.
- Ligands were employed at millimolar (or sub millimolar) concentration.
- The on-resonance irradiation frequency used was δ 0.0 ppm.

- A high signal-to-noise ratio was achieved with a low number of scans (256), what ensured a short acquisition time (30 minutes, short enough to assure cell integrity and viability).
- The spectra were acquired at 298K using 3 s of saturation time.

The final results of the *on-cell* STD experiments are reported in Figure 6.12 together with the epitope maps. Indeed, STD spectra acquired in presence of wt CFT073 cells and **Man$_6$** or **Man$_{18}$** showed that both D-mannose units and the dendrimer core-scaffold are involved in the interaction, since they display clear STD contributions.

In order to validate these results, several controls were performed. First, the binding of **Man$_6$** and **Man$_{18}$** was tested at the presence of the CFT073 Δ*fimH* strain. The loss of Man binding activity (resulting from the targeted deletion of *fimH* gene) acts as correct negative control, since no STD NMR signals were detected in this case (Figure 6.12)
In addition, *on-cell* STD experiments were performed with samples containing *E. coli* BL21 cells and, again, no binding to was detected. These data are also in accordance with the yeast agglutination assay results and can be explained by the fact that BL21 strain does not express pili when grown in rich media (Figure 6.12).
Moreover, STD NMR experiments acquired on-cell using the **core-scaffold** lacking sugar units as ligand displayed absence of interactions with the CFT073 strain (Figure 6.12).
With the support of all the controls performed, it is possible to state that the binding event observed for **Man$_6$** or **Man$_{18}$** on CFT073 cells was:

- Mediated by FimH (demonstrated through with the experiments using CFT073 Δ*fimH* and BL21 strains)
- Dependent on the D-mannose moieties exposed in the ligand structure (demonstrated with the experiments using the core scaffold)

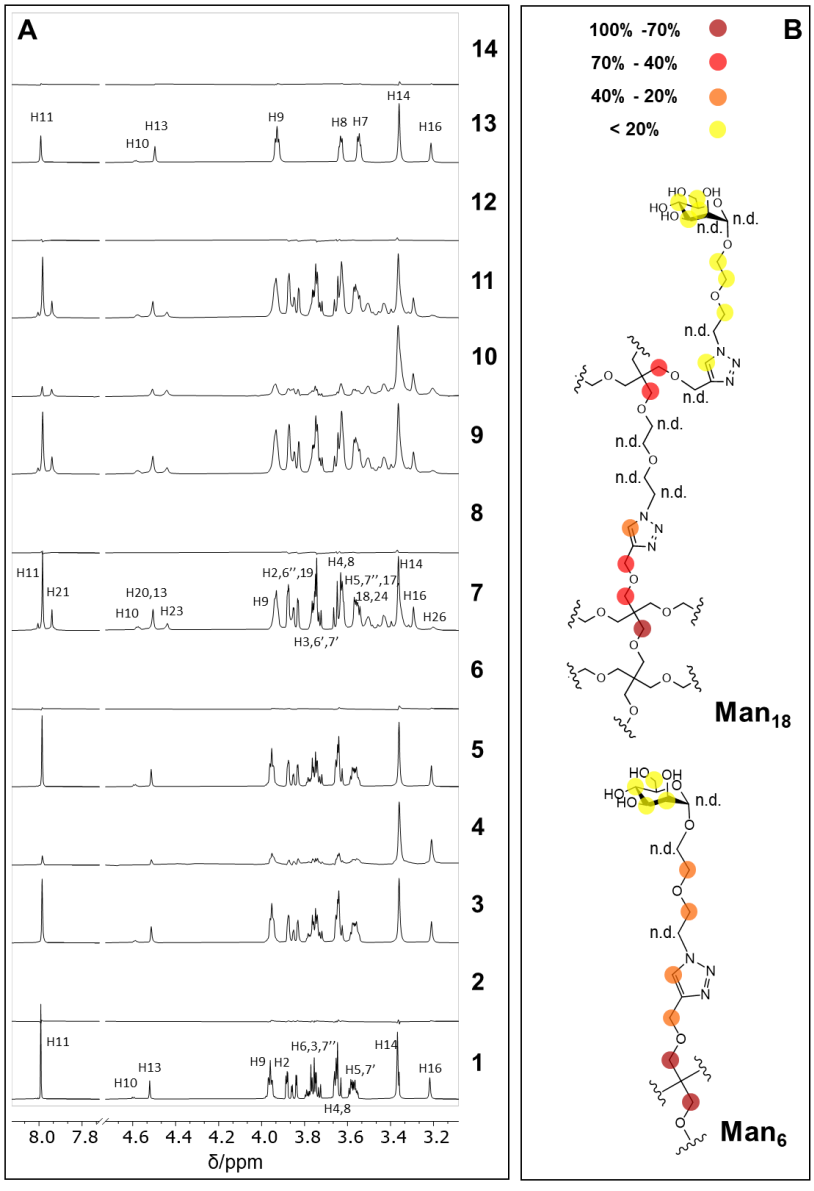


Figure 6.12. A) 1H and STD NMR spectra acquired in presence of ligands and different bacterial strains; cells were always used at a fixed concentration of 2 × 10^9 cells in each NMR tube 1) 1H spectrum of **Man$_6$** (1 mM); 2) STD spectrum of **Man$_6$** (1 mM); 3) 1H spectrum of **Man$_6$** (1mM) and CFT073 cells; 4) STD spectrum of **Man$_6$** (1mM) and CFT073 cells; 5) 1H spectrum of **Man$_6$** (1 mM) and CFT073*ΔfimH* cells; 6) STD spectrum of **Man$_6$** (1 mM) and CFT073*ΔfimH* cells; 7) 1H spectrum of **Man$_{18}$** (0.25 mM); 8) STD spectrum of a sample containing **Man$_{18}$** (0.25 mM); 9) 1H spectrum of **Man$_{18}$** (0.25 mM) and CFT073 cells; 10) STD spectrum of **Man$_{18}$** (0.25 mM) and CFT073 cells; 11) 1H spectrum of **Man$_{18}$** (0.25 mM) and CFT073*ΔfimH* cells; 12) STD spectrum of **Man$_{18}$** (0.25 mM) and CFT073*ΔfimH* cells; 13) 1H spectrum of core-scaffold (1 mM) and CFT073 cells; 14) STD spectrum of core-scaffold (1 mM) and CFT073 cells. B) **Man$_6$** and **Man$_{18}$** binding epitopes obtained from the STD NMR *on-cell* experiments acquired in presence of CFT073 cells. The expression of the results is shown in percentage. The relative STD effect of some protons were not determined (n.d.) due to peak overlapping or proximity to solvent resonance.

6.4 The interaction of siglec-10 with glycomimetics: a view by *on-cell* NMR

In this section, further applications of *on-cell* NMR will be presented. The study is focused on the inspection of the binding of siglec-10 to a synthetic molecule (**F9**), a sialic acid glycomimetic. The aim behind this strategy was to examine, through *on-cell* NMR, molecular recognition events that take place at the surface of mammalian cells. In particular, we chose HEK293 cells, which have been widely used in the lab to generate well-folded glycoproteins. [44]

The project is still ongoing in our laboratory, within the framework of a project in the Chemical Glycobiology group dedicated to the siglec family of receptors (coordinated by Dr. June Ereño-Orbea). [45] Other key persons in this project have been Dr. Ana Ardà, in the NMR side, and Maria Pia Lenza, a former postdoc in the group, in the protein expression, purification, and manipulation side. The target molecule, the so-called **F9** glycomimetic (Figure 6.13), selected through a screening campaign based on the synergic employment of high-throughput synthesis and microarray selection, was kindly provided by Prof. James H. Paulson (Scripps Research Institute, USA). [46]

6.4.1 General context

Although galectins carry out their biological function also in the extracellular micro-environment, they are soluble and not anchored to the cell membrane. [47] Therefore, they are not proper targets to detect their binding properties through *on-cell* NMR methods. To curb the problem, strategies to artificially anchor their CRDs to the cell membrane or the creation of chimeras can been explored. [48–51] However, in this particular case, we decided to change the target protein, focusing on a lectin membrane receptor, such as the siglec family, which are available in the lab.

As commented in the Introduction chapter, the sialic acid-binding immunoglobulin (Ig)-like lectins (siglecs) family is composed of 15 members in humans. [45,52] They are receptors of the innate immune system and preferentially bind sialic acid moieties connected to other sugar units through α2-3, α2-6, and α2-8 linkages. [53–56] The key establishment of a salt bridge between a conserved arginine residue of the protein and the carboxyl group of the sialic acid is common for all siglecs. [52] As frequently found

for lectin-sugar interactions, the affinity of these lectins for monomeric sialic acid-containing ligands is rather weak (from the high micromolar to the low millimolar range). [45]

We focused on **siglec-10**, besides its biological relevance, also because, among CD33-related, is the one with the higher number of extracellular domains (fove), making the CRD more exposed to the extracellular environment to interact with exogenous ligands. [57] Briefly, siglec-10 is expressed on immune cells of the myeloid lineage as well as on B cells, where it regulates innate and adaptive immune responses to tissue injury, sepsis, and viral invasion. [45,58–61] Recently, siglec-10 has also been shown to promote immune evasion of CD24-expressing tumour cells. [58]

Concerning its binding preferences, a minor preference for α2-3 linked sialosides over their α2-6 counterparts has been described, a lower specificity when compared to other siglecs. [62] Significant efforts have been employed to develop chemically modified sialylated ligands able to bind siglecs with elevate affinity and selectivity and then modulate their biological actions. A limited number of high affinity ligands for siglec-10 have been proposed. The glycomimetic **F9**, containing a triazole-linked adamantane at the C-9 end of the 9-amino-Neu5Acα2-3Lac trisaccharide, did emerge from a click-and-pick approach, and showed selective and strong binding to liposomes displaying ligand **F9** to Siglec-10^{+} human peripheral blood cells. [46]

Since the *on-cell* STD-NMR strategy is also favoured when using ligands with high specificity for the target receptor, the **F9** glycomimetic was selected. This type of NMR protocol has already been employed by us and by other to detect, at atomic level, interactions occurring on the cell surface. [24,39,40,63,64]

Nevertheless, the *on-cell* NMR experiments surmount some of the unfavourable aspects described for the *in-cell* NMR analysis (Chapter 6.2.1), since:

- The target receptor can be overexpressed directly in its natural cell line.
- The selective labelling of the protein is not required when using standard 1D-STD NMR experiments.
- The problems arising from the large tumbling of the protein are minimized. In fact the extracellular portions of the receptor still display a relatively fast motion, with adequate effective rotational motion correlation times to provide sharp NMR signals of their bound ligands. The viscosity of the extracellular

environment in the NMR tube is lower than that within the cell, considering also that the cells suffer several washes before the NMR measurements.

However, some challenges still remain:

- The existence of non-specific interactions with other receptors can undermine the study. Both *cis* and *trans* interactions between receptors can occur. Nevertheless, depending on the specific case, different solutions can be designed and applied.
- Cell vitality is an issue, because it has to be maintained during the acquisition. Thus, the experiments have to be set with short recording times and this point obviously compromises the acquisition of complex 2D experiments. Different methods have been proposed to overcome this phenomenon and the natural cell sedimentation process, including the use of bioreactors specifically designed for NMR experiments or the employment of methylcellulose hydrogels. [17,20] However, every single case have to be evaluated, since some mammalian cell lines can result sensitive to these type of treatments.

Therefore, these factors were considered when analysing of the interaction of the overexpressed full length siglec-10 on mammalian HEK293 cells in the presence and absence of the **F9** glycomimetic using *on-cell* STD-NMR experiments (Figure 6.13).

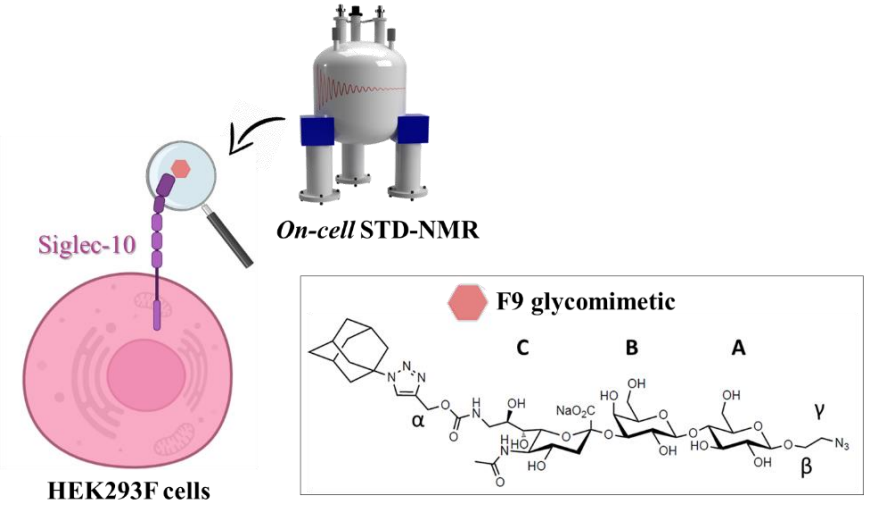


Figure 6.13. Above, representation of the strategy adopted in this chapter to detect on-cell binding events. Below, structure of the **F9** glycomimetic employed in the study. [46]

6.4.2 Overexpression of siglec-10 in HEK293F cell lines

The chosen hosts for the overexpression of siglec-10 were HEK293 cells, as they are a cell line particularly easy to transfect and to maintain in culture. Nowadays, many subtypes and derivatives of HEK293 cells are available. [65] We chose HEK293F, which are adapted (from the parental HEK293T attached cells) to live in in suspension culture in serum-free media.

Thus, the full length siglec-10 receptor was overexpressed in this cell line. The effective overexpression was proved by FACS analysis (data not shown). From the comparison of non-transfected and transfected cells, it emerged that the hits generated by the binding of the antibody anti-siglec10 to the transfected cells were considerably higher.

The protocol for the transfection is described in the Method chapter; however, it is worthy to mention that FectoPRO was chosen as transfection agent, mixed with an equal amount of the vector and gently added to the flask containing the cells. Two days after the transfection, the cells were harvested and washed twice with deuterated PBS before the NMR measurements.

6.4.3 *On-cell* STD-NMR experiments

The conditions for the *on-cell* STD-NMR experiments (Figure 6.14) were carefully optimized and the final setup was designed as follow:

- 2 x 10^6 cells were used for each NMR sample. Their vitality was measured before and after the experiments and decreased by 20-25%.
- The **F9** mimetic was employed at millimolar concentration (1 mM per NMR tube).
- The on-resonance irradiation frequency used was set at δ -0.8 ppm, adjusted in order to keep it at a safe frequency difference to the last signal arising from the glycomimentic (δ 1.5 ppm).
- The experimental parameters were set to ensure a short recording time (1 hour).
- The spectra were acquired at 286 K using 2 s of saturation time.

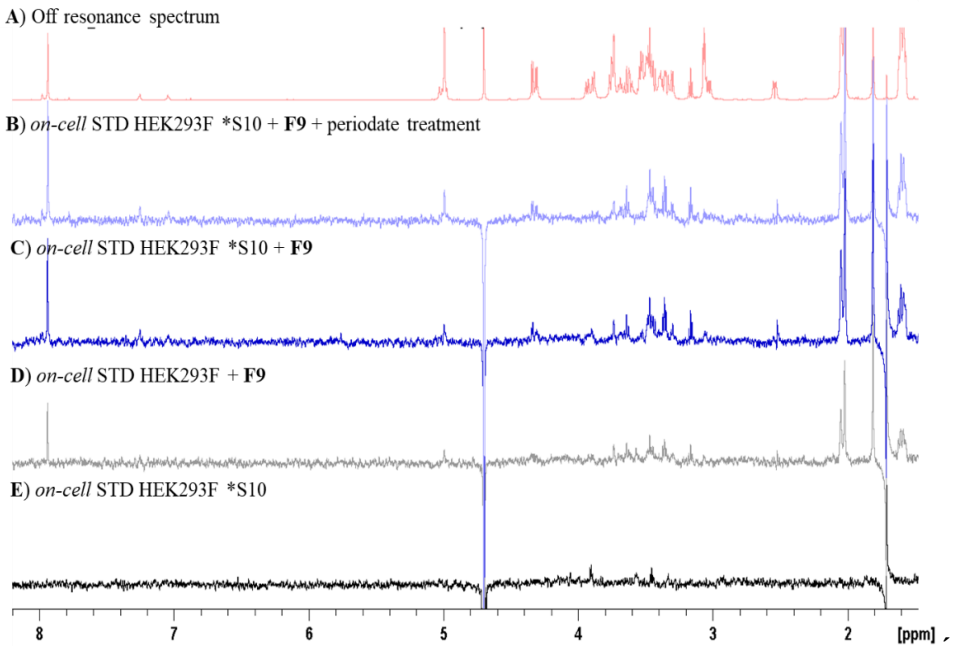


Figure 6.14. Overview of the *on-cell* STD NMR results. The concentration of cells (HEK293F) was set at $2x10^6$ cells/tube with ligand **F9** concentrated, when present, at 1 mM. A) off-resonance spectrum acquired for the sample containing transfected cells with **F9**. B) *On-cell* STD-NMR experiment acquired with transfected cells, **F9** and after mild periodate treatment. C) *On-cell* STD-NMR experiment acquired with transfected cells and **F9**. D) *On-cell* STD-NMR experiments acquired with non-transfected cells and **F9**. E) *On-cell* STD-NMR experiment acquired with transfected cells without ligand.

Additional STD NMR experiments were carried out after treating the cells with sodium periodate ($NaIO_4$). Indeed, sialic acid-terminated glycolipids and glycoproteins exposed on mammalian cell surfaces are natural ligands for siglecs and may interact with them when they are displayed on the same cell surface (*cis*) or at two different cells (*trans*). A possible strategy to avoid these undesired binding event is to treat the transfected cells with sodium periodate. [64,66] The periodate disrupts the *cis*-binding process since it truncates the key interacting vicinal diols of the sialic acid moieties through oxidation. This reaction provides modified molecules that do not bind to siglecs (Figure 6.15) [66]

Moreover, in order to obtain robust and non-ambiguous results, different control experiments were also performed using i) non-transfected cells, ii) in absence of the ligand, or iii) with the isolated purified receptor.

All the experiments were carried out employing the same identical set-up, as follows:

- *On-cell* STD-NMR experiments with transfected HEK293F cells (overexpressing siglec-10) and the **F9** mimetic after mild periodate treatment. (Figure 6.14, B).
- *On-cell* STD-NMR experiments with transfected HEK293F cells (overexpressing siglec-10) and the **F9** mimetic. (Figure 6.14, C).
- *On-cell* STD-NMR experiment with non-transfected HEK293F cells and the **F9** mimetic. (Figure 6.14, D).
- *On-cell* STD-NMR experiment with transfected HEK293F cells (overexpressing siglec-10) without any added ligand (Figure 6.14, E).
- *In-vitro* STD-NMR with the purified siglec-10_{d1d2} lectin with the **F9** mimetic.

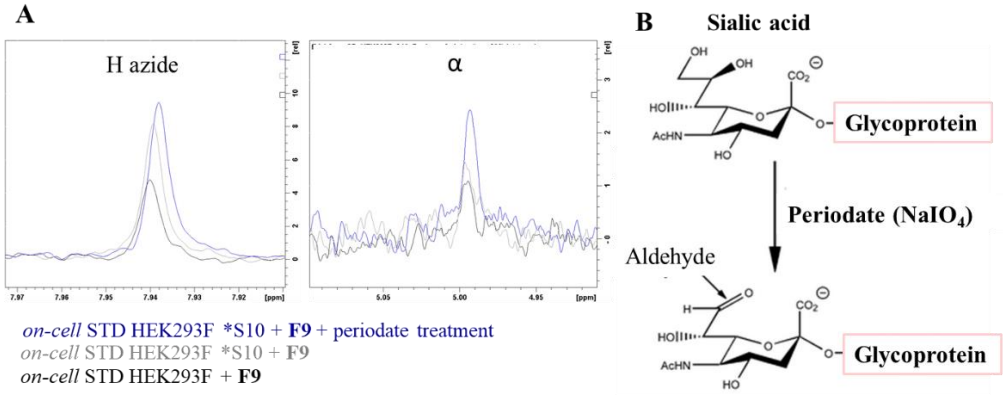


Figure 6.15. A) Zoom-in of two STD proton signals of **F9** mimetic in the *on-cell* STD experiments of cells transfected and undergone to periodate treatment with **F9** (blu), cells transfected with **F9** (grey) and non-transfected cells with **F9** (black). Details of the experiments described in Figure 6.14 B) Chemical action of the mild periodate treatment on the sialic acid.

From the comparison of the entire set of spectra, it can be affirmed that the *on-cell* STD-NMR experiments were successful, since a selective profile was obtained (Figure 6.14), as detailed below. However, the inspection of the results allows concluding that a complete specificity for siglec-10 was not achieved, since a weak STD profile also appeared in the experiments using the non-transfected cells. This evidence suggests that other unknown entities transiently bind to **F9** in the absence of an elevate presentation of siglec-10 on the cells. Doubtless, the STD intensity is way lower than that observed in the experiments recorded when siglec-10 is overexpressed on the cell surface, strongly suggesting that the detected STD NMR enhancement is indeed due to the presence of the siglec-10 receptor on the cellular membranes.

Focusing on the experiments performed after the treatment with mild periodate, it can be perceived that the STD NMR signal increased when compared to the untreated sample that has the intact sialic acids, without disruption of the diols that may mediate the *cis*-interactions. (Figure 6.14).

From the detailed comparison of each STD NMR experiment, a general trend can be deduced: the observed STD intensities are continuously increasing when passing from the non-transfected cells, to the transfected ones and to those transfected and treated with periodate (Figure 6.15). This tendency can be explained by the existence of the interaction of the extracellular siglec-10 with **F9** and with the involvement of siglec-10 in *cis*-interactions that are disrupted with mild periodate treatment. Under these conditions, more receptors are available for the interaction with the **F9** mimetic.

The STDD (Saturation Transfer Double Difference) spectrum, generated by subtracting the *on-cell* STD obtained without receptor overexpression (Figure 6.14, D) to that acquired with transfected cells and periodate treatment (Figure 6.14, B), displayed a clear STD profile that was used to obtain the final binding epitope (Figure 6.16).

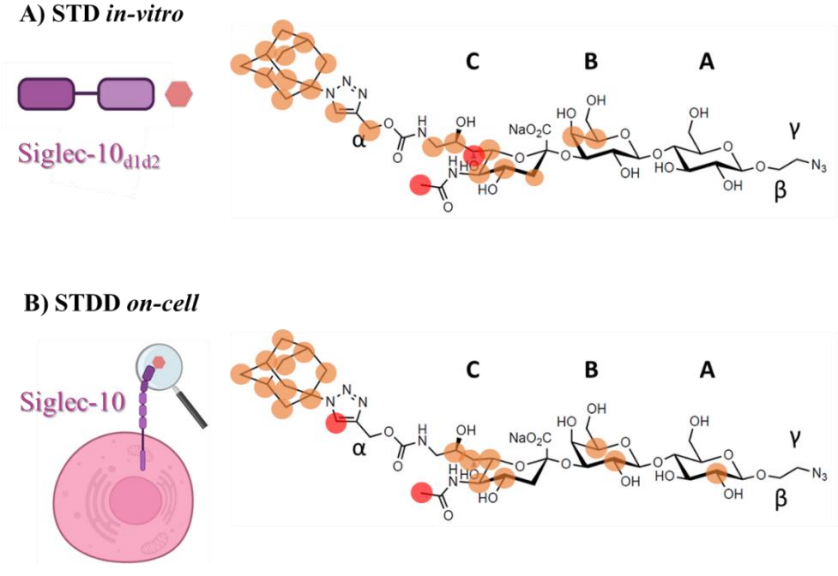


Figure 6.16. STD binding epitope obtained for the *in-vitro* experiment with siglec-$_{10d1d2}$ (concentrated 30 μM) and **F9** (0.9 mM) (A) and for the STDD spectrum (B) generated by subtracting the *on-cell* STD obtained without receptor overexpression (Figure 6.14, D) to the one acquired with transfected cells and periodate treatment (Figure 6.14, B).

The **F9** epitope interacting with siglec-10 on the extracellular environment include all the protons of the adamantane moiety as well as the proton of the azide linker. H4, H5, H6, H7, H8, and the Ac moiety also displayed STD intensities. Minor contributions from the Gal unit (H2 and H5) and Glc moieties (H2) were also observed.
Fittingly, the binding epitope is consistent with that obtained from the *in-vitro* STD experiments (Figure.16). Regarding the sialic acid, the recognition pattern is similar to that reported for the interaction of siglec-10 with α2-3sialyllactosamine. [62]

The same repertory of experiments was also performed for α2-6 sialyl lactose. In this case, the resulting STD profile was not reliable (data not shown). In fact, the STD intensities arose from all the sugar rings, including the Glc moiety. Thus, the putative binding epitope was improbable for the interaction with siglec-10. Very probably, the the α2-6 sialyl lactose ligand interacts with a variety of receptors (including siglec 10 and other undefined partners on the cell surface) with similar low affinities, hampering the achievement of realistic data on its recognition by siglec-10 *on-cell*. These findings confirm that the use of a specific inhibitor with *on-cell* NMR experiments is of fundamental importance.

Although the results reported are still preliminary and need additional investigations, it is tempting to propose that the *on-cell* STD-NMR experiments have indeed demonstrated the interaction of siglec-10 exposed on the cell surface with a selective glycomimetic molecule, characterizing the binding epitope. Additionally, this binding epitope is consistent with that obtained *in-vitro* for the same system.

As a future plan, new experiments will be performed employing methylcellulose hydrogels to reduce the sedimentation rate of the cells. If this method is not successful for cells adapted to live in suspension, the possibility to use adherent cells (HEK293T) will be evaluated, since this type of cells has been shown to tolerate these conditions, even increasing their lifespan. [20]

6.5 *On-cell* galectin-glycan recognition using genetically glycoengineered mammalian cell lines

The work presented in this section gathers the results obtained during a secondment (2 months) carried out in 2022 at the Copenhagen Center for Glycomics (University of Copenhagen, Department of Cellular and Molecular Medicine, Denmark) under the supervision of Professor Hans Wandall (https://icmm.ku.dk/english/research-groups/wandall-group/). During the stay, I had the possibility to actively participate in some steps of a global work on this topic, already ongoing in the group.

The major aims of the training stay were:

- Learn how to produce controlled cellular knockouts targeting specific enzymes involved in glycan biosynthesis; to this purpose, the cutting-edge technology of clustered regularly interspaced short palindromic repeats with the CRISPR-associated protein 9 (CRISPR-Cas9) has been used. [67–69]
- Characterize, through flow cytometry, the binding of a set of selected galectins to genetically engineered cell lines presenting homogeneous and well-defined glycosylation on their surface.

In order to achieve these goals, I worked together with the laboratory technician Karin Uch Hansen and with the student assistant Frederikke Moll Engersgaard, who trained and assisted me for the glycoengineering protocol and the FACS experiments, respectively.

6.5.1 General context

Glycans decorate our cellular membranes and influence the organization and function of many proteins, especially the ones expressed on the cell surface. [70] They also modulate the interaction with the surroundings and the formation of multicellular structures such as tissue and organs. [71,72]

As commented in the Introduction Chapter, the biosynthesis of the rich assortment of glycans (the glycome) found on mammalian glycoproteins, glycolipids, and proteoglycans is a complex process involving more than 200 different **glycosyltransferases** (GTs) operating in 16 major pathways (Figure 6.17). [73] The possibility of performing a precise genome engineering of these enzymes on selected

cell lines, thus modifying the entire glycome, is a fascinating tool for glycosciences. Indeed, by exploiting precise targeting of glycogenes, it is possible to rationally modulate the glycosylation capacity of a selected cell line. [67,68,74,75]

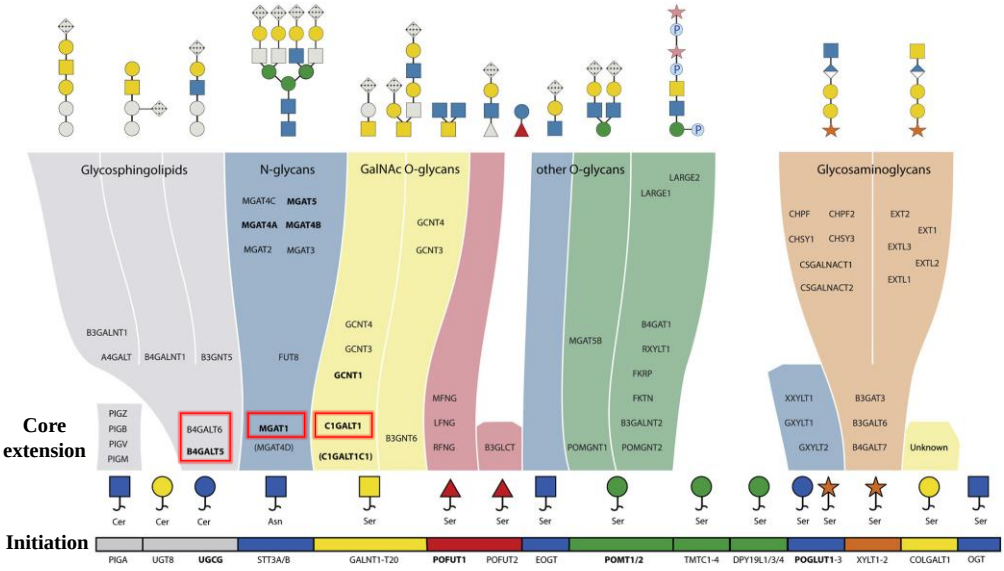


Figure 6.17. Schematic overview of the glycosylation biosynthetic pathways with main acting GTs reported. Image taken from Marinova et al., 2021. [69]

Diverse controlled **knockouts** targeting specific enzymes involved in glycan biosynthesis have been obtained by exploiting different technologies, including zing finger nuclease (ZFNs) and transcription activator-like effector nucleases (TALEN), but, lately, mostly clustered regularly interspaced short palindromic repeats with the CRISPR-associated protein 9 (**CRISPR-Cas9**) have been used. [73,76–79]
This type of strategy has opened new avenues in the area, not only for the dissection and design of well controlled glycoproteins for therapeutic purposes, but also for the study of glycan interactions with target biomolecules at the cell level. [76,80,81]

The availability of such potent tools has captured the attention of several laboratories. In particular, a few years ago, my host research group at CCG developed a comprehensive protocol to obtain Chinese hamster ovary (CHO) cells lines expressing different **homogeneous glycosylation patterns**. [76] Moreover, Wandall's team decided to exploit the glycoengineering technology for the analysis of **glycan-galectin interactions in a cellular context**. Precisely, they have recently used this tool to probe

the binding of selected galectins to CHO cells presenting different complex N-linked glycans and core-1 O-glycans. [80]
They performed an exquisite genetic deconstruction of cellular glycosylation, demonstrating that bi-antennary N-glycans are sufficient for high-affinity binding of galectins and that sialylation is an important checkpoint for their binding ability. They also proved that galectin-8N can binds the core-1 O-glycans with high affinity. [80]

Based on these prior achievements, they decided to further expand this workflow and to characterize the interaction of a wider panel of galectins with a new generation of cell lines. In particular, the new designed cell lines differed from the ones employed in the previous work in terms of: [80]

i. Type of cell used (human embryonic kidney 293 -**HEK293-** and immortal human keratinocyte –**HaCaT-** cells instead of CHO cells); the choice lies in the fact that both HaCaT and HEK293 are widely used human cells lines with slightly different glycosylation capacity among them and with respect to the more simplified glycosylation pattern of CHO cells. [82]
ii. Glycosyltransferases knock-out (KO). In fact, the KOs were planned to obtain cells presenting only N-glycans, complex O-glycans or glycosphingolipds (instead of focusing mainly on the modifications of N-glycans' antennae). The idea is to implement the past knowledge of stable lines bearing only a single KO with combinatorial KOs and thus discern the role of each player in the recognition phenomenon.

I participated in the glycoengineering (through CRISPR-Cas9 methodology) of the HaCaT cells (Chapter 6.5.2), whereas the library of glycoengineered HEK293 cells was already available in the laboratory. Indeed, I had the opportunity to use them for preliminary flow cytometry experiments with galectins (Chapter 6.5.3).

6.5.2 Generation of glycoengineered HaCaT cell lines

At the time of designing KOs of GTs, it is crucial to know whether they can be differentiated in non-redundant and redundant enzymes. In fact, some steps in the glycosylation pathways in mammalian cells are regulated by a **single GT (non-**

redundant). Therefore, the loss of function of such genes leads to global changes in the glycan structures. Contrarily, some other glycosylation steps that are regulated by **multiple isoenzymes (redundant)**, where the loss of function of a single isoenzyme gene may lead to no or only subtle effects on the glycans. [68,73,81] Usually, elongation, branching and capping are covered by partial redundancies, while the very first core extension is less interested by this phenomenon (Figure 6.17).

HaCaT cell lines bearing only N-glycans, O-glycans or glycospgingolipids (GSL) were obtained performing the combinatorial KOs of multiple genes. The detailed experimental procedure, based on the employment of CRISPR-Cas9 technology, is described in Chapter 7.

The following mutant cells lines were produced (Figure 6.18):

- HaCaT KO *mgat1/c1galt1*
- HaCaT KO *mgat1/B4galt5/6*
- HaCaT KO *c1galt1/B4galt5/6*

All the KO enzymes act in the core extension steps of glycan synthesis (Figure 6.17).

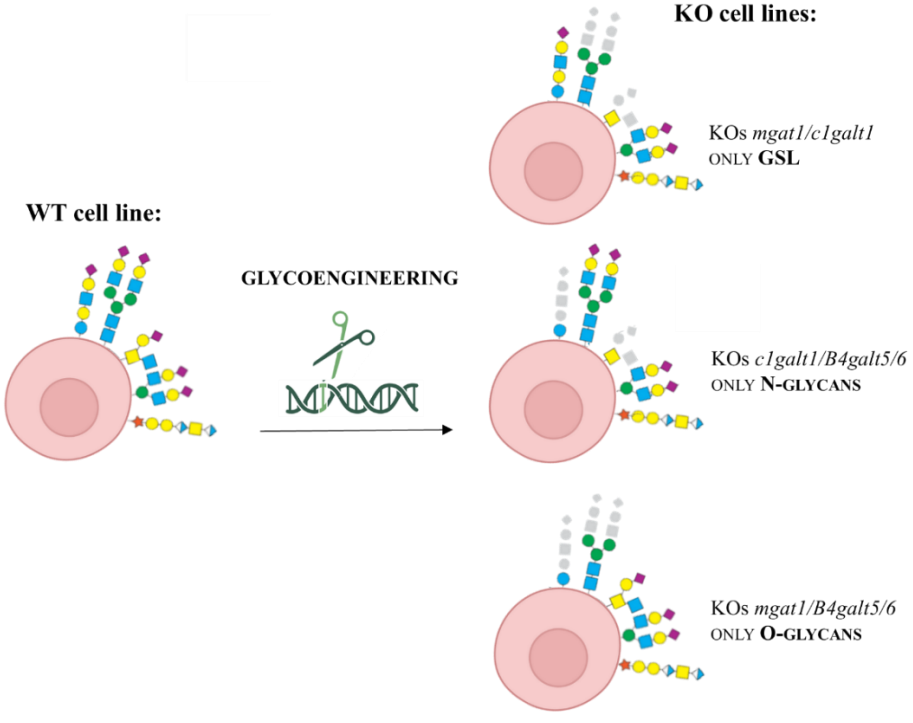


Figure 6.18. Overview of the different glycolengineered cells lines produced for this work with the respective genes KO indicated

It should be pointed out that none of the cell lines produced has a controlled synthesis of non-mucin type O-glycosylations (O-fucosylation, O-mannosylation, or O-glucosylation) and glycosaminoglycans (GAGs), which that are expected to be present.

6.5.2.1 Phenotypical validation of glycoengineered HaCaT cells

Once isolated the desired combinatorial KOs, they were phenotypically examined through FACS based assay. In detail, their binding to three control biotinylated lectins (PNA, VVA and PHA-L) was scrutinized through fluorescence detection (see Chapter 7 for the experimental details).

Each lectin recognizes specific epitopes:

- Biotinylated *Phaseolus vulgaris* Leucoagglutinin (**PHA-L**) preferably binds to GlcNAc residues presented on the 6 branches of N-glycans, and it is expected to bind to wt cells and only to N-glycan presenting lines.
- Biotinylated Peanut Agglutinin (**PNA**) preferably binds to β-Gal terminal moieties presented on Tn-mucin-type O-glycans, and it is expected to bind to wt cells and also to O-glycan presenting lines, but only after treatment with neuraminidase (in order to cleave the sialic acid cappings).
- Biotinylated *Vicia villosa* lectin (**VVA**) preferably binds to GalNAc residues presented on truncated mucin-type O-glycans. It is expected to bind to cells with inactive O-glycans synthesis (only N-glycans or glycosphingolipids-presenting cell lines).

The FACS assay proved the quality of the obtained glycoengineered HaCaT cell lines (Figure 6.19). In fact, the binding profiles of the control lectins were in accordance with the expected glycan presentation on the different lines and, moreover, was almost identical to the profile obtained previously testing the same type of glycoengineering on HEK293 cells (Figure 6.19).

Specifically, those cells only expressing N-glycans displayed binding to PHA and VVA and low binding to PNA. Fittingly, the cell line that only presented glycosphingolipids showed low binding to PHA and PNA, and high affinity to VVA. To test cells expressing only mucin-type O-glycans, the culture was previously treated

with neuraminidase, to cleave the sialic acid terminal moieties, and then incubated with PNA. The cells only showed high affinity for PNA after the enzymatic treatment. No binding to the other lectins, including PNA without previous neuraminidase incubation, was observed.

In a further step, the glycoprofiles resulting from the combinatorial enzymatic knockouts will be assessed using mass spectrometry (ongoing experiments).

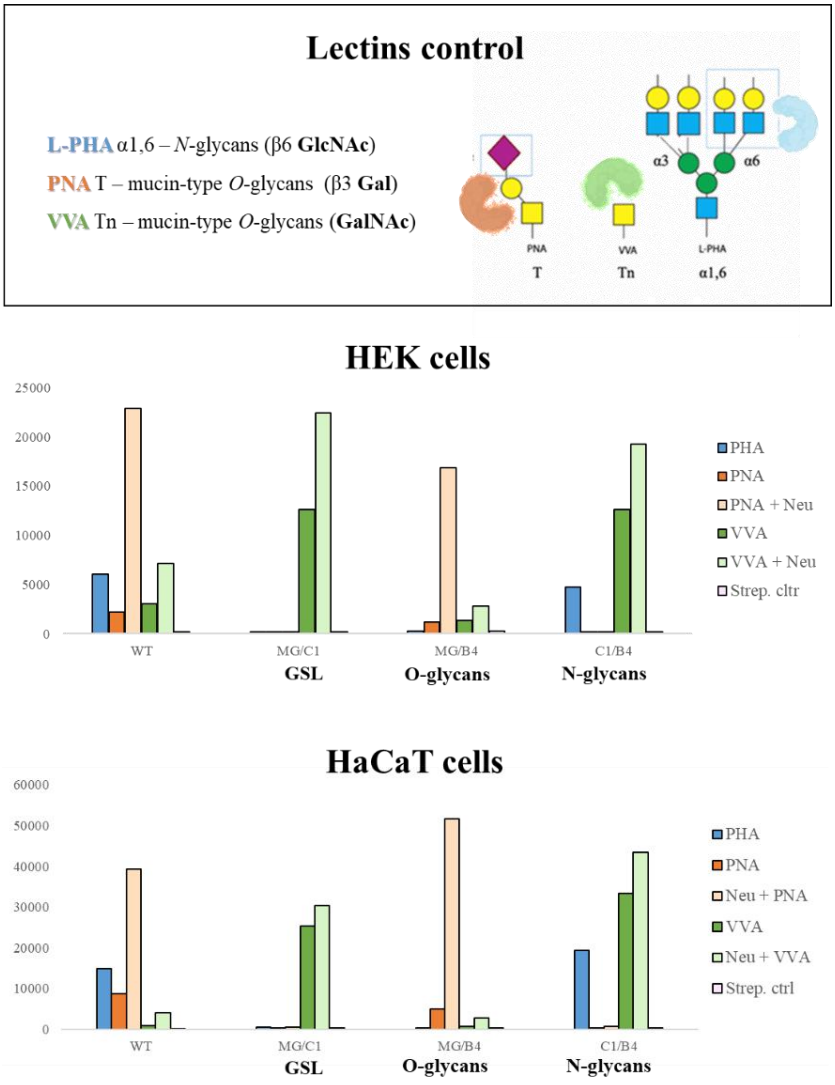


Figure 6.19. Lectin control cell-surface binding to different glycoengineered HEK and HaCaT cells. Above: schematic overview of the lectin control binding sites. Below: Fluorescence quantification from FACS assay for each cell line treated with the respective lectin control. The results are expressed in terms of MFI (Median Fluorescence Intensity).

6.4.3 Binding of galectins to glycoengineered HEK293 cell lines

In a further step, the initial results about the specific relevance of complex N-glycans, O-glycans, and glycosphingolipids (GSL) in the interaction with galectins were obtained through flow cytometry. The experimental details are discussed in Chapter 7 During the secondment, galectins -1, -3, -8C, and -8N conjugated to NSH-fluorescein were tested for their capacity to recognize HEK293 KO cell lines. The initial data are reported in Figure 6.20.

Lactose 100 mM was used to treat cells as negative control. In particular, the background fluorescence of lactose-treated samples was subtracted for all the tested concentrations of galectin tested to provide the final results.

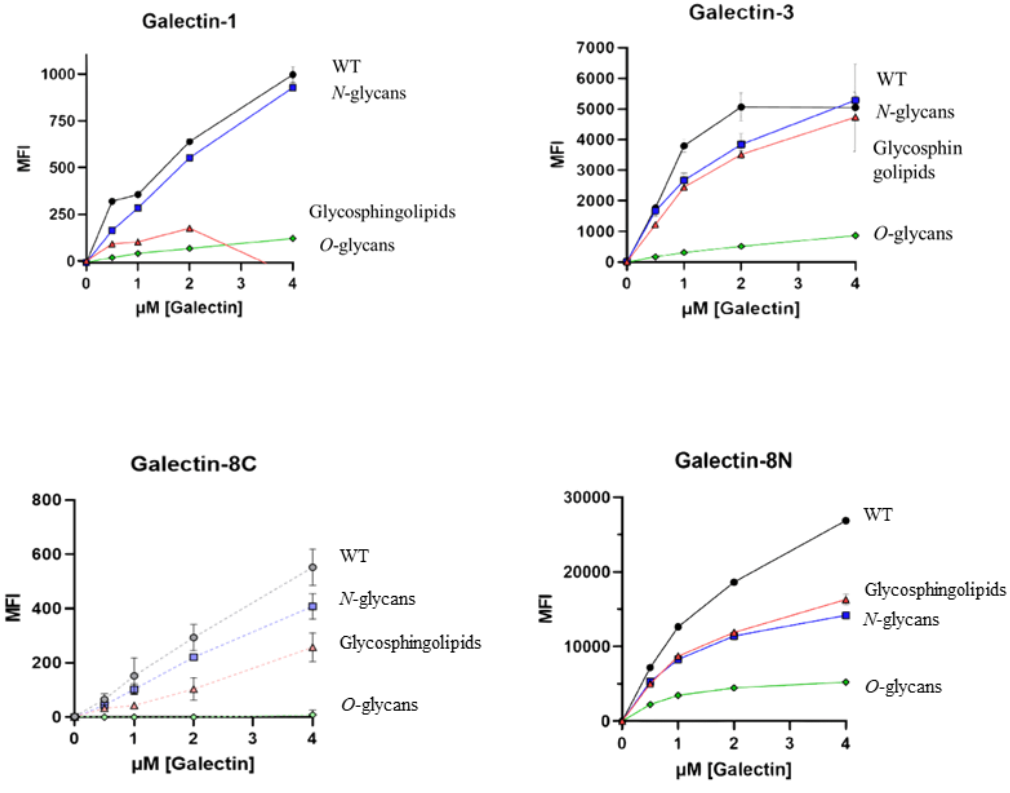


Figure 6.20. Preliminary data of the galectin cell-surface binding experiments to different glycoengineered HEK cells. Fluorescence quantification obtained from FACS analysis for each cell line treated with the increasing galectin concentrations. Results expressed in terms of MFI (Median Fluorescence Intensity) after subtracting background fluorescence in the presence of 100 mM lactose as negative control.

Although the experiments should be further repeated to obtain robust average values, from the observation of these FACS results it can be deduced that all galectins bound to the cells presenting only N-glycans (*c1galt1/B4galt5/6* KOs).
Moreover, galectin-1 bound with high affinity to the N-glycans presenting cells, apart from to the wt line. On the other hand, galectin-3 bound to both N-glycans and glycosphingolipids (*mgat1/c1galt1* KOs) in the cellular context, with highest affinity for N-glycans. Similarly, galectin-8C binds to N-glycans and GSL, but with diverse affinities (higher affinity for N-glycans than to glycosphingolipids). Interestingly, galectin-8N, apart from binding to N-glycans and GSL, resulted the only galectin, among the tested ones, which showed also association with mucin-type O-glycans (*mgat1/B4galt5/6 KOs*).

New sets of FACS assays will be performed in the next months using the HaCaT cells KOs produced during the secondment. This aspect is interesting since the host group has found that, generally, the HaCaT cells' glycosylation pattern displays extended poly-LacNAc elongation, a higher degree of sialylation, and an increased degree of complexity for O-glycans structures, when compared to the HEK cells.
Galectin specificity toward N-glycans is particularly dependent on branching and terminal modifications. In particular, poly-LacNAc extensions on tri- and tetra-antennary N-glycans have been proposed to be of major importance for providing specificity to galectin- 1, -3, and -8. [83]
Also, sialylation of terminal galactose residues has been proved to be important for some galectins. For example, while glycans capped with 2,6-sialic acid globally block galectin binding, whereas galectin-8N has a particularly high affinity for galactose residues with a terminal 2,3-sialic acid. [80]
In this context, additional information concerning the unique specificity of galectin-8N to bind O-glycans with high affinity will be obtained by performing additional experiments with systems displaying increasing O-glycans complexity. [80]

Finally, the array of galectins will be extended with more members of the family (galectin-4, galectin-7 and galectin-9, already available at CIC bioGUNE).

6.6 Conclusions

The observation of **intracellular events** is always fascinating. In the panorama of protein-sugar recognition processes, the task is quite difficult, since lectins are generally promiscuous and can interact with several partners inside the cells. Here a methodology for the detection of galectin-7 injected into zebrafish oocytes have been proposed and its intracellular detection alone or in the presence of a glycomimetic ligand has been demonstrated by NMR.

The extracellular matrix (ECM) is an intricate network composed by proteins, glycoproteins, and other components that provides physical support for cells but also determines biological functions. Its complexity, unknown for a long time, can be compared to the one observed in the intracellular compartments.
Knowing that, the available methods to study ***on-cell*** **phenomena** are usually an approximation, because several manipulations of the cells are done right after the measurements. However, they can be considered good types of approximation, since allow to study receptors directly embedded into membranes and without have to purifying them. In this chapter methodologies to study extracellular binding events involving not only NMR but also other emerging techniques based on precise genome engineering, have been explored.

With the aim to find inhibitors of the FimH-mediated adhesion of the uropathogenic *E. coli* strains to host epithelium, the *on-cell* STD-NMR methodology has been employed. The protocol allowed assessing the ligand-lectin binding directly on the surface of intact cells, mimicking the physiological environment. D-mannose decorated dendrimers were synthetized to reproduce the multiple epitope presentation of the mannosides naturally recognized by FimH. This strategy permitted to identify the structural determinants of the binding of new potential inhibitors of adhesion under physiologically relevant conditions and at atomic level.
Similarly, the recognition event of the siglec-10 receptor with a glycomimetic with marked selectivity, employing mammalian HEK293F, cells have been unravelled. Although further validations have to be performed, this approached allowed to obtain

a reliable first information concerning the binding epitope of this extracellular interaction in cellular conditions.

To finish, preliminary experiments have allowed analysing the galectin-binding specificities *versus* HEK cells exposing only N-glycans, mucin-type O-glycans, or glycosphingolipids. To achieve this aim, controlled glycoengineered HaCaT and HEK cell lines were successfully obtained.

These results represent the first step towards the study of combinatorial effects of glycosylation on the interaction with galectins in an extracellular context. Once completed and optimized, this tool will provide an array of cells presenting tailored glycans in their native context, which will be used to interrogate carbohydrate ligands of lectins of interest.

The future interpretation of such significant outcomes will hopefully shed light on how galectins control the organization, endocytosis, and residence time of membrane proteins. Moreover, this unique approach to glycomics is a promising tool to uncover and dissect how glycosylation can modulate galectin recognition in health and disease events. [74]

6.7 References

1. Serber, Z.; Dötsch, V. In-Cell NMR Spectroscopy†. *Biochemistry* **2001**, *40*, 14317–14323, doi:10.1021/BI011751W.
2. Tochio, H. Watching Protein Structure at Work in Living Cells Using NMR Spectroscopy. *Curr. Opin. Chem. Biol.* **2012**, *16*, 609–613, doi:10.1016/J.CBPA.2012.10.022.
3. Freedberg, D.I.; Selenko, P. Live Cell NMR. *https://doi.org/10.1146/annurev-biophys-051013-023136* **2014**, *43*, 171–192, doi:10.1146/ANNUREV-BIOPHYS-051013-023136.
4. Luchinat, E.; Banci, L. In-Cell NMR: A Topical Review. *urn:issn:2052-2525* **2017**, *4*, 108–118, doi:10.1107/S2052252516020625.
5. Theillet, F.X.; Luchinat, E. In-Cell NMR: Why and How? *Prog. Nucl. Magn. Reson. Spectrosc.* **2022**, *132–133*, 1–112, doi:10.1016/J.PNMRS.2022.04.002.
6. Serber, Z.; Keatinge-Clay, A.T.; Ledwidge, R.; Kelly, A.E.; Miller, S.M.; Dötsch, V. High-Resolution Macromolecular NMR Spectroscopy inside Living Cells. *J. Am. Chem. Soc.* **2001**, *123*, 2446–2447, doi:10.1021/JA0057528/ASSET/IMAGES/LARGE/JA0057528F00002.JPEG.
7. Selenko, P.; Serber, Z.; Gadea, B.; Ruderman, J.; Wagner, G. Quantitative NMR Analysis of the Protein G B1 Domain in Xenopus Laevis Egg Extracts and Intact Oocytes. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11904–11909, doi:10.1073/PNAS.0604667103/SUPPL_FILE/04667TABLE1.PDF.
8. Zhu, W.; Guseman, A.J.; Bhinderwala, F.; Lu, M.; Su, X.C.; Gronenborn, A.M. Visualizing Proteins in Mammalian Cells by 19F NMR Spectroscopy. *Angew. Chemie Int. Ed.* **2022**, *61*, e202201097, doi:10.1002/ANIE.202201097.
9. Ogino, S.; Kubo, S.; Umemoto, R.; Huang, S.; Nishida, N.; Shimada, I. Observation of NMR Signals from Proteins Introduced into Living Mammalian Cells by Reversible Membrane Permeabilization Using a Pore-Forming Toxin, Streptolysin O. *J. Am. Chem. Soc.* **2009**, *131*, 10834–10835, doi:10.1021/JA904407W/SUPPL_FILE/JA904407W_SI_001.PDF.
10. Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; et al. High-Resolution Multi-Dimensional NMR Spectroscopy of Proteins in Human Cells. *Nat. 2009 4587234* **2009**, *458*, 106–109, doi:10.1038/nature07839.
11. Bertrand, K.; Reverdatto, S.; Burz, D.S.; Zitomer, R.; Shekhtman, A. Structure of Proteins in Eukaryotic Compartments. *J. Am. Chem. Soc.* **2012**, *134*, 12798–12806, doi:10.1021/JA304809S/SUPPL_FILE/JA304809S_SI_001.PDF.
12. Banci, L.; Barbieri, L.; Bertini, I.; Luchinat, E.; Secci, E.; Zhao, Y.; Aricescu, A.R. Atomic-Resolution Monitoring of Protein Maturation in Live Human Cells by NMR. *Nat. Chem. Biol. 2013 95* **2013**, *9*, 297–299, doi:10.1038/nchembio.1202.
13. Luchinat, E.; Banci, L. A Unique Tool for Cellular Structural Biology: In-Cell NMR *. *J. Biol. Chem.* **2015**, *291*, 3776–3784, doi:10.1074/jbc.R115.643247.
14. Binolfi, A.; Limatola, A.; Verzini, S.; Kosten, J.; Theillet, F.X.; May Rose, H.; Bekei, B.; Stuiver, M.; Van Rossum, M.; Selenko, P. Intracellular Repair of Oxidation-Damaged α-Synuclein Fails to Target C-Terminal Modification Sites. *Nat. Commun. 2016 71* **2016**, *7*, 1–10, doi:10.1038/ncomms10251.
15. Barbieri, L.; Luchinat, E.; Banci, L. Characterization of Proteins by In-Cell NMR Spectroscopy in Cultured Mammalian Cells. *Nat. Protoc. 2016 116* **2016**, *11*, 1101–1111, doi:10.1038/nprot.2016.061.
16. Luchinat, E.; Letizia Barbieri, ·; Cremonini, M.; Banci, · Lucia Protein In-Cell NMR Spectroscopy at 1.2 GHz. *J. Biomol. NMR* **2021**, *75*, 97–107, doi:10.1007/s10858-021-00358-w.

17. Sharaf, N.G.; Barnes, C.O.; Charlton, L.M.; Young, G.B.; Pielak, G.J. A Bioreactor for In-Cell Protein NMR. *J. Magn. Reson.* **2010**, *202*, 140–146, doi:10.1016/J.JMR.2009.10.008.
18. Kubo, S.; Nishida, N.; Udagawa, Y.; Takarada, O.; Ogino, S.; Shimada, I. A Gel-Encapsulated Bioreactor System for NMR Studies of Protein–Protein Interactions in Living Mammalian Cells. *Angew. Chemie* **2013**, *125*, 1246–1249, doi:10.1002/ANGE.201207243.
19. Inomata, K.; Kamoshida, H.; Ikari, M.; Ito, Y.; Kigawa, T. Impact of Cellular Health Conditions on the Protein Folding State in Mammalian Cells. *Chem. Commun.* **2017**, *53*, 11245–11248, doi:10.1039/C7CC06004A.
20. Ausgabe, D.; Mateos, B.; Sealey-Cardona, M.; Balazs, K.; Konrat, J.; Staffler, G.; Konrat, R.; Mateos, B.; Sealey-Cardona, M.; Konrat, J.; et al. NMR Characterization of Surface Receptor Protein Interactions in Live Cells Using Methylcellulose Hydrogels. *Angew. Chemie* **2020**, *132*, 3914–3918, doi:10.1002/ANGE.201913585.
21. Sánchez-López, C.; Labadie, N.; Lombardo, V.A.; Biglione, F.A.; Manta, B.; Jacob, R.S.; Gladyshev, V.N.; Abdelilah-Seyfried, S.; Selenko, P.; Binolfi, A. An NMR-Based Biosensor to Measure Stereospecific Methionine Sulfoxide Reductase Activities in Vitro and in Vivo**. *Chem. – A Eur. J.* **2020**, *26*, 14838–14843, doi:10.1002/CHEM.202002645.
22. Thole, J.F.; Fadero, T.C.; Bonin, J.P.; Stadmiller, S.S.; Giudice, J.A.; Pielak, G.J. Danio Rerio Oocytes for Eukaryotic In-Cell NMR. *Biochemistry* **2021**, *60*, 451–459, doi:10.1021/ACS.BIOCHEM.0C00922/SUPPL_FILE/BI0C00922_SI_002.AVI.
23. Nesmelova, I. V.; Berbís, M.Á.; Miller, M.C.; Cañada, F.J.; André, S.; Jiménez-Barbero, J.; Gabius, H.J.; Mayo, K.H. 1H, 13C, and 15N Backbone and Side-Chain Chemical Shift Assignments for the 31 KDa Human Galectin-7 (P53-Induced Gene 1) Homodimer, a pro-Apoptotic Lectin. *Biomol. NMR Assign.* **2012**, *6*, 127–129, doi:10.1007/S12104-011-9339-9/FIGURES/2.
24. Palmioli, A.; Sperandeo, P.; Bertuzzi, S.; Polissi, A.; Airoldi, C. On-Cell Saturation Transfer Difference NMR for the Identification of FimH Ligands and Inhibitors. *Bioorg. Chem.* **2021**, *112*, doi:10.1016/J.BIOORG.2021.104876.
25. Finch, R. Bacterial Resistance—the Clinical Challenge. *Clin. Microbiol. Infect.* **2002**, *8*, 21–32, doi:10.1046/J.1469-0691.8.S.3.3.X.
26. Bunduki, G.K.; Heinz, E.; Phiri, V.S.; Noah, P.; Feasey, N.; Musaya, J. Virulence Factors and Antimicrobial Resistance of Uropathogenic Escherichia Coli (UPEC) Isolated from Urinary Tract Infections: A Systematic Review and Meta-Analysis. *BMC Infect. Dis.* **2021**, *21*, 1–13, doi:10.1186/S12879-021-06435-7/TABLES/5.
27. Mulvey, M.A.; Lopez-Boado, Y.S.; Wilson, C.L.; Roth, R.; Parks, W.C.; Heuser, J.; Hultgren, S.J. Induction and Evasion of Host Defenses by Type 1-Piliated Uropathogenic Escherichia Coli. *Science (80-.).* **1998**, *282*, 1494–1497, doi:10.1126/SCIENCE.282.5393.1494.
28. Martinez, J.J.; Mulvey, M.A.; Schilling, J.D.; Pinkner, J.S.; Hultgren, S.J. Type 1 Pilus-Mediated Bacterial Invasion of Bladder Epithelial Cells. *EMBO J.* **2000**, *19*, 2803–2812, doi:10.1093/EMBOJ/19.12.2803.
29. Jones, C.H.; Pinkner, J.S.; Roth, R.; Heuser, J.; Nicholes, A. V.; Abraham, S.N.; Hultgren, S.J. FimH Adhesin of Type 1 Pili Is Assembled into a Fibrillar Tip Structure in the Enterobacteriaceae. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 2081–2085, doi:10.1073/PNAS.92.6.2081.
30. Schilling, J.D.; Mulvey, M.A.; Hultgren, S.J. Structure and Function of Escherichia Coli Type 1 Pili: New Insight into the Pathogenesis of Urinary Tract Infections. *J. Infect. Dis.* **2001**, *183*, S36–S40, doi:10.1086/318855.
31. Waksman, G.; Hultgren, S.J. Structural Biology of the Chaperone–Usher Pathway of Pilus Biogenesis. *Nat. Rev. Microbiol. 2009 711* **2009**, *7*, 765–774, doi:10.1038/nrmicro2220.

32. Sauer, M.M.; Jakob, R.P.; Eras, J.; Baday, S.; Eriş, D.; Navarra, G.; Bernèche, S.; Ernst, B.; Maier, T.; Glockshuber, R. Catch-Bond Mechanism of the Bacterial Adhesin FimH. *Nat. Commun. 2016 71* **2016**, *7*, 1–13, doi:10.1038/ncomms10738.
33. Schembri, M.A.; Kjaergaard, K.; Sokurenko, E. V.; Klemm, P. Molecular Characterization of the Escherichia Coli FimH Adhesin. *J. Infect. Dis.* **2001**, *183*, S28–S31, doi:10.1086/318847.
34. Hartmann, M.; Lindhorst, T.K. The Bacterial Lectin FimH, a Target for Drug Discovery – Carbohydrate Inhibitors of Type 1 Fimbriae-Mediated Bacterial Adhesion. *European J. Org. Chem.* **2011**, *2011*, 3583–3609, doi:10.1002/EJOC.201100407.
35. Mydock-McGrane, L.K.; Hannan, T.J.; Janetka, J.W. Rational Design Strategies for FimH Antagonists: New Drugs on the Horizon for Urinary Tract Infection and Crohn's Disease. *http://dx.doi.org/10.1080/17460441.2017.1331216* **2017**, *12*, 711–731, doi:10.1080/17460441.2017.1331216.
36. Sarshar, M.; Behzadi, P.; Ambrosi, C.; Zagaglia, C.; Palamara, A.T.; Scribano, D. FimH and Anti-Adhesive Therapeutics: A Disarming Strategy Against Uropathogens. *Antibiot. 2020, Vol. 9, Page 397* **2020**, *9*, 397, doi:10.3390/ANTIBIOTICS9070397.
37. Palmioli, A.; Panigati, M.; Bernardi, A. Glycodendron–Rhenium Complexes as Luminescent Probes for Lectin Sensing. *Org. Biomol. Chem.* **2018**, *16*, 8413–8419, doi:10.1039/C8OB01838C.
38. Palmioli, A.; La Ferla, B. Glycofunctionalization of Poly(Lactic- Co-Glycolic Acid) Polymers: Building Blocks for the Generation of Defined Sugar-Coated Nanoparticles. *Org. Lett.* **2018**, *20*, 3509–3512, doi:10.1021/ACS.ORGLETT.8B01287/ASSET/IMAGES/LARGE/OL-2018-01287J_0003.JPEG.
39. Palmioli, A.; Sperandeo, P.; Polissi, A.; Airoldi, C. Targeting Bacterial Biofilm: A New LecA Multivalent Ligand with Inhibitory Activity. *ChemBioChem* **2019**, *20*, 2911–2915, doi:10.1002/CBIC.201900383.
40. Mari, S.; Serrano-Gómez, D.; Cañada, F.J.; Corbí, A.L.; Jiménez-Barbero, J. 1D Saturation Transfer Difference NMR Experiments on Living Cells: The DC-SIGN/Oligomannose Interaction. *Angew. Chemie* **2005**, *117*, 300–302, doi:10.1002/ANGE.200461574.
41. Mirelman, D.; Altmann, G.; Eshdat, Y. Screening of Bacterial Isolates for Mannose-Specific Lectin Activity by Agglutination of Yeasts. *J. Clin. Microbiol.* **1980**, *11*, 328–331, doi:10.1128/JCM.11.4.328-331.1980.
42. Firon, N.; Ofek, I.; Sharon, N. Interaction of Mannose-Containing Oligosaccharides with the Fimbrial Lectin of Escherichia Coli. *Biochem. Biophys. Res. Commun.* **1982**, *105*, 1426–1432, doi:10.1016/0006-291X(82)90947-0.
43. Sharon, N.; Lis, H. Lectins: Cell-Agglutinating and Sugar-Specific Proteins. *Science (80-.).* **1972**, *177*, 949–959, doi:10.1126/SCIENCE.177.4053.949/ASSET/788C20DD-3CF0-40AA-9DE9-07F05F091A71/ASSETS/SCIENCE.177.4053.949.FP.PNG.
44. Unione, L.; Moure, M.J.; Lenza, M.P.; Oyenarte, I.; Ereño-Orbea, J.; Ardá, A.; Jiménez-Barbero, J. The SARS-CoV-2 Spike Glycoprotein Directly Binds Exogeneous Sialic Acids: A NMR View. *Angew. Chemie Int. Ed.* **2022**, *61*, e202201432, doi:10.1002/ANIE.202201432.
45. Lenza, M.P.; Atxabal, U.; Oyenarte, I.; Jiménez-Barbero, J.; Ereño-Orbea, J. Current Status on Therapeutic Molecules Targeting Siglec Receptors. *Cells 2020, Vol. 9, Page 2691* **2020**, *9*, 2691, doi:10.3390/CELLS9122691.
46. Rillahan, C.D.; Schwartz, E.; McBride, R.; Fokin, V. V.; Paulson, J.C. Click and Pick: Identification of Sialoside Analogues for Siglec-Based Cell Targeting. *Angew. Chemie Int. Ed.* **2012**, *51*, 11014–11018, doi:10.1002/ANIE.201205831.
47. Johannes, L.; Jacob, R.; Leffler, H. Galectins at a Glance. *J. Cell Sci.* **2018**, *131*, doi:10.1242/JCS.208884/57083.

48. Medof, M.E.; Nagarajain, S.; Tykocinski, M.L. Cell-Surface Engineering with GPI-Anchored Proteins. *FASEB J.* **1996**, *10*, 574–586, doi:10.1096/FASEBJ.10.5.8621057.
49. Liao, K.W.; Chou, W.C.; Lo, Y.C.; Roffler, S.R. Design of Transgenes for Efficient Expression of Active Chimeric Proteins on Mammalian Cells. *Biotechnol. Bioeng.* **2001**, *73*, 313–323, doi:10.1002/BIT.1064.
50. Wang, T.Y.; Leventis, R.; Silvius, J.R. Artificially Lipid-Anchored Proteins Can Elicit Clustering-Induced Intracellular Signaling Events in Jurkat T-Lymphocytes Independent of Lipid Raft Association. *J. Biol. Chem.* **2005**, *280*, 22839–22846, doi:10.1074/jbc.M502920200.
51. Paulick, M.G.; Bertozzi, C.R. The Glycosylphosphatidylinositol Anchor: A Complex Membrane-Anchoring Structure for Proteins. *Biochemistry* **2008**, *47*, 6991–7000, doi:10.1021/BI8006324/ASSET/IMAGES/LARGE/BI-2008-006324_0003.JPEG.
52. Paulson, J.C.; MacAuley, M.S.; Kawasaki, N. Siglecs as Sensors of Self in Innate and Adaptive Immune Responses. *Ann. N. Y. Acad. Sci.* **2012**, *1253*, 37–48, doi:10.1111/J.1749-6632.2011.06362.X.
53. Duan, S.; Paulson, J.C. Siglecs as Immune Cell Checkpoints in Disease. *https://doi.org/10.1146/annurev-immunol-102419-035900* **2020**, *38*, 365–395, doi:10.1146/ANNUREV-IMMUNOL-102419-035900.
54. Crocker, P.R.; Varki, A. Siglecs in the Immune System. *Immunology* **2001**, *103*, 137, doi:10.1046/J.0019-2805.2001.01241.X.
55. MacAuley, M.S.; Crocker, P.R.; Paulson, J.C. Siglec-Mediated Regulation of Immune Cell Function in Disease. *Nat. Rev. Immunol. 2014 1410* **2014**, *14*, 653–666, doi:10.1038/nri3737.
56. Gonzalez-Gil, A.; Schnaar, R.L. Siglec Ligands. *Cells* **2021**, *10*, 1260, doi:10.3390/CELLS10051260.
57. Whitney, G.; Wang, S.; Chang, H.; Cheng, K.Y.; Lu, P.; Zhou, X.D.; Yang, W.P.; McKinnon, M.; Longphre, M. A New Siglec Family Member, Siglec-10, Is Expressed in Cells of the Immune System and Has Signaling Properties Similar to CD33. *Eur. J. Biochem.* **2001**, *268*, 6083–6096, doi:10.1046/J.0014-2956.2001.02543.X.
58. Chen, G.Y.; Tang, J.; Zheng, P.; Liu, Y. CD24 and Siglec-10 Selectively Repress Tissue Damage - Induced Immune Responses. *Science (80-.).* **2009**, *323*, 1722–1725, doi:10.1126/SCIENCE.1168988/SUPPL_FILE/CHEN.SOM.PDF.
59. Chen, G.Y.; Brown, N.K.; Zheng, P.; Liu, Y. Siglec-G/10 in Self–Nonself Discrimination of Innate and Adaptive Immunity. *Glycobiology* **2014**, *24*, 800–806, doi:10.1093/GLYCOB/CWU068.
60. Bandala-Sanchez, E.; Bediaga, N.G.; Goddard-Borger, E.D.; Ngui, K.; Naselli, G.; Stone, N.L.; Neale, A.M.; Pearce, L.A.; Wardak, A.; Czabotar, P.; et al. CD52 Glycan Binds the Proinflammatory B Box of HMGB1 to Engage the Siglec-10 Receptor and Suppress Human T Cell Function. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 7783–7788, doi:10.1073/PNAS.1722056115/SUPPL_FILE/PNAS.1722056115.SAPP.PDF.
61. Yin, S.S.; Gao, F.H. Molecular Mechanism of Tumor Cell Immune Escape Mediated by CD24/Siglec-10. *Front. Immunol.* **2020**, *11*, 1324, doi:10.3389/FIMMU.2020.01324/BIBTEX.
62. Forgione, R.E.; Di Carluccio, C.; Guzmán-Caldentey, J.; Gaglione, R.; Battista, F.; Chiodo, F.; Manabe, Y.; Arciello, A.; Del Vecchio, P.; Fukase, K.; et al. Unveiling Molecular Recognition of Sialoglycans by Human Siglec-10. *iScience* **2020**, *23*, doi:10.1016/J.ISCI.2020.101231.
63. Claasen, B.; Axmann, M.; Meinecke, R.; Meyer, B. Direct Observation of Ligand Binding to Membrane Proteins in Living Cells by a Saturation Transfer Double Difference (STDD) NMR Spectroscopy Method Shows a Significantly Higher Affinity of Integrin AIIbβ 3 in Native Platelets than in Liposomes. *J. Am. Chem. Soc.* **2005**, *127*, 916–919, doi:10.1021/JA044434W/SUPPL_FILE/JA044434WSI20040914_060615.PDF.

64. Madge, P.D.; Maggioni, A.; Pascolutti, M.; Amin, M.; Waespy, M.; Bellette, B.; Thomson, R.J.; Kelm, S.; Von Itzstein, M.; Haselhorst, T. Structural Characterisation of High Affinity Siglec-2 (CD22) Ligands in Complex with Whole Burkitt's Lymphoma (BL) Daudi Cells by NMR Spectroscopy. *Sci. Rep.* **2016**, *6*, doi:10.1038/SREP36012.
65. Tan, E.; Chin, C.S.H.; Lim, Z.F.S.; Ng, S.K. HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. *Front. Bioeng. Biotechnol.* **2021**, *9*, doi:10.3389/FBIOE.2021.796991.
66. Razi, N.; Varki, A. Masking and Unmasking of the Sialic Acid-Binding Lectin Activity of CD22 (Siglec-2) on B Lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 7469–7474, doi:10.1073/PNAS.95.13.7469.
67. Steentoft, C.; Bennett, E.P.; Schjoldager, K.T.B.G.; Vakhrushev, S.Y.; Wandall, H.H.; Clausen, H. Precision Genome Editing: A Small Revolution for Glycobiology. *Glycobiology* **2014**, *24*, 663–680, doi:10.1093/GLYCOB/CWU046.
68. Narimatsu, Y.; Büll, C.; Chen, Y.H.; Wandall, H.H.; Yang, Z.; Clausen, H. Genetic Glycoengineering in Mammalian Cells. *J. Biol. Chem.* **2021**, *296*, doi:10.1016/J.JBC.2021.100448.
69. Marinova, I.N.; Wandall, H.H.; Dabelsteen, S. Protocol for CRISPR-Cas9 Modification of Glycosylation in 3D Organotypic Skin Models. *STAR Protoc.* **2021**, *2*, 100668–100668, doi:10.1016/J.XPRO.2021.100668.
70. Collins, B.E.; Paulson, J.C. Cell Surface Biology Mediated by Low Affinity Multivalent Protein–Glycan Interactions. *Curr. Opin. Chem. Biol.* **2004**, *8*, 617–625, doi:10.1016/J.CBPA.2004.10.004.
71. Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27*, 3–49, doi:10.1093/GLYCOB/CWW086.
72. Marth, J.D.; Grewal, P.K. Mammalian Glycosylation in Immunity. *Nat. Rev. Immunol. 2008 811* **2008**, *8*, 874–887, doi:10.1038/nri2417.
73. Narimatsu, Y.; Joshi, H.J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y.H.; Schjoldager, K.T.; Steentoft, C.; et al. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cell* **2019**, *75*, 394-407.e5, doi:10.1016/J.MOLCEL.2019.05.017.
74. Wisnovsky, S.; Bertozzi, C.R. Reading the Glyco-Code: New Approaches to Studying Protein–Carbohydrate Interactions. *Curr. Opin. Struct. Biol.* **2022**, *75*, doi:10.1016/J.SBI.2022.102395.
75. Clausen, H.; Wandall, H.H.; DeLisa, M.P.; Stanley, P.; Schnaar, R.L. Glycosylation Engineering. *Essentials Glycobiol.* **2022**, doi:10.1101/GLYCOBIOLOGY.4E.56.
76. Yang, Z.; Wang, S.; Halim, A.; Schulz, M.A.; Frodin, M.; Rahman, S.H.; Vester-Christensen, M.B.; Behrens, C.; Kristensen, C.; Vakhrushev, S.Y.; et al. Engineered CHO Cells for Production of Diverse, Homogeneous Glycoproteins. *Nat. Biotechnol. 2015 338* **2015**, *33*, 842–844, doi:10.1038/nbt.3280.
77. Narimatsu, Y.; Joshi, H.J.; Yang, Z.; Gomes, C.; Chen, Y.H.; Lorenzetti, F.C.; Furukawa, S.; Schjoldager, K.T.; Hansen, L.; Clausen, H.; et al. A Validated GRNA Library for CRISPR/Cas9 Targeting of the Human Glycosyltransferase Genome. *Glycobiology* **2018**, *28*, 295–305, doi:10.1093/GLYCOB/CWX101.
78. Dabelsteen, S.; Pallesen, E.M.H.; Marinova, I.N.; Nielsen, M.I.; Adamopoulou, M.; Rømer, T.B.; Levann, A.; Andersen, M.M.; Ye, Z.; Thein, D.; et al. Essential Functions of Glycans in Human Epithelia Dissected by a CRISPR-Cas9-Engineered Human Organotypic Skin Model. *Dev. Cell* **2020**, *54*, 669-684.e7, doi:10.1016/J.DEVCEL.2020.06.039.
79. Steentoft, C.; Vakhrushev, S.Y.; Vester-Christensen, M.B.; Schjoldager, K.T.B.G.; Kong, Y.; Bennett, E.P.; Mandel, U.; Wandall, H.; Levery, S.B.; Clausen, H. Mining the O-Glycoproteome Using Zinc-Finger Nuclease–Glycoengineered SimpleCell Lines. *Nat. Methods 2011 811* **2011**, *8*, 977–982, doi:10.1038/nmeth.1731.

80. Nielsen, M.I.; Stegmayr, J.; Grant, O.C.; Yang, Z.; Nilsson, U.J.; Boos, I.; Carlsson, M.C.; Woods, R.J.; Unverzagt, C.; Leffler, H.; et al. Galectin Binding to Cells and Glycoproteins with Genetically Modified Glycosylation Reveals Galectin-Glycan Specificities in a Natural Context. *J. Biol. Chem.* **2018**, *293*, 20249–20262, doi:10.1074/JBC.RA118.004636.
81. Schjoldager, K.T.; Narimatsu, Y.; Joshi, H.J.; Clausen, H. Global View of Human Protein Glycosylation Pathways and Functions. *Nat. Rev. Mol. Cell Biol. 2020 2112* **2020**, *21*, 729–749, doi:10.1038/S41580-020-00294-X.
82. Croset, A.; Delafosse, L.; Gaudry, J.P.; Arod, C.; Glez, L.; Losberger, C.; Begue, D.; Krstanovic, A.; Robert, F.; Vilbois, F.; et al. Differences in the Glycosylation of Recombinant Proteins Expressed in HEK and CHO Cells. *J. Biotechnol.* **2012**, *161*, 336–348, doi:10.1016/J.JBIOTEC.2012.06.038.
83. Moure, M.J.; Gimeno, A.; Delgado, S.; Diercks, T.; Boons, G.J.; Jiménez-Barbero, J.; Ardá, A. Selective 13C-Labels on Repeating Glycan Oligomers to Reveal Protein Binding Epitopes through NMR: Polylactosamine Binding to Galectins. *Angew. Chemie Int. Ed.* **2021**, *60*, 18777–18782, doi:10.1002/ANIE.202106056.

CHAPTER 7

MATERIALS AND METHODS

7.1 Proteins

All the proteins used in this PhD thesis were produced in our lab (galectins and siglec-10), except LSECtin that was provided by the group of Dr. Franck Fieschi (IBS, University of Grenoble). I also present herein the methodology used to obtain mammalian cells overexpressing siglec-10. In this case, the protocol lacks the purification step since intact cells were directly used for the *on-cell*-NMR experiment.

7.1.1 Galectin expression and purification in *E. coli*

All the analysed galectins (galectin-1, galectin-3 CRD, galectin-7 and galectin-8C and –N terminal domain) were expressed and purified in our laboratory (Table 7.1).

7.1.1.1 Plasmid generation

The sequence of the gene encoding for sequence of a specific galectin was inserted into a pET21 expression vector (Thermo Fisher Scientific) and, after the design of the construct, the plasmids were synthetized by GeneScript Biotech. They were amplified in our lab using competent DH5α cells and QIAprep Spin Miniprep Kit from QIAGEN.

Galectin	*Uniprot code*	*Amino acid residues*	*Vector*
Galectin-1	P09382	1-135	pET22b
Galectin-3 CRD	P17931	114-250	pET21a
Galectin-7	P47929	1-136	pET22b
Galectin-8C term	O00214	184-317	pET21a
Galectin-8N term	O00214	20-155	pET21b

Table 7.1 Galectins produced for the work presented in the PhD thesis, their Uniprot code, as well as the number of amino acid residues cloned and the vector that was employed.

pET21 plasmid is a vector for inducible expression of genes in bacteria, endowed with some fundamental elements:

- the gene for ampicillin resistance necessary for the selection of the transformed cells;
- the *lacI* gene for the synthesis of the lac repressor, LacI;

- and the Open Reading Frame (ORF, the site where the gene of interest is cloned in the plasmid) under the control of T7 promoter and the LacO operator (Figure 7.1).

The pET family of plasmids is highly employed for the expression of recombinant proteins in *E. coli*. The system is based on the conjunct action of T7 RNA polymerase and lac operon (T7lac promoter system). The BL21(DE3) *E. coli* cells used in our laboratory are a competent cell line suitable for the expression of heterologous genes. They include in their genome the sequence for the T7 RNA polymerase under the control of lacUV5 promoter, as well as *lacI* gene encoding for LacI repressor (Figure 7.1). [1]

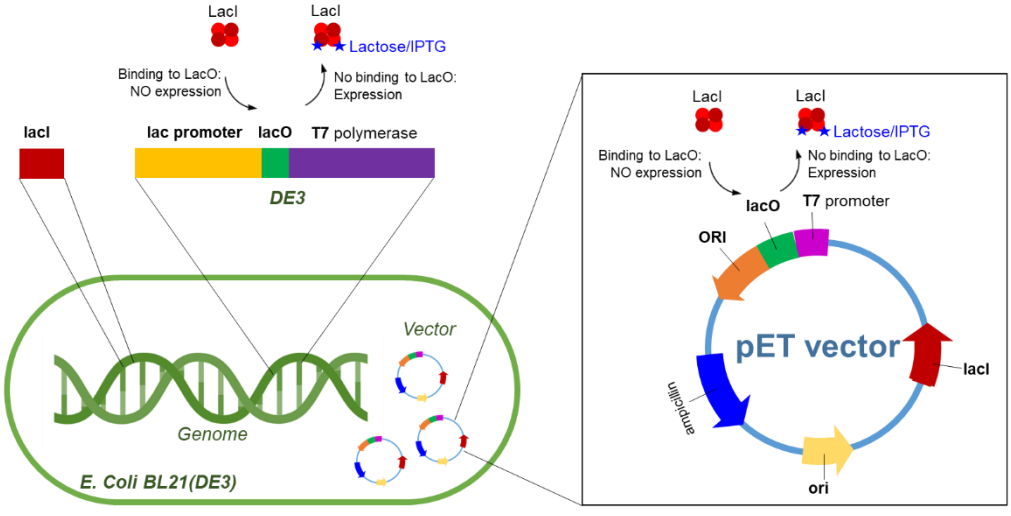


Figure 7.1. Simplified representation of the system responsible of exogenous protein overexpression in BL21(DE3) *E. coli* strain after transfection with pET vector and induction with IPTG.

In the absence of lactose (or analogous), the LacI repressor binds to the LacO operator site and prevents the expression of the gene located downstream. When lactose is present, it binds to a pocket of LacI, disrupting its binding to LacO and allowing the transcription. Therefore, in the absence of the inductor, neither T7 polymerase nor the cloned gene are expressed, since the LacI repressor binds to the upstream-located LacO of both genes. The expression of T7 polymerase is induced by addition of the lactose analogue Isopropyl-β-D-1-thio-galactopyranoside (IPTG) to the cell media that positively acts on the lac promoter, dissociating its binding to the LacI repressor. Once T7 is produced, it can drive the overexpression of the gene of interest that is contained

in the exogenous plasmid under control of the strong T7 promoter. Moreover, the site is now accessible.

With this strategy, *E. coli* is capable of producing high levels of the target protein that, after the subsequent purification steps, can be used for further experiments (Figure 7.1).

7.1.1.2 Transformation of E. coli cells using the heat-shock method

The transformation protocol, based on an induced heat shock stress to the cells, has been common for all the galectins employed in this PhD Thesis. BL21(DE3) *E. coli* competent cells were chosen as host and transformed with the respective expression vector using the heat shock method. After the addition of 10 μg of DNA to the batch of cells, the sample was left 30 min on ice. Then, the cells were placed at 42 °C for 90 s and subsequently 5 min on ice. With this action, the plasmid can be incorporated into the bacteria passing through the pores transiently formed on the bacterial wall. The cells were then diluted into 2 mL of Luria-Bertani broth (LB) media and grown at 37 °C with shaking. 2 hours later, they were incubated overnight on agar plates containing the antibiotic ampicillin at a concentration of 100 μg/mL. Under the pressure of antibiotic selection, only the cells that have internalized the expression construct can survive and form colonies that are visible after overnight incubation of the plate at 37 °C.

7.1.1.3 Galectin expression protocols

The methods followed for the expression of the galectin differ for unlabelled galectins (7.1.1.3.1)and ^{15}N labelled galectins (7.1.1.3.2).

7.1.1.3.1 Expression of unlabelled galectins

After the transformation, a single colony was selected from the plate and inoculated with 200 mL of LB and ampicillin at 100 μg/mL overnight at 37 °C with shaking. The required amount of the culture to achieve a final OD_{600} value of 0.1 was then added to 2 L of fresh LB medium containing ampicillin. Cells were grown with shaking at 37 °C until the OD_{600} value reached 0.6-1.2 and subsequently induced with 1 mM of IPTG.

The induced culture was then harvested for 3 h at 37 °C and centrifuged at 5500 rpm for 20 min. The pellet was frozen and stored at -20 °C, or purified as detailed below.

7.1.1.3.2 Expression of ^{15}N labelled galectins

For the expression of ^{15}N uniformly labelled galectins, after the first selection on the agar plate, a single colony was inoculated into 5 mL of LB medium containing 100 µg/mL of ampicillin for 6 hours at 37 °C with shaking. The culture was centrifuged at 4400 rpm for 5 minutes and the pellet was re-suspended in 1 mL of M9 unlabelled medium containing ampicillin (Table 7.2), transferred into a flask with 200 mL of the same medium and then incubated overnight at 37°C with shaking. The required amount of the overnight culture to achieve a final OD_{600} value of 0.1 was then added to 2 L of the fresh M9 labelled medium (^{15}N-NH_4Cl as the only nitrogen source) containing ampicillin (Table 7.2). Cells were grown at 37 °C until the OD_{600} value reached 0.6-1.2, then induced with 1 mM IPTG and again grown for 3 h at 37 °C. The induced culture was harvested by centrifugation at 5500 rpm for 20 min. The pellet of the ^{15}N labelled galectin was frozen and stored at -20 °C, or purified as explained in the purification section.

M9 media recipe for 1L	
Salt stock (stock 10X) For 1 L: 60 g Na_2PO_4 39 g KH_2PO_4 5 g NaCl	100 mL (final concentration 1X)
NH_4Cl (^{15}N labelled or not)	1 g (final concentration 1 g/L)
Glucose	3 g (final concentration 3 g/L)
$MgSO_4$ (stock 1M)	1 mL (final concentration 1 mM)
$CaCl_2$ (stock 100 mM)	1 mL (final concentration 0.1 mM)
Biotin (stock 10 mg/mL)	1 mL (final concentration 0.01 mg/mL)
Thiamine (stock 10 mg/mL)	1 mL (final concentration 0.01 mg/mL)
Trace elements (stock 100X) For 1 L: 5 g EDTA	2 mL (final concentration 0.2X)

0.83 g $FeCl_3$ X $6H_2O$ 84 mg $ZnCl_2$ 13 mg $CuCl_2$ x $2H_2O$ 10 mg $CoCl_2$ x $6H_2O$ 10 mg H_2BO_3 1.6 mg $MnCl_2$ x $6H_2O$	
Ampicillin (stock 100 mg/mL)	1 mL (final concentration 100 µg/mL)

Table 7.2 M9 medium recipe

7.1.1.4 Galectin purification

The purification protocol was identical for all the unlabelled and ^{15}N labelled galectins. The pellet obtained from the expression in BL21 *E. coli* cells was suspended in a lysis buffer containing 22 mM Tris-HCl (pH 7.5), 5 mM EDTA, and 1 mM phenylmethylsulfonyl fluoride (PMSF, protease inhibitor). 10 mL of the lysis buffer were used for each g of pellet. The solution was left in ice for 30 minutes with shaking. The cell suspension was lysed by sonication (60% amplitude, 12 x 20 s, with 59 s intervals between each burst) paying attention to always maintain the cells on ice during the process. The crude extract was ultracentrifuged at 35000 rpm for 1 hour at 4 °C. The soluble fraction containing the protein was loaded onto 5mL α-Lactose-Agarose resin (Sigma-Aldrich) previously equilibrated with the equilibration buffer (50 mM TRIS (pH 7.2), 150 mM NaCl). The loaded column was washed with 100 mL of equilibration buffer and then the lectin was eluted with 7 mL of the elution buffer (150 mM α-Lactose pH 7.4 in PBS 1X).

The purity of the galectins was checked by 4-12% SDS-PAGE and by LC-MS. To eliminate the lactose from the protein sample, a series of dialysis and washes with centrifuge filters (Sartorius Vivaspin 6 5000 MWCO) using fresh buffer (50 mM sodium phosphate, 150 mM NaCl, pH 7.4) were performed. The absence of lactose was checked by ^1H NMR standard experiments.

For galectin-1, galectin-8 C terminal domain, and galectin-8 FL, the final buffer and the lysis buffer also included 2 mM of dithiothreitol (DTT) reducing agent, justified by the presence of cysteine residues in those lectins, which are exposed to the solvent and could generate non-specific dimers or aggregates through intermolecular

disulphide bonds. Moreover, in the final buffer and in the lysis buffer for galectin-8, 0.1% of NaN_3 was added, to avoid bacterial growth.

The sample of unlabelled galectin-1 and galectin-3 CRD used for the cryo-microscopy and DLS measurements (Chapter 4) were further purified through size-exclusion chromatography using a Superdex 10/300 75 Increase column. The elution peaks corresponding to the dimer and monomer forms of the lectins were collected and concentrated for the following analysis.

7.1.2 Siglec-10 expression in HEK mammalian cells and purification

The soluble portion of siglec-$10_{d1\text{-}d2}$ used to obtain the *in vitro* binding epitope with the F9 glycomimetic ligand described in Chapter 6 was produced through overexpression in mammalian cells. The pHLsec vector [2] optimized for expression in human cells and containing the sequence for d1-d2 extracellular domain of siglec 10 (amino acid residues 17-537, Uniprot code Q96LC7) was designed in our laboratory and synthetized by GenScript Biotech. The transient expression of siglec-$10_{d1\text{-}d2}$ in HEK293F suspension cells (Thermo Fisher Scientific) was achieved by transfection of the DNA with the transfection reagent FectoPRO (Polyplus Transfections). 200 mL of cell culture in Freestyle medium, contained into a sterile culture flask with vented cap, were grown in a Minitron Infors HT orbital shaker incubator at 310 K, 125 rpm, 70% of humidity and 8% CO_2 until reaching 1 x 10^6 cells/mL with a viability > 97%. For transfection, 50 μg of Endotoxin-free DNA plasmid and 50 μg of FectoPRO (Polyplus) transfection reagent were added to the cells. The transfected cells were incubated with the same conditions described above for 7 days. Cells were harvested by centrifugation at 6371 x *g* for 20 min and the supernatant was filtered using a 0.22 μm Polysterene sterile filters (Corning). The supernatant was passed through a HisTrap Ni-NTA column (GE Healthcare). The eluted fractions containing siglec-$10_{d1\text{-}d2}$ were concentrated and separated on a Superdex 75 Increase 10/300 size exclusion column (GE Healthcare). Siglec-10 d1-d2's purity was checked by 4-12% SDS-PAGE and by LC-MS.

7.1.3 *On-cell* expression of FL Siglec-10 in HEK mammalian cells

For the *on-cell* NMR experiments described in Chapter 6, the FL siglec-10 was overexpressed on the membrane of HEK293F cell lines. The pHLsec vector [2] optimized for expression in human cells containing siglec-10 FL sequence was designed in our laboratory and synthetized by GenScript Biotech. 50 mL of cell culture in Freestyle medium, contained into a sterile culture flask with vented cap, were grown in a Minitron Infors HT orbital shaker incubator at 310 K, 125 rpm, 70% of humidity and 8% CO_2 until reaching 1 x 10^6 cells/mL with viability > 97%. For transfection, 15 µg of Endotoxin-free DNA plasmid and 15 µg of FectoPRO (Polyplus) transfection reagent were added to the cells. The transfected cells were incubated using the same conditions described above. 24 hours after incubation, filtered sodium butyrate (at a final concentration 5 mM) was added to the culture. 48 hours after transfection, the cells were harvested and prepared for the *on-cell* NMR experiments. For the blank experiments, in parallel to the flask of transfected cells, a second flask containing non-transfected cells was grown under the same experimental conditions.

7.2 Ligands

LacNAc, **Galili**, the **H type II**, and the **B type II** antigens were purchased from Elycityl (references GLY008, GLY074-2, GLY031-2 and GLY038-2).

All the LacNAc derivatives (multivalent monomeric scaffolds and glycopolymers) analysed in Chapter 4 (**1**-**6**), were synthetized by Prof. Pavla Bojarová and coworkers (Institute of Macromolecular Chemistry of the Czech Academy of Sciences). [3]

The thiodigalactoside (TDG) lactose analogue employed for the *in-cell* NMR experiments (Chapter 6) was purchased from Sigma Aldrich (CAS number 51555-87-4).

The GlcNAcβ1-2Man disaccharide as well as the **G0**, **LDN3** and **LDN6** N-glycans used to study the binding preferences of LSECtin (Chapter 5) were chemo-enzymatically synthesized in the laboratory of Prof. Niels C. Reichardt (CIC biomaGUNE, San Sebastian, Spain). [4,5]

The **Man_4**, **Man_6**, **Man_{18}** and **Gal_6** oligosaccharides used for the study described in Chapter 6 had been previously synthetized by using click chemistry methods in the BioOrg NMR Lab at the University of Milano-Bicocca.

The glycomimetic **F9** studied in Chapter 6 was kindly provided by Prof. James H. Paulson (Scripps Research Institute, USA). [6]

7.3 Cell lines used for *on-cell* experiments

7.3.1 Commercial cell lines

Human embryonic kidney (HEK) cells in suspension were used for the *on-cell* NMR experiments described in Chapter 6. HEK293F (Thermo Fisher Scientific) is an immortalized cell line cloned from the HEK293 cells and adapted to grow in commercial medium (FreeStyle ™).

For the *on-cell* NMR experiments (chapter 6) with *E. coli*, the standard *E. coli* BL21 [1], the *E. coli* K-12 MG1655 [7], and uropathogenic *E. coli* CFT073 (ATCC 700928) [8] strains were used.

7.3.2 Mutated cell lines

During the course of this PhD thesis, three different cell lines have been engineered to obtain *ad hoc* permanent modifications in the genome. In particular, the engineering of *E. coli* uropathogenic CFT073 strain to obtain the *ΔfimH* variant (used in Chapter 6) was carried out in the Department of Pharmacological and Biomolecular Sciences (University of Milano), under the supervision of Prof. Alessandra Polissi.

The genome mutation on HEK293 and HaCaT cells to obtain specific knock outs of glycosyltransferases were performed in the Copenaghen Center for Glycomics (CCG, University of Copenhagen), under the supervision of Prof. Hans Wandall. I participated in both protocols during my secondments.

7.3.2.1 Uropathogenic CFT073 ΔfimH

The deletion mutant CFT073 Δ*fimH E. coli* was generated to have a control of the uropathogenic strain not expressing the FimH lectin, with the consequent lack of mannose binding ability. During this workflow, the cells were transformed through electroporation.

The **electroporation protocol** consists in the exposition of the cells to an electrical pulse to create temporary pores in the cell wall. Through these pores, exogenous molecules, such as nucleic acids, can pass into cells. Before electroporation, cells need

to be made electrocompetent through a series of washing steps with ice-cold 10% glycerol. Following the protocol, electrocompetent cells were then prepared in LB and SOC media, separated in aliquots of 50 µL, and mixed with 1 µg/mL of the desired plasmid. The mixture was then transferred to a chilled electroporation cuvette and a voltage of 750 V for 6 ms was applied. 1 mL of SOC medium was added to the cells and the Eppendorf was incubated at 37 °C for 1 h. The electroporated cells were then selected on agar plates according to the type of selection (antibiotic or thermosensitive) given by the plasmid.

The procedure to obtain CFT073 Δ*fimH E. coli* strain can be divided into two parts:

- Deletion of *fimH* gene through homologous recombination, obtaining CFT073 Δ*fimH::kan* mutant;
- "cure" of the strain through elimination of the *kan* cassette, obtaining the final CFT073 Δ*fimH* mutant.

The first step was carried out using the λ Red Recombination system. [9] This methodology allows to efficiently disrupt chromosomal genes in *E. coli* with homologous recombination. The λ enzyme catalyses the homologous recombination of a given substrate (donor) with a target DNA sequence. The donor DNA has been designed to target by homology the *fimH* gene and to substitute it with the gene for kanamycin resistance (*kan* cassette). Moreover, the *kan* cassete was designed to be flanked by FRT sequences. To generate this fragment, a standard PCR was performed using the pkD4 plasmid as template (carrying *kan* resistance) and two hybrid primer paisr. The primers were synthetized by MWG Eurofins and designed as follows:

- *AP699*
  (5′CATTCAGGCAGTGATTAGCATCACCTATACCTACAGCTGAACC CAAAGAGGTGTAGGCTGGAGCTGCTTC-3')
- *AP700*
  (5′TAGCTTCAGGTAATATTGCGTACCTGCATTAGCAATGCCCTGTG ATTTCTATGGGAATTAGCCATGGTCC-3′).

APP699 and AP700 include the upstream and downstream *fimH* homology flanking regions and the priming sequences for pKD4 plasmid (underlined). The final amplified fragment was purified after standard PCR.

The competent CFT073 cells were transformed with Red helper pKD46 plasmid and the selection was carried out at 30 °C (since pKD46 is temperature-sensitive). Then, CFT073/pKD46 cells were made competent and transformed with 150 ng of the PCR-purified plasmid. The CFT073 Δ*fimH::kan* mutants were colony-purified on LB with kanamycin at 37 °C. The replacement of the *fimH* gene with the *kan* cassette was verified by colony-PCR.

To be cured from the *kan* resistance, CFT073 Δ*fimH::kan* cells were transformed by electroporation with pCP20 plasmid, which confers ampicillin resistance, temperature-sensitive replication, and thermal activation of flippase recombinase (FLP) synthesis at 42 °C. FLP allows FRT-flanking *kan* gene's excision through recombination at the FRT sites. After the transformation of CFT073 Δ*fimH::kan* cells with pCP20 by electroporation, the resistant transformants were selected at 30 °C. To allow simultaneous FLP induction and loss of the pCP20 plasmid, the colonies were further purified non-selectively at 42 °C. The final CFT073Δ*fimH* mutant clones were tested for loss of kanamycin and ampicillin resistance and confirmed by colony-PCR.

7.3.2.2 Glycosylation mutant of HEK293 and HaCaT cells

The glycosylation HEK293 and HaCaT mutants used for flow cytometry experiments described in Chapter 6 were produced by using the CRISPR/Cas9 technology [10–14]. While the library of HEK293 mutants was already available in the Copenhagen Center for Glycomics, I directly participated in engineering the HaCaT cells.

Three different mutant cells lines were produced:

- HaCaT KO *mgat1/c1galt1*
- HaCaT *KO mgat1/b4galt5/6*
- HaCaT *KO c1galt1/b4galt5/6*

The HEK293 cells mutants used for FACS, has been previously engineered with the same genetic mutations.

The workflow can be summarized as follow:

1) **Guides**: the CRISPR/Cas9 gene editing requires the design of a guide RNA (gRNA), which was carried out using the CHOPCHOP prediction tool. The best guide for each

target gene was chosen among those proposed in order to have a minimal number of off targets and the best calculated efficiency to induce disruption of the gene target. Together with the guide (forward and reverse) two additional *ad hoc* primers were predicted by the software. Guides and primers' sequence were exported and adjusted with the following nucleotide additions:

- Oligo_1F (Guide F) CACCG-
- Oligo_1R (Guide R) AAAC-C
- 1C_FamFw AGCTGACCGGCAGCAAAATTG-
- 1B_Rev

The sequences were synthetized by TAG Copenhagen A/S (Frederiksberg, Denmark).

2) **Lentiplasmid construct with the oligo of interest**: The lentiplasmid used for this protocol is a Lenti-CRISPR-v2 puro vector (Addgene). The Oligo_1F and Oligo_1R were annealed with a PCR Annealing program of 95 °C 5 min, ramped down to 25 °C at a rate of 5 °C min^{-1}, and then diluted 1:200 in milliQ water. In parallel, the lentivector was digested with the Bsmb1 restriction enzyme and left 2 h at 37 °C. The annealed oligos and the digested vector were ligated using T4 DNA ligase (ThermoFisher Scientific) at RT for 2.5 h.

3) **Amplification of the lentiplasmid:** The competent *E. coli* cells were transformed using the heat shock method and the ligation mix. The day after, 15 positive colonies were selected, lysated, and used to setup a standard PCR colony. The PCR product was checked on 2.5% agarose gel and then 3 of the positive colonies were inoculate on 200 mL of LB (and selection) overnight at 37 °C with shaking. The amplified DNA was purified following the Midi Prep protocol (Kit Endotoxin-free plasmid DNA purification NucleoBond Xtra Midi EF/Maxi EF).

4) **Production of the lentiviral particles in HEK293T cells**: HEK293T (purchased bu ATCC) cells in good shape were transfected with the packaging vector (psPAX2, Addgene), envelop vector (pVSV-G, Addgene), and the lentiplasmid. The transfection was performed with the PEI (…) transfected agent. Three days later, the lentiviral

particles accumulated in the culture medium were harvested and filtered through 0.45 µm microfilter (Sartorius).

5) **Transduction**: the target cells for the 30% confluent knock-out (HEK or HaCaT) were transducted with the filtered media containing virus particles and polybrene to enhance the transduction efficiency. At this point, the lentivirus particles transfer their genetic content into the host cells (guide and CRISPR/Cas9 machinery). The media was frequently changed for 2-4 days. Then, the antibiotic selection (1 week) was started and only cells containing the lentivirus genome survived. During the antibiotic selection steps, the bulk of cells was frequently checked by **IDAA** (Indel Detection by Amplicon Analysis). IDAA provides single base resolution and allows selecting indels resulting in frameshifts and functional knock-outs (KOs). After the antibiotic selection, the cells were scaled down to single-cell into 96 wells and, through IDAA, a few good pure clones were identified, isolated, and scaled-up. The final KO cells were maintained in culture into T175 wells and used to carry out further phenotypical tests before being used for specific experiments.

7.4 NMR

The NMR experiments performed at CIC BioGUNE were acquired using Bruker AVANCE 2 600 MHz or Bruker AVANCE 2 800 MHz NMR spectrometers (Bruker Inc.; Billerica, MA, US) equipped with a TCI cryo-probe.
The NMR experiments recorded at the bioNMR lab of the University of Milano-Bicocca (Chapter 4 and Chapter 6) were performed using a Bruker AVANCE III 600 MHz NMR spectrometer, equipped with a 5 mm QCI cryo-probe.

All the NMR samples were prepared in a final volume of 500 or 400 µL and transferred to precision 5 mm or Shigemi NMR tubes, respectively (New Era Enterprises, Vineland, USA). The pH of the buffer was measured with a Crison Basic 20 pH meter (Crison Instruments SA, Barcelona, Spain) and adjusted with the required amount of NaOH and HCl or NaOD and DCl.

7.4.1 ^{1}H-NMR titration

^{1}H-NMR titrations were employed to study the interaction of LSECtin with the disaccharide, the non-elongated N-glycan, **LDN3** and **LDN6** (Chapter 5). The soluble CRD portion of LSECtin was dissolved in deuterated buffer containing tris-d_{11} 25 mM, NaCl 150 mM, $CaCl_2$ 4 mM, dithiothreitol-d_{10} (DTT-d_{10}) 2 mM at pD 8.4.

The set of experiments for the disaccharide and the N-glycan were acquired using the 600 MHz spectrometer, with LSECtin at a concentration of 120 µM and 298 K. Starting from the apo sample containing only LSECtin, a series of ^{1}H-NMR spectra was recorded using (lectin:ligand ratios of 1:0, 1:1, 1:2, 1:5, 1:10, and 1:20. Obviously, for the last point of the titration with ratio 1:20, the actual concentration of the ligand in the NMR tube was 2.4 mM. The sequence zgesgp was selected from Bruker's library, using 32 scans.

The titrations for **LDN3** and **LDN6** were acquired using the 800 MHz spectrometer, with LSECtin at a concentration of 82 µM and 298 K. The same methodology was adopted and ^{1}H-NMR spectra recorded. LDN3 was titrated with lectin:ligand ratios of 1:0, 1:1, 1:2, 1:4, and 1:20; while the ratios used for LDN6 were 1:0, 1:2, 1:4, 1:6, and 1:34. The same sequence and number of scans were employed.

7.4.2 Saturation Transfer Difference NMR (STD NMR)

STD-NMR have been extensively used along this PhD Thesis. [15,16] In all the experiments, the final STD NMR spectra were obtained by subtracting the on-resonance spectrum from the off-resonance spectrum. The analysis was carried out using the ^{1}H NMR signal with the strongest STD intensity as reference (100% of STD effect). On this basis, the relative STD intensities for the other protons were estimated from the comparison of the corresponding integrals.

The STD NMR Amplification Factor (STD-AF) and the percentage of STD (STD%) were obtained comparing the intensities of the STD NMR spectrum to the off-resonance one. Reference experiments were also acquired for samples containing only the protein as well as only the ligands under the same experimental conditions to assess the authenticity of the binding.

In Chapter 3, the samples of **galectin-1** were prepared in deuterated phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, pD 7.4), using 2 mM of DTT-d_{10}, and a standard lectin:ligand ratio of 1:50 or 1:100. The concentration of the lectin was 50 μM. For **H type-II antigen**, the ratio was 1:138. The experiments were recorded using the 600 MHz NMR spectrometer at 298 K. The stddiff.3 sequence was chosen from the Bruker library, which includes spoil and T2 filters to remove the background protein signals. The STD NMR spectra were acquired with 1024 scans, 2 s of saturation time and 2 s of relaxation delay. The spin-lock filter was 20 ms. The on- and off-resonance spectra were registered in the interleaved mode with the same number of scans. The on-resonance frequency was set for the aliphatic region between δ 0.55 ppm and δ 0.85 ppm and between δ 7.67 ppm and δ 7.73 ppm for the aromatic region. The off-resonance frequency was always set at δ 100 ppm.
No signals were detected in the blank STD NMR spectra of the ligands alone, with exceptions for the acetyl and methyl group of GlcNAc, GalNAc and Fuc moieties respectively, which displayed weak STD NMR intesnsities likely due to direct irradiation or slow relaxation effects.

For the study of the interaction of **galectin-1 and galectin-3 (CRD) with the multivalent sugar ligands** (Chapter 4), the STD NMR experiments were performed using the 600 MHz NMR spectrometer equipped with the QCI cryo-probe at 298 K. The samples were prepared in deuterated phosphate saline buffer (50 mM sodium phosphate, 150 mM NaCl, pD 7.4). To the sample containing galectin-1, 2 mM DTT-d_{10} was added to the buffer. The ligand/lectin ratio was always 1:30 with both galectin-1 (100 μM) and Gal-3 (50 μM). The stddiffesgp.3 sequence was chosen from Bruker library, with spoil and T2 filter, as well as excitation sculpting. The STD NMR spectra were acquired with a total of 1024 scans, 2 s of saturation time and 2 s of relaxation delay. The spin-lock filter was 30 ms. The on- and off-resonance spectra were registered in the interleaved mode with the same number of scans. The on-resonance frequencies were set at δ 0.00 ppm, δ -0.5 ppm, or δ -1 ppm, while the off-resonance frequency was δ 100 ppm.

The STD-NMR experiments for **LSECtin (CRD) and the disaccharide** (Chapter 6) were acquired using the 600 MHz NMR spectrometer at 310 K. The samples were prepared in deuterated phosphate saline buffer (Tris-d_{11} 25 mM, NaCl 150 mM, $CaCl_2$ 4 mM, DTT-d_{10} 2mM at pD 8.4) with a lectin/ligand ratio of 1:70, with LSECtin at 120 μM. The stddiffesgp.3 sequence was chosen from Bruker library with spoil and T2 filters and excitation sculpting. The STD NMR spectrum was recorded with a total of 1024 scans, 2 s of saturation time and 2 s of relaxation delay. The spin-lock filter was 40 ms. The on- and off-resonance spectra were registered in the interleaved mode. The on-resonance frequency was set at δ 0.83 ppm for aliphatic irradiation and at δ 6.8 ppm for the aromatic alternative, with the off-resonance frequency at δ 100.

For the analysis of the recognition event between **LSECtin (CRD) and the N-glycans** (Chapter 6), the STD-NMR experiments were carried out using the 800 MHz NMR spectrometer at 310 K. The samples were prepared in deuterated phosphate saline buffer (Tris-d_{11} 25 mM, NaCl 150 mM, $CaCl_2$ 4 mM, DTT-d_{10} 2mM at pD 8.4). The ligand/lectin ratio was set 1:34, with LSECtin at 120 μM. The stddiffesgp.3 sequence was again chosen in the interleave mode, with 2880 scans, 2 s of saturation time, 15 s of relaxation delay, and a spin-lock of 40 ms. The on-resonance frequency was set at δ 0.83 ppm or δ 0.6 ppm for the aliphatic and δ 6.8 ppm for the aromatic irradiations, with the off-resonance frequency at δ 100 ppm.

7.4.2.1 STD-NMR experiments on-cell

In Chapter 6, *on-cell* STDs were carried out both in *E. coli* and in HEK293 cells. The preparation of the NMR samples is rather different compared to that required for the *in-vitro* samples and, therefore it is described in detail in this section.

For the ***on-cell* STD NMR experiments with *E. coli***, the cultures for CFT073, CFT073*ΔfimH* mutant and BL21 were grown at 37 ºC until the OD_{600} value reached 0.5–0.7 (mid-logarithmic phase). Bacterial cells were harvested and washed four times in 10 mM NaH_2PO_4 (pH 7.2) by centrifuging 10 min at 6000 *g* and finally re-suspended in the same buffer to a final concentration of 1–5 × 10^9 cells/ml. The number of living cells was evaluated through serial dilutions of the sample plated on solid LB medium and incubated at 37 ºC for 16–18 h. Cells were harvested by centrifugation and re-

suspended in deuterated PBS at a concentration of 2×10^9 cells/mL; the ligands were added at this point (1 mM for the core-scaffold and **Man_6**, and 0.25 mM for **Man_{18}**). Then, the final sample was transferred into a NMR tube.
The NMR experiments were acquired directly using the 600 MHz NMR spectrometer equipped with the QCI cryo-probe. STD NMR spectra were acquired at 298K using the standard pulse sequence from the Bruker's library with an on-resonance frequency of δ 0.0 ppm and an off-resonance frequency of δ 30.0 ppm. A train of 40 Gaussian-shaped pulses of 50 ms each was employed, separated by a 1 ms delay, with 256 scans. A spin-lock of 30 ms was applied to erase the broad cell resonances.

For the ***on-cell* STD NMR experiments with the *mammalian cells*** (Chapter 6), the cultures of HEK293F cells were grown 48 hours after transfection and then harvested. The cells were counted and the precise volume required to have $2x10^6$ total cells was picked up from each flask. The withdrawn cells were centrifuged for 5 min at 800 rpm and room temperature (RT). The pellet was re-suspended in 1 mL of PBS 1X pH 7.4. The sample was centrifuged again and the pellet was re-suspended in 1 mL of deuterated PBS pD 7.4. After a last centrifugation, the cells were re-suspended into 500 μL of deuterated PBS pD 7.4 and transferred into the NMR tube (with or without the ligand).
The NMR experiments were acquired immediately after the sample preparation in 800 MHz NMR spectrometer. The stddiffesgp.3 sequence was again chosen (with spoil and T2 filters, and excitation sculpting). The STD NMR spectra were acquired at 286K with 1080 scans, 2 s of saturation time and 2.4 s of relaxation delay. The length of experiment (2h) was set in order to guarantee the vitality of the cells in the tube (mortality <10%). The spin-lock filter was set 30 ms. The on- and off-resonance spectra were registered in the interleaved mode with the on-resonance frequency at δ -0.8 ppm and the off-resonance at δ 100 ppm. ´
The **periodate treatment** was performed resuspending HEK293F cells in PBS pH 4 containing freshly dissolved 2mM $NaIO_4$ and incubated 4°C in the dark for 30 minutes. After 30 minutes, the excess of periodate was quenched by adding 10 uL of 20% deuterated glycerol followed by immediate washing with deuterated buffer optimised for cell NMR experiments.

7.4.3 ^{1}H-^{15}N HSQC-based titrations

Heteronuclear Single Quantum Coherence (HSQC) NMR experiment allows to correlate ^{15}N nuclei with the attached ^{1}H through the one bond scalar coupling $^{1}J_{NH}$. Through this receptor-based NMR method the perturbations at different regions of the receptor (in terms of changes in the chemical shifts and/or intensities of the cross peaks) upon ligand binding can be monitored.

7.4.3.1 Chemical Shift Perturbation (CSP) Analysis

For the Chemical Shift Perturbation (CSP) NMR analysis, uniformly ^{15}N-labelled proteins were employed. [17] Different concentrations of the selected ligand were systematically added to the protein sample and, at each point, a ^{1}H-^{15}N HSQC NMR experiment was acquired. The set of spectra was then analysed using the CcpNMR Analysis 2.4.2 software. [18] The average of ^{1}H and ^{15}N CSP were calculated for each NH groups of the protein backbone using the formula [19]:

$$\Delta\delta\ (ppm) = \left(\frac{[\Delta\delta H_2 + (0.14\Delta\delta N)^2]}{2}\right)^{\frac{1}{2}} \qquad (1)$$

The results were plotted in graphics and the standard deviation was calculated and represented as a line in the plot.

The ^{1}H-^{15}N HSQC experiments recorded for the CSP analysis of **galectin-1 with BGA** presented in Chapter 3, were acquired using the 800 MHz spectrometer equipped with the TCI cryo-probe. The samples were prepared using ^{15}N-labelled galectin-1 at 100 μM dissolved in 90% phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, 2 mM of DTT-d_{10}, pH 7.4) and 10% of D_2O. The titration was carried out at 298K with the following protein:sugar ratios: 1:0.5, 1:1, 1:3, 1:5, 1:10, and 1:15. At each point, a standard HSQC spectrum (Bruker library) was recorded (80 scans) with 200 (t1) x 1024 (t2) complex data points in ^{15}N and ^{1}H dimensions.

The ^{1}H-^{15}N HSQC experiments for monitoring the interaction between **galectin-1 and galectin-3 CRD with ligands 1-3** (Chapter 4) were acquired using the 600 MHz

spectrometer equipped with the QCI cryo-probe. Uniformly ^{15}N-labelled galectins were employed at 100 μM for galectin-1 and 50 μM for galectin-3 CRD. The buffer was 90% phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, pH 7.4) and 10% of D_2O. 2 mM of DTT-d_{10} was added to the buffer of galectin-1. The titration of ligands **1–3** employed protein:sugar ratios of 1:0, 1:0.1, 1:0.25, 1:0.5, 1:1, 1:2.5, and 1:5. The ^{1}H-^{15}N HSQC experiments were recorded at 298K using the standard Bruker sequence, with 192 (t1) and 2048 (t2) complex data points for ^{15}N the and ^{1}H dimensions, respectively and 64 scans.

7.4.3.2 Cross-peak intensity analysis

The ^{1}H-^{15}N HSQC experiments described for the CSP analysis were also used to follow the cross peaks' intensity variation during the titration of ligands **2-4** to galectin-1 and galectin-3 CRD (Chapter 4).

The experiments were acquired using the 800 MHz spectrometer by adding 1.5 equivalents of each ligand to a sample containing only galectin-1 and 4.5 equivalents to the sample of galectin-3 CRD. This addition drastically diminished the number of detectable cross peaks of the lectins). The subsequent recovery of the signals' intensity was monitored by adding 5, 10, 20, 50, and 75 equivalents of LacNAc (used as competitor of the target ligands) for galectin-1 and 1, 5, 10 and 15 equivalents for galectin-3 CRD. The ^{1}H-^{15}N HSQC experiments were recorded at 298K using the standard Bruker sequence (32 scans), with 256 (t1) and 1536 (t2) complex data points for the ^{15}N and ^{1}H dimensions, respectively. The experiments were analysed using the CcpNMR Analysis 2.4.2 software. [18] The intensity lists for each cross peak was obtained for all the spectra and processed to create a histogram chart. The total intensities were normalized to those measured in the spectra recorded for the lectin alone, which displayed the maximum intensities (given the 100% value for each peak). The intensity plots to be analysed were obtained applying the formula:

$$\Delta I = \frac{Ii - If}{\Delta Imax} \qquad (2)$$

7.4.4 *In-cell* ^{1}H-^{15}N HMQC NMR with zebrafish oocytes

The *in-cell* NMR experiments described in Chapter 6 required a special sample preparation. For each trial (three in total), two different samples were prepared and immediately sent to Biobide (San Sebastian, Spain) to be microinjected into living wild type Zebrafish (*Danio rerio*) oocytes:

- Control: ^{15}N labelled galectin-7 at a concentration of 2-2.5 mM
- Sample: ^{15}N labelled galectin-7 at a concentration of 2-2.5 mM with 30 or 60 equivalents of thiodigalactoside (**TDG**).

The protein was washed against PBS 1X to remove the remaining lactose that still was present after the purification process. 10% of D_2O and 0.01% of phenol red solution were added to the final 40 µL-samples. The presence of phenol red is required to visually guide the operator during the microinjection process. The same day of the preparation, the lectin were shipped to Biobide to be microinjected the following day.

The injections into the oocytes was performed by Andrea Weiner in Biobide's facilities, according to EU standards of animal welfare on animal used for scientific purposes (2010/63/EU), compiled with national regulations for the care of experimental animals and were approved as described in national regulations (RD 53/2013) by the local and regional committees.

To obtain the zebrafish oocytes, adult female zebrafish were individually arranged in breeding tanks the day after the injection. On the morning of the injection, the fishes were anesthetized and the oocytes were obtained by abdominal stripping. The oocytes were collected in E3 media (specific for Zebrafish embryo) and cleaned of dead or misshapen units.

The injections were performed following an established procedure: 3 nL of the protein samples were injected into a single oocyte's yolk using a glass needle that penetrates the chorion (Figure 7.2). A total of 700 oocytes for each trial (350 with the control solution and 350 with the sample) were microinjected. After the injection, oocytes were allowed to recover in E3 media at 28.5 ºC in the incubator. 1 h post injection, damaged or not well injected oocytes (those in which phenol red solution was localized in the yolk rather than in the blastodisc) were removed. Well injected oocytes were

transferred to a 15 mL falcon tube containing E3 buffer with 1.5% Ficoll and immediately delivered to CIC bioGUNE.

The oocytes were then transferred into a 5 mm Shigemi NMR tube using a plastic Pasteur pipette with large tip, taking care not to directly bump them during the process. The NMR tubes were filled to the standard volume height of 400 µL and 10% of D_2O was added (Figure 7.2).

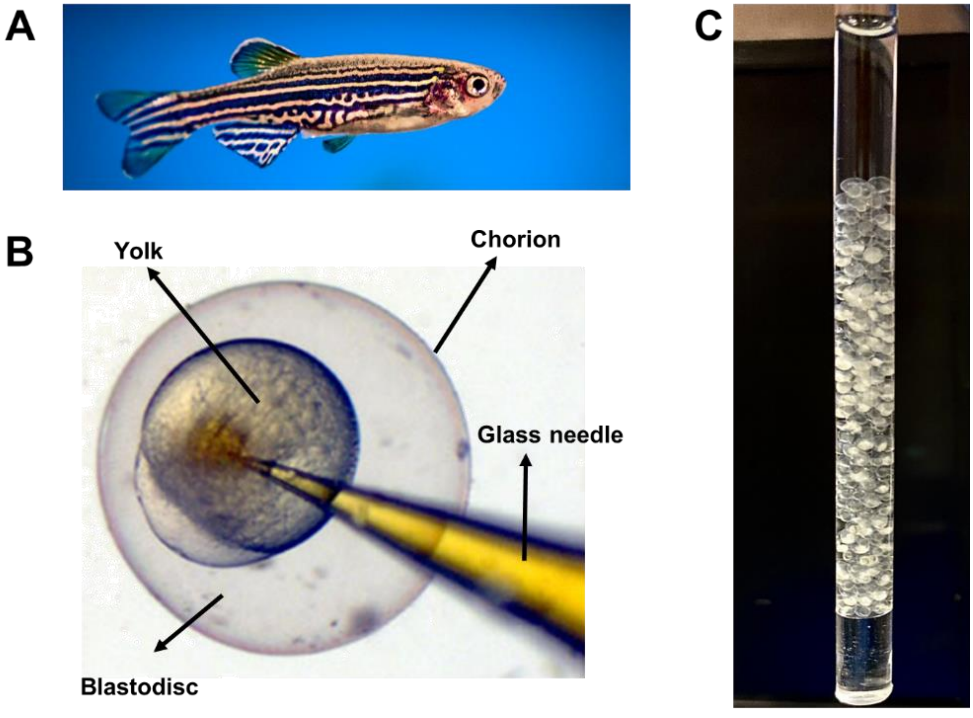


Figure 7.2. A) Zebrafish oocyte. B) Zebrafish oocyte during microinjection. Yolk, chorion and glass needle used to deliver the sample are pointed out. C) 5 mm Shigemi tube filled with injected Zebrafish oocytes.

Then, ^{1}H-^{15}N HMQC experiments were recorded (64 scans) using the 800 MHz spectrometer at 301K, using an in-house sequence, with 128 (t1) and 640 (t2) complex data points for the ^{15}N and ^{1}H dimensions, respectively. The analysis of the cross peak intensities of the spectra was performed as detailed in Chapter 7.4.3.2.

7.4.5 CLEAN Chemical Exchange (CLEANEX-PM) experiments.

CLEANEX-PM experiments [20] presented in Chapter 3 were acquired using the 800 MHz spectrometer equipped with the TCI cryo-probe at 298K. These experiments

provide information on the exchange rates of NH groups of the protein with the bulk water. The NMR samples were prepared using uniformly ^{15}N-labelled galectin-1 (1 mM) in 90% phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, 2 mM DTT-d_{10}, pH 7.4) and 10% D_2O. The CLEANEX spectra were recorded, in a pseudo-3D mode, for galectin-1 alone and in presence of **LacNAc** (12 equivalents) and the **B-type II antigen** (10 equivalents) (Figure 7.3).

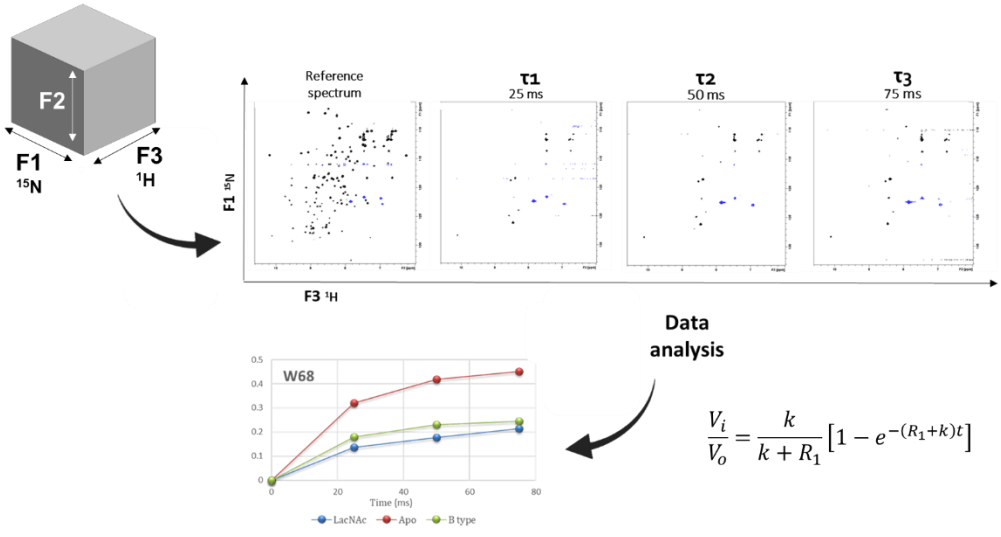


Figure 7.3. Visual representation of the process followed for the acquisition and analysis of CLEANEX-PM experiments.

The setup was optimized using 2048 (t3) x 128 (t1) points for the ^{15}N and ^{1}H dimensions, respectively, and 4 points in the t2 dimension, corresponding to 3 different mixing times (25, 50, and 75 ms) and the reference spectrum. The ratio between the peak intensities in the CLEANEX spectra (V_i) and the reference spectra (V_0) provides the estimation of the exchange rate. The analysis was performed with the CcpNMR Analysis 2.4.2 software [18] and the obtained ratios were fitted as function of the mixing time. The exchange rates, k_{ex}, were obtained by fitting to the equation (3):

$$\frac{V_i}{V_o} = \frac{k}{k + R_1}\left[1 - e^{-(R_1+k)t}\right] \qquad (3)$$

In the formula, R_1 is the effective NH relaxation rate during CLEANEX.

7.4.6 ^{1}H-^{15}N long-range HMQC experiments.

The investigation of the behaviour of the histidine (His) side chains, to understand the role of His44 and His52 of galectin-1 during the binding event, was carried out using long-range HMQC experiments (Chapter 1) [21,22]. This experiment allows to detect $^{2}J_{NH}$ correlations and therefore, to assess the tautomeric form of the imidazole ring of a given His residue. The NMR samples were prepared using uniformly ^{15}N-labelled galectin-1 at 160-260 μM in 90% phosphate-buffered saline (50 mM sodium phosphate, 150 mM NaCl, 2 mM DTT-d_{10}, pH 7.4) and 10% D_2O. The HMQC spectra were recorded at 298 K for galectin-1 in its *apo* form and for galectin-1 in presence of **LacNAc** and the **H type-II** antigen with a 1:10 lectin:ligand ratio. The experiment (96 scans) was set with 1024 (t2) x 160 (t1) data points for the ^{1}H and ^{15}N dimensions, respectively.

7.4.7 ^{15}N CPMG Relaxation Dispersion NMR experiments.

The dynamic features of galectin-1 and galectin-3 CRD in the presence and absence of the ligand were monitored by transverse relaxation dispersion NMR experiments [23] in the 800 MHz spectrometer (Chapter 3).

The NMR samples were prepared using uniformly ^{15}N-labelled galectin-1 or galectin-3 at 450 and 650 μM in 90% phosphate buffer saline (50 mM sodium phosphate, 150 mM NaCl, pH 7.4) with 10% D_2O. DTT-d_{10} 2 mM was added as reducing agent only for galectin-1. Different datasets were collected for galectin-1 and galectin-3 CRD apo and in presence of ligands **LacNAc** and **the B-type II** antigen, using a protein:ligand ratio of 1:20.

The experiments were carried out using a relaxation compensated pulse pseudo 3D-CPMG sequence (80 ms of total CPMG time) and variable effective fields of 25, 50 (x2), 75, 100, 125, 150, 200, 250, 350, 450, 600, 800 (x2), and 1000 Hz. A data matrix of 2048 (t3) x 180 (t2) was recorded for the ^{1}H and ^{15}N dimensions, respectively, while 16 points in the F1 dimension were collected. The data were analysed using the CcpNMR Analysis 2.4.2 software [18] and fit to the Carver-Richards equation using in-house Matlab© scripts. Finally, a collective fitting was done using different clustering residues.

7.4.8 2D ROESY experiments

To detect the presence of chemical exchange-mediated cross peaks, 2D ROESY (Rotating frame Overhause Effect SpectroscopY) experiments were also carried out for LSECtin in the presence of the disaccharide and the non-elongated **G0** glycan (Chapter 5) using the 800 MHz spectrometer with the TCI cryo-probe. The samples were prepared in deuterated phosphate-buffer saline (buffer (Tris-d_{11} 25 mM, NaCl 150 mM, $CaCl_2$ 4 mM, DTT-d_{10} 2mM at pD 8.4). The lectin:ligand molar ratio was 1:10 (with LSECtin CRD at 120 μM). The ROESY-NMR experiments were acquired at 298K, using an in-house sequence with 4096 (t2) x 256 (t1) points, 96 scans and 150 ms of mixing time.

7.4.9 The estimation of kinetic parameters from EXSY experiments

2D-NOESY experiments were acquired in the 800 MHz spectrometer equipped with the TCI cryo-probe. The samples were those described above for the ROESY experiment (Chapter 7.4.8). All experiments were acquired at 298 K with an in-house NOESY sequence. Two 2D-NOESY (0 and 100 ms) were recorded for each sample (LSECtin:disaccharide and LSECtin:**G0**) with 32 scans. Once the exchange cross peaks had been identified by analysing the ROESY experiments, cross- and diagonal-peaks were integrated in the NOESY experiments and the obtained intensities were analysed using the ExsyCalc software, which allowed deducing the pseudo-first-order exchange rates (k_{on} and k_{off}). [24,25]

7.5 Isothermal Titration Calorimetry (ITC)

The Isothermal Titration Calorimetry (ITC) experiments presented in Chapter 3 were performed using a MicroCal PEAQ-ITC calorimeter. The samples containing galectin-1 were concentrated at 100-200 μM and then buffer-exchanged to PBS 1X pH 7.4, using 1 mM TCEP as reducing agent. The sugar stocks (5-9 mM) were prepared dissolving the powder in the same buffer. Both protein and sugar solutions were degasified right before the experiment was carried out. 300 μL of the protein sample were placed (for each measurement) into the ITC cell by using a syringe, with extreme care to avoid the generation of bubbles. The reference cell of the calorimeter was filled

with fresh degasified water and changed after 2 days of work on the machine. 70 µL of the solution containing the ligand were placed in a small Eppendorf and then located in the dedicated position of the instrument. The experiments were set up using the MicroCal PEAQ-ITC Control Software through modification of the stir speed (750 rpm), initial delay (s), number of injections, volume (µL) and duration (s) of each injection (s) and the spacing between them (s). The temperature was 298K and the Reference Power (µcal/s) was set at a value of 10.

During the automated experiments, small amounts of the sugar solution (from 1 up to 3 µL) were withdrawn and added into the cell containing the protein solution. Since the released or absorbed heat is measured at each titration point by the instrument, binding constants (K_D), the reaction stoichiometry (n), the enthalpy (ΔH) and the entropy (ΔS) of the recognition process can be accurately determined. The analysis of the curves was performed with MicroCal Origin 7 Analysis Software. K_Ds and the corresponding thermodynamic parameters were obtained from the fit of the titration profile to a single-site binding and to a sequential binding model.

7.6 Cryo-EM

The Cryo-EM experiments described in Chapter 4 were performed at the Electron Microscopy and Macromolecular Crystallography Platform of CIC bioGUNE under the supervision of Professor Nicola Abrescia and Dr. Isaac Santos. Different samples were prepared as follows:

- Galectin-3 CRD at 0.5 mg/mL;
- Galectin-1 at 0.5 mg/mL;
- Ligand **4** at 0.17 mg/mL;
- Galectin-3 CRD at 0.5 mg/mL with ligand **4** at 0.17 mg/mL (ratio 1:5);
- Galectin-1 at 0.25 mg/mL and ligand **4** at 0.13 mg/mL (ratio 1:5).

In all cases, the buffer was PBS 1X with pH 7.4. 4 µL of each sample were pipetted onto R2/2 cupper 300-mesh (Quantifoil) grids prior to vitrification. The grids had previously been plasma-cleaned, using the BAL-TEC MED 020 coating system. The

samples were vitrified using a Leica EM GP2 plunge freezer by preincubating them in a chamber at 95% of humidity and 281K for 30 s (blotting conditions: 1.5 s and 43 mm of offset). The images were acquired with a JEM-2200FS/CR (JEOL Ltd., Tokyo, Japan) electron microscope, operating at 200 kV at liquid nitrogen temperature, and equipped with a K2 Summit direct detection camera (Gatan, Inc., Pleasanton, CA, USA). Dose fractionated movies were recorded with the Gatan Digital MicrographTM software and the motion correction of frames was achieved within the same software. The movies were collected at a defocus range from -1.8 µm to -3.0 µm with a final dose of ~40 e–/Å2 at a nominal magnification of 30,000X, producing a pixel size of 1.28Å at the specimen.

7.7 Dynamic Light Scattering (DLS)

The Dynamic Light Scattering (DLS) measurements reported in Chapter 4 were carried out using a Malvern Nano-S Zeta-Sizer spectrometer (Malvern Instruments, Worcestershire, UK). The hydrodynamic radii of different mixtures were measured by quasi-elastic light scattering employing standard acryl-cuvettes. The employed galectin samples were:

- Galectin-3 CRD at 0.5 mg/mL;
- Galectin-1 0.5 at mg/mL;
- Galectin-3 CRD at 0.5 mg/mL and ligand **4** at 1.5 mg/mL (ratio 5:1);
- Galectin-1 0.5 mg/mL and ligand **4** at 1.5 mg/mL (ratio 5:1).

All samples were suspended in PBS 1X buffer pH 7.4 and the experiments were acquired at RT.

7.8 Yeast Agglutination Assay

A yeast agglutination assay was performed to test the ability of the *E. coli* strains presented in Chapter 6 to prevent the FimH-mediated yeast agglutination. [26] Thus, common baker's yeast (*Saccharomyces cerevisiae*) was suspended in PBS at a concentration of 10 mg/mL. Bacterial strains were grown overnight at 310K on LB

broth without agitation, washed, and re-suspended in PBS at an OD_{600} value of 1. Equal volumes of yeast cell suspension and two-fold serial dilutions of *E. coli* strain suspensions in PBS were mixed in a 96-well plate and incubated at RT for 2 h. Agglutination was followed and monitored visually.

The procedure was repeated in presence of the ligands to determine the minimal concentration of test compound (MIC, Minimal Inhibitory Concentration) able to inhibit yeast agglutination induced by wild-type *E. coli* CFT073 strain.
90 µL of bacterial suspension were transferred on a 96-well microplate and mixed with 10 µL of two-fold serial dilutions of the target compound in PBS (31-500 µM final concentration) for 10 min under orbital shaking. Equal volumes (100 µl) of yeast cell suspension were mixed to each well and the plates were incubated at RT for 2 h. Agglutination was monitored visually, and the MIC was obtained from the lowest concentration of each molecule that was still able to inhibit the yeast agglutination. All experiments were repeated three times.

7.9 Fluorescence Activated Cell Sorting (FACS)

The FACS experiments described in Chapter 6 were performed with the knock outs of the HEK293 suspension cells (KO *mgat1/c1galt1*, KO *mgat1/b4galt5/6 and* KO *c1galt1/b4galt5/6*). Those cells were tested for their ability to bind a selected variety of NHS-fluorescein labelled galectins through flow cytometry.

The **cell lines** were grown in suspension culture until a density of 0.5–2 10^6 cells/ml was reached. Trypan blue staining was used to assess cell viability and only the cell cultures with at least 90% of viability were considered for the experiment.
The preparation of the cells was different for the untreated cells and for those treated with neuraminidase. In both cases the cells were always kept in ice:

- Untreated cells: cells were collected and centrifuged for 3 minutes at 500 *g*, then washed with 2 mL of PBS pH 7.4. 200 mM of lactose were added (Sigma) to remove bound endogenous galectins or binders. This addition was followed by two additional washes with PBS. The final pellet was suspended in the required volume to set up a 96-well plate with 0.4 x 10^6 cells in each well.

- Cells treated with neuraminidase: cells were collected and centrifuged for 5 min at 500 *g*. The pellet was re-suspended in PBS with 1% BSA containing 150 mU/mL of neuraminidase. The preparation was incubated at 310K with rotation for 30 minutes. The cells were then centrifuged at 500 *g* for 3 min and the pellet was re-suspended in 5 mL of PBS + 1% BSA, with 200 mM of lactose (Sigma). They were then washed twice with PBS + 1% BSA to remove the lactose excess. The final pellet was suspended in the required volume to set up a 96-well plate with 0.4 x 10^6 cells in each well.

The **labelling of galectins** had previously been achieved in the CCG center using NHS-fluorescein (ThermoFisher Scientific), a fluorophore that is excited at 494 nm and emits at 518 nm. Briefly, the NHS esters react with primary amino groups (-NH_2) at pH 7-9, forming stable amide bonds. The NHS-fluorescein, dissolved in DMSO, was directly added at a 10-fold molar excess to a 2 mg/ml of galectin solution dissolved in the coupling buffer (20 mM HEPES). The mixture was kept in the dark and incubated at RT with continuous mixing for 1 h. The labelled galectins were separated from unreacted dye by buffer exchange with PBS on a PD10 column (GE Healthcare). The degree of labelling was calculated according to the manufacturer's instructions.

Round-shaped 96-well **plates** (ThermoFisher Scientific Nunclon Delta surface) were seeded with 200 µL of cell suspension (following the plate-tables reported in Figure 7.4) and then spun down (400 *g* for 5 minutes). During the entire preparation of the plates, they were always maintained on ice. The pellet on the bottom of the wells was re-suspended with the adequate solution depending on the organization of the plate.
Plate 1 and plate 2 were incubated 1 h with the NHS-fluorescein labelled galectins (re-suspended in PBS 1X pH 7.4, with or without 200 mM of lactose, depending on the map). The galectins were then employed at final concentrations of 0, 1, 2, 4, and 8 µM. After the incubation with galectins, the cells were washed twice and then 50 µg/mL of propidium iodide (PI, Life Technologies, Inc.) were added to identify the permeable cells. The use of PI allows testing the cell viability through dye exclusion, since living cells have membranes that are intact and exclude a variety of dyes that easily penetrate the damaged, permeable membranes of non-viable cells. In particular, PI binds to

double stranded DNA by intercalating between base pairs; it is excited at 488 nm and emits at a maximum of 617 nm. It can be used in combination with other fluorochromes excited at 488 nm.

After PI addition, plates 1 and 2 were ready for the flow cytometry experiments.

For plate 3, the cells were incubated at 277 K for 30 min under shaking with 1:1000 biotinylated-control lectins:

- Biotinylated *Peanut Agglutinin* (PNA) with stock concentration of 2 mg/mL (Vector Laboratories).
- Biotinylated *Vicia Villosa* (VVA) with stock concentration of 2 mg/mL (Vector Laboratories).
- Biotinylated *Phaseolus Vulgaris Leucoagglutinin* (PHA-L) with stock concentration of 2 mg/mL (Vector Laboratories).

Plate 3 was subsequently washed twice with PBS and incubated with 1:1000 streptavidin (Invitrogen, stock 2 mg/mL) at 277 K for 30 minutes with shaking. The plate was washed twice and the pellet was finally re-suspended in 100 µL of PBS + 1% BSA right before flow analysis.

The **fluorescence cytometry** experiments were performed using an LSRII cytometer (BD Biosciences) and the data were analysed using the FlowJo software. Fluorescence data were collected using logarithmic amplification on 30,000 light scatter-gated events (cell counts).

PLATE 1 WT-MG/C1	1	2	3	4	5	6
A	WT Gal 0	WT Gal 0.5	WT Gal 1	WT Gal 2	WT Gal 4	WT - NS (NO PI)
B	WT Gal 0	WT Gal 0.5	WT Gal 1	WT Gal 2	WT Gal 4	WT - NS (NO PI)
C	WT Gal 0 +lac	WT Gal 0.5 +lac	WT Gal 1 +lac	WT Gal 2 +lac	WT Gal 4 +lac	WT Gal 2 (NO PI)
D	WT Gal 0 +lac	WT Gal 0.5 +lac	WT Gal 1 +lac	WT Gal 2 +lac	WT Gal 4 +lac	WT Gal 2 (NO PI)
E	MG/C1 Gal 0	MG/C1 Gal 0.5	MG/C1 Gal 1	MG/C1 Gal 2	MG/C1 Gal 4	MG/C1-NS(NO PI)
F	MG/C1 Gal 0	MG/C1 Gal 0.5	MG/C1 Gal 1	MG/C1 Gal 2	MG/C1 Gal 4	MG/C1-NS(NO PI)
G	MG/C1 Gal 0+lac	MG/C1 Gal 0.5+lac	MG/C1 Gal 1+lac	MG/C1 Gal 2+lac	MG/C1 Gal 4+lac	MG/C1 Gal 2 (NO PI)
H	MG/C1 Gal 0+lac	MG/C1 Gal 0.5+lac	MG/C1 Gal 1+lac	MG/C1 Gal 2+lac	MG/C1 Gal 4+lac	MG/C1 Gal 2 (NO PI)

PLATE 2 C1/B4- MG/B4	1	2	3	4	5	6
A	C1/B4 Gal 0	C1/B4 Gal 0.5	C1/B4 Gal 1	C1/B4 Gal 2	C1/B4 Gal 4	C1/B4 - NS (NO PI)
B	C1/B4 Gal 0	C1/B4 Gal 0.5	C1/B4 Gal 1	C1/B4 Gal 2	C1/B4 Gal 4	C1/B4 - NS (NO PI)
C	C1/B4 Gal 0 +lac	C1/B4 Gal 0.5 +lac	C1/B4 Gal 1 +lac	C1/B4 Gal 2 +lac	C1/B4 Gal 4 +lac	C1/B4 Gal 2 (NO PI)
D	C1/B4 Gal 0 +lac	C1/B4 Gal 0.5 +lac	C1/B4 Gal 1 +lac	C1/B4 Gal 2 +lac	C1/B4 Gal 4 +lac	C1/B4 Gal 2 (NO PI)
E	MG/B4 Gal 0	MG/B4 Gal 0.5	MG/B4 Gal 1	MG/B4 Gal 2	MG/B4 Gal 4	MG/B4-NS(NO PI)
F	MG/B4 Gal 0	MG/B4 Gal 0.5	MG/B4 Gal 1	MG/B4 Gal 2	MG/B4 Gal 4	MG/B4-NS(NO PI)
G	MG/B4 Gal 0+lac	MG/B4 Gal 0.5+lac	MG/B4 Gal 1+lac	MG/B4 Gal 2+lac	MG/B4 Gal 4+lac	MG/B4 Gal 2 (NO PI)
H	MG/B4 Gal 0+lac	MG/B4 Gal 0.5+lac	MG/B4 Gal 1+lac	MG/B4 Gal 2+lac	MG/B4 Gal 4+lac	MG/B4 Gal 2 (NO PI)

PLATE 3 Lectins Ctrl	1	2	3	4	5	6
A	WT PHA	WT PNA	WT VVA	WT Strep	WT Neu PNA	WT Neu VVA
B	WT PHA	WT PNA	WT VVA	WT Strep	WT Neu PNA	WT Neu VVA
C	MG/C1 PHA	MG/C1 PNA	MG/C1 VVA	MG/C1 Strep	MG/C1 Neu PNA	MG/C1 Neu VVA
D	MG/C1 PHA	MG/C1 PNA	MG/C1 VVA	MG/C1 Strep	MG/C1 Neu PNA	MG/C1 Neu VVA
E	MG/B4 PHA	MG/B4 PNA	MG/B4 VVA	MG/B4 Strep	MG/B4 Neu PNA	MG/B4 Neu VVA
F	MG/B4 PHA	MG/B4 PNA	MG/B4 VVA	MG/B4 Strep	MG/B4 Neu PNA	MG/B4 Neu VVA
G	C1/B4 PHA	C1/B4 PNA	C1/B4 VVA	C1/B4 Strep	C1/B4 Neu PNA	C1/B4 Neu VVA
H	C1/B4 PHA	C1/B4 PNA	C1/B4 VVA	C1/B4 Strep	C1/B4 Neu PNA	C1/B4 Neu VVA

Figure 7.4. Organization of the 96-well plates used to perform the cytofluorimetry experiments.

7.10 References

1. Jeong, H.; Barbe, V.; Lee, C.H.; Vallenet, D.; Yu, D.S.; Choi, S.H.; Couloux, A.; Lee, S.W.; Yoon, S.H.; Cattolico, L.; et al. Genome Sequences of Escherichia Coli B Strains REL606 and BL21(DE3). *J. Mol. Biol.* **2009**, *394*, 644–652, doi:10.1016/J.JMB.2009.09.052.
2. Aricescu, A.R.; Lu, W.; Jones, E.Y. Biological Crystallography A Time-and Cost-Efficient System for High-Level Protein Production in Mammalian Cells. *Res. Pap. Acta Cryst* **2006**, *62*, 1243–1250, doi:10.1107/S0907444906029799.
3. Tavares, M.R.; Bláhová, M.; Sedláková, L.; Elling, L.; Pelantová, H.; Konefał, R.; Etrych, T.; Křen, V.; Bojarová, P.; Chytil, P. High-Affinity N-(2-Hydroxypropyl)Methacrylamide Copolymers with Tailored N-Acetyllactosamine Presentation Discriminate between Galectins. *Biomacromolecules* **2020**, *21*, 641–652, doi:10.1021/ACS.BIOMAC.9B01370/SUPPL_FILE/BM9B01370_SI_001.PDF.
4. Echeverria, B.; Serna, S.; Achilli, S.; Vivès, C.; Pham, J.; Thépaut, M.; Hokke, C.H.; Fieschi, F.; Reichardt, N.C. Chemoenzymatic Synthesis of N-Glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-Type Lectin Receptors. *ACS Chem. Biol.* **2018**, *13*, 2269–2279, doi:10.1021/ACSCHEMBIO.8B00431/ASSET/IMAGES/LARGE/CB-2018-00431Q_0007.JPEG.
5. Serna, S.; Etxebarria, J.; Ruiz, N.; Martin-Lomas, M.; Reichardt, N.C. Construction of N-Glycan Microarrays by Using Modular Synthesis and On-Chip Nanoscale Enzymatic Glycosylation. *Chem. – A Eur. J.* **2010**, *16*, 13163–13175, doi:10.1002/CHEM.201001295.
6. Rillahan, C.D.; Schwartz, E.; McBride, R.; Fokin, V. V.; Paulson, J.C. Click and Pick: Identification of Sialoside Analogues for Siglec-Based Cell Targeting. *Angew. Chemie Int. Ed.* **2012**, *51*, 11014–11018, doi:10.1002/ANIE.201205831.
7. Blattner, F.R.; Plunkett, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; et al. The Complete Genome Sequence of Escherichia Coli K-12. *Science* **1997**, *277*, 1453–1462, doi:10.1126/SCIENCE.277.5331.1453.
8. Welch, R.A.; Burland, V.; Plunkett, G.; Redford, P.; Roesch, P.; Rasko, D.; Buckles, E.L.; Liou, S.R.; Boutin, A.; Hackett, J.; et al. Extensive Mosaic Structure Revealed by the Complete Genome Sequence of Uropathogenic Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 17020–17024, doi:10.1073/PNAS.252529799.
9. Datsenko, K.A.; Wanner, B.L. One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 6640–6645, doi:10.1073/PNAS.120163297/ASSET/103C1B8D-D302-4337-8E67-9F9084156407/ASSETS/GRAPHIC/PQ1201632006.JPEG.
10. Chandrasegaran, S.; Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J. Mol. Biol.* **2016**, *428*, 963–989, doi:10.1016/J.JMB.2015.10.014.
11. Narimatsu, Y.; Joshi, H.J.; Yang, Z.; Gomes, C.; Chen, Y.H.; Lorenzetti, F.C.; Furukawa, S.; Schjoldager, K.T.; Hansen, L.; Clausen, H.; et al. A Validated GRNA Library for CRISPR/Cas9 Targeting of the Human Glycosyltransferase Genome. *Glycobiology* **2018**, *28*, 295–305, doi:10.1093/GLYCOB/CWX101.
12. Narimatsu, Y.; Joshi, H.J.; Nason, R.; Van Coillie, J.; Karlsson, R.; Sun, L.; Ye, Z.; Chen, Y.H.; Schjoldager, K.T.; Steentoft, C.; et al. An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Mol. Cell* **2019**, *75*, 394-407.e5, doi:10.1016/J.MOLCEL.2019.05.017.
13. Dabelsteen, S.; Pallesen, E.M.H.; Marinova, I.N.; Nielsen, M.I.; Adamopoulou, M.; Rømer, T.B.; Levann, A.; Andersen, M.M.; Ye, Z.; Thein, D.; et al. Essential Functions of Glycans in Human Epithelia Dissected by a CRISPR-Cas9-Engineered Human

Organotypic Skin Model. *Dev. Cell* **2020**, *54*, 669-684.e7, doi:10.1016/J.DEVCEL.2020.06.039.

14. Marinova, I.N.; Wandall, H.H.; Dabelsteen, S. Protocol for CRISPR-Cas9 Modification of Glycosylation in 3D Organotypic Skin Models. *STAR Protoc.* **2021**, *2*, 100668–100668, doi:10.1016/J.XPRO.2021.100668.
15. Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, doi:10.1021/ja0100120.
16. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788, doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q.
17. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-.).* **1996**, *274*, 1531–1534, doi:10.1126/SCIENCE.274.5292.1531.
18. Vranken, W.F.; Boucher, W.; Stevens, T.J.; Fogh, R.H.; Pajon, A.; Llinas, M.; Ulrich, E.L.; Markley, J.L.; Ionides, J.; Laue, E.D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **2005**, *59*, 687–696, doi:10.1002/PROT.20449.
19. Williamson, M.P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16, doi:10.1016/J.PNMRS.2013.02.001.
20. Hwang, T.L.; Van Zijl, P.C.M.; Mori, S. Accurate Quantitation of Water-Amide Proton Exchange Rates Using the Phase-Modulated CLEAN Chemical EXchange (CLEANEX-PM) Approach with a Fast-HSQC (FHSQC) Detection Scheme. *J. Biomol. NMR* **1998**, *11*, 221–226, doi:10.1023/A:1008276004875.
21. Pelton, J.G.; Torchia, D.A.; Meadow, N.D.; Roseman, S. Tautomeric States of the Active-Site Histidines of Phosphorylated and Unphosphorylated IIIGlc, a Signal-Transducing Protein from Escherichia Coli, Using Two-Dimensional Heteronuclear NMR Techniques. *Protein Sci.* **1993**, *2*, 543–558, doi:10.1002/PRO.5560020406.
22. Van Dijk, A.A.; Scheek, R.M.; Dijkstra, K.; Wolters, G.K.; Robillard, G.T. Characterization of the Protonation and Hydrogen Bonding State of the Histidine Residues in IIAmtl, a Domain of the Phosphoenolpyruvate-Dependent Mannitol-Specific Transport Protein. *Biochemistry* **1992**, *31*, 9063–9072, doi:10.1021/BI00152A050.
23. Neudecker, P.; Lundström, P.; Kay, L.E. Relaxation Dispersion NMR Spectroscopy as a Tool for Detailed Studies of Protein Folding. *Biophys. J.* **2009**, *96*, 2045–2054, doi:10.1016/J.BPJ.2008.12.3907.
24. Latham, M.P.; Zimmermann, G.R.; Pardi, A. NMR Chemical Exchange as a Probe for Ligand-Binding Kinetics in a Theophylline-Binding RNA Aptamer. *J. Am. Chem. Soc.* **2009**, *131*, 5052–5053, doi:10.1021/JA900695M/SUPPL_FILE/JA900695M_SI_001.PDF.
25. Perrin, C.L.; Dwyer, T.J. Application of Two-Dimensional NMR to Kinetics of Chemical Exchange. *Chem. Rev.* **1990**, *90*, 935–967, doi:10.1021/CR00104A002/ASSET/CR00104A002.FP.PNG_V03.
26. Dreux, N.; Denizot, J.; Martinez-Medina, M.; Mellmann, A.; Billig, M.; Kisiela, D.; Chattopadhyay, S.; Sokurenko, E.; Neut, C.; Gower-Rousseau, C.; et al. Point Mutations in FimH Adhesin of Crohn's Disease-Associated Adherent-Invasive Escherichia Coli Enhance Intestinal Inflammatory Response. *PLoS Pathog.* **2013**, *9*, doi:10.1371/JOURNAL.PPAT.1003141.

CHAPTER 8

CONCLUSIONS

8.1 General Conclusions

The detailed conclusions of the works discussed in this Thesis are reported at the end of Chapters 3, 4, 5, and 6.

1. From a global perspective, the structural and dynamics features of the interaction between a variety of lectin-sugar systems have been herein deciphered at atomic detail employing a plethora of complementary experimental techniques assisted by computational methods.

2. The recognition of the histo blood group saccharide antigens by the dimeric galectin-1 has been decoded using a combination of ITC, MD and NMR methodologies. An affinity decrease was detected when passing from the LacNAc disaccharide to the more complex tri- and tetrasaccharide antigens. The ITC analysis highlighted that the binding events occur with negative cooperativity and that are always favoured by entropy. NMR and MD demonstrated the presence of concerted motions of the protein related with allostery. Such phenomenon was noticed only upon LacNAc binding, while no concerted clustered residues were detected in other cases (galectin-1 with lower affinity binders or galectin-3 with LacNAc). These results provide insights into the functional role of the dimeric architecture of the lectin.

3. The interaction of two human galectins of biomedical interest (galectin-1 and the CRD of galectin-3) with a panel of glycopolymers has been deeply analysed. The detailed picture of the recognition event has been provided at different degrees of resolution using NMR, DLS, and cryo-EM. Those particular glycopolymers bring a clear discrimination between the galectins, being galectin-1 the preferred receptor. Cryo-EM results visually testify that different networks are formed by the monomeric and dimeric galectins with the glycoclusters and confirm the presence of cross-linked intermolecular entities mediated by the prototype galectin. This selectivity based on the structure of a multivalent molecule paves the way to the development of compounds of biomedical interest.

4. The fine details of the recognition processes taking place in solution between the C-type lectin LSECtin and a variety of N-glycans has been unravelled by NMR. The results obtained in solution strongly differ from those obtained when the N-glycans are immobilized on a solid surface. In fact, under array conditions, substitutions in the 3-arm of biantennary N-glycans completely abrogate the binding to LSECtin, whereas the same substitution at the 6-arm is tolerated. On the contrary, in solution both substitutions are equally accepted without precluding the binding. This dramatic difference is due to the different presentation of the epitopes under immobilized settings, highlighting how different techniques can influence the final outcome. This compelling example points out the tremendous difficulty of translating *in vitro* results to the *in vivo* environment.

5. A compendium of promising results obtained adopting *in-cell* and *on-cell* strategies have been presented, emphasising how the standard reductionistic approach employed in chemistry labs has to be complemented with experiments towards the natural-like context.

8.2 Scientific publications during this dissertation

- Lete, M. G., Franconetti, A., **Bertuzzi. S.,** Delgado, S., Azkargorta, M., Elortza, F., Millet, M., Jiménez-Osés, G., Arda, A. and Jiménez-Barbero, J. NMR Investigations of Protein-Carbohydrate Interactions: The Recognition of Glycans by Galectins Engineered with Fluorotryptophan Residues. *Chem Eur J.***,** *just accepted***.**

- **Bertuzzi, S.,** Peccati, F., Serna, S., Artschwager, R., Notova, S., Thépaut, M., Jiménez-Osés, G., Fieschi, F., Reichardt, N.C., Jiménez-Barbero, J. and Ardà, A. Immobilization of Biantennary N-Glycans Leads to Branch Specific Epitope Recognition by LSECtin. *ACS Cent. Sci.* **2022**, *8(10)*, 1415–1423. 10.1021/acscentsci.2c00719

- Pietri, G.P., Tontini, M., Brogioni, B., Oldrini, D., Robakiewicz, S., Henriques, P., Calloni, I., Abramova, V., Santini, L., Malić, S., Miklić, K., Lisnic, B., **Bertuzzi, S.,** Unione, L., Balducci, E., de Ruyck, J., Romani, M.R., Jiménez-Barbero, J., Bouckaert, J., Jonjic, S., Rovis, T. and Adamo, R.. Elucidating the Structural and Minimal Protective Epitope of the Serogroup X Meningococcal Capsular Polysaccharide. *Frontiers in Molecular Biosciences.* **2021**, *8*, 895. https://doi.org/10.3389/fmolb.2021.745360

- Lassfolk, R., **Bertuzzi, S.,** Ardá, A., Wärnå, J., Jiménez-Barbero, J., Leino, R. Kinetic Studies of Acetyl Group Migration between the Saccharide Units in an Oligomannoside Trisaccharide Model Compound and a Native Galactoglucomannan Polysaccharide. *ChemBioChem*. **2021**, *22*, 2986–2995. https://doi.org/10.1002/cbic.202100374

- **Bertuzzi S.,** Gimeno, A., Martinez-Castillo, A., G. Lete, M., Delgado, S., Airoldi, C., Rodrigues Tavares, M., Bláhová, M., Chytil, P., Křen, V., G. A. Abrescia, N., Ardá, A., Bojarová, P., Jiménez-Barbero, J. Cross-linking effects dictate the preference of galectins to bind LacNAc-decorated HPMA copolymers. *International Journal of Molecular Sciences.* **2021**, *22 (11)*, 6000. https://doi.org/10.3390/ijms22116000

- Palmioli, A., Sperandeo, P., **Bertuzzi, S**., Polissi, A., Airoldi, C. On-cell saturation transfer difference NMR for the identification of FimH ligands and inhibitors. *Bioorganic Chemistry*. **2021**, *112 (7)*, 104876. https://doi.org/10.1016/j.bioorg.2021.104876

- **Bertuzzi, S.,** Gimeno, A., Núñez-Franco, R., Bernardo-Seisdedos, G., Delgado, S., Jiménez-Osés, G., Millet, O., Jiménez-Barbero, J., Ardá, A. Unravelling the Time Scale of Conformational Plasticity and Allostery in Glycan Recognition by Human Galectin-1. *Chemistry - A European Journal.* **2020**, *26*, 15643-15653. https://doi.org/10.1002/chem.202003212

- **Bertuzzi, S.,** Quintana, J. I., Ardá, A., Gimeno, A., & Jiménez-Barbero, J. Targeting Galectins With Glycomimetics. *Frontiers in Chemistry*. **2020**, *8*, 593. https://doi.org/10.3389/fchem.2020.00593

- Gimeno, A., Delgado, S., Valverde, P., **Bertuzzi, S**., Berbis, M., Echavarren, J., Lacetera, A., Martín-Santamaria, S., Surolia A., Cañada, F., Jimenez-Barbero, J., Ardá, A. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood Group Antigens by Human Galectin-3. *Angewandte Chemie International Edition*. **2019**, *58*, 7268-7272. https://doi.org/10.1002/anie.201900723

Book chapter

- Arda, A., Atxabal, U., **Bertuzzi, S.,** Canales, A., Cañada, F.J., Gimeno, A., Ereño-Orbea, J., Fernández de Toro, B., Franconetti, A., Gómez-Redondo, M., Lete, M., Lenza, M.P., Martínez, J.D., Moure, M.J., Poveda, A., Quintana, J.I., Valverde P. and Jiménez-Barbero, J. Recent advances in the application of NMR methodologies to analyze the conformation, dynamics, and interactions of saccharides. *Carbohydrate Chemistry: Chemical and Biological Approaches*. **2020**, *44*, 170-194. https://doi.org/10.1039/9781788013864-00170

8.3 Contributions to congress during this dissertation

Poster Presentations:

- "Immobilization of biantennary N-glycans can lead to branch specific epitope recognition by LSECtin" Summer School in Biomedical Glycoscience (Jaca, Spain, **2022**). [Author] (Presented by Sara Bertuzzi).

- "Cross-linking effects dictate the preference of galectins to bind LacNAc-derived HMPA copolymer" 9th Ibero-American NMR meeting and 7th Iberian NMR meeting GERMN, (online, **2021**). [Author] (Presented by Sara Bertuzzi).

- "Structural insights tn the Recognition of glycans by Human Galectin-1" Virtual EUROMAR International Congress, (online, **2020**). [Author] (Presented by Sara Bertuzzi).

- "Insight into the interaction between human galectin-1 and the blood group antigens. An NMR view" 3rd Glycobasque Meeting, (Bilbao, Spain, **2020**). [Author] (Presented by Sara Bertuzzi).

- "NMR insights on the recognition mechanism of histo-blood group antigens by human galectin-1" V GEQB Chem Bio congress (Granada, Spain, **2020**). [Author] (Presented by Sara Bertuzzi).

- "NMR in molecular recognition: the case of human galectin-1 and the blood group antigens" EUROMAR International congress, (Berlin, Germany, **2019**). [Author] (Presented by Sara Bertuzzi).

- "NMR in molecular recognition: the case of human galectin-1 and the blood group antigens"; XXXVII Bienal of RSEQ (Real Sociedad Española de Quimica), (San Sebastián, Spain, **2019**). [Author] (Presented by Sara Bertuzzi).

- "Molecular recognition of blood group antigens by hGalectin-1" X Spanish Drug Discovery Network Meeting, (Bilbao, Spain, **2018**). [Author] (Presented by Sara Bertuzzi).

- "Molecular recognition of blood group antigens by hGalectin-1" 29th International Carbohydrate Symposium (ICS), (Lisbon, Portugal, **2018**). [Author] (Presented by Sara Bertuzzi).

Oral Presentations:

- "Changes in glycan presentation dictate lectin binding: in solution vs on array" 5th Glycobasque Meeting, (San Sebastián, Spain, **2022**). [Author] (Presented by Sara Bertuzzi).

- "Immobilization of biantennary N-glycans leads to branch specific epitope recognition by LSECtin" 11th GERMN Biennial NMR Meeting (Almeria, Spain, **2022**). [Author] (Presented by Sara Bertuzzi).

- "Immobilization of glycans can lead to branch specific epitope recognition" Challenges in Chemical Synthesis 2022 (Leioa, Spain, **2022**). [Author] (Presented by Sara Bertuzzi).

- "Cross linking effects dictate the preference of galectins to bind LacNAc Derived HMPA copolymers" 4th Glycobasque Meeting, (San Sebastián, Spain, **2021**). [Author] (Presented by Sara Bertuzzi).

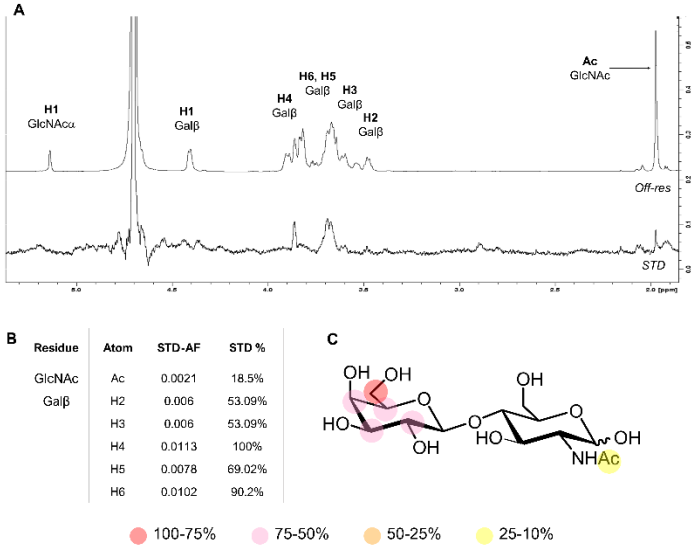
- "NMR view of the interaction of histo blood group antigens with human galectin 1" 2nd Glycobasque Meeting, (Bilbao, Spain, **2019**). [Author] (Presented by Sara Bertuzzi).

- "Human Galectin-1: molecular recognition of blood group antigens" Introductory Workshop on Biomedical Glycoscience, (San Sebastián, Spain, **2019**). [Author] (Presented by Sara Bertuzzi).

CHAPTER 9

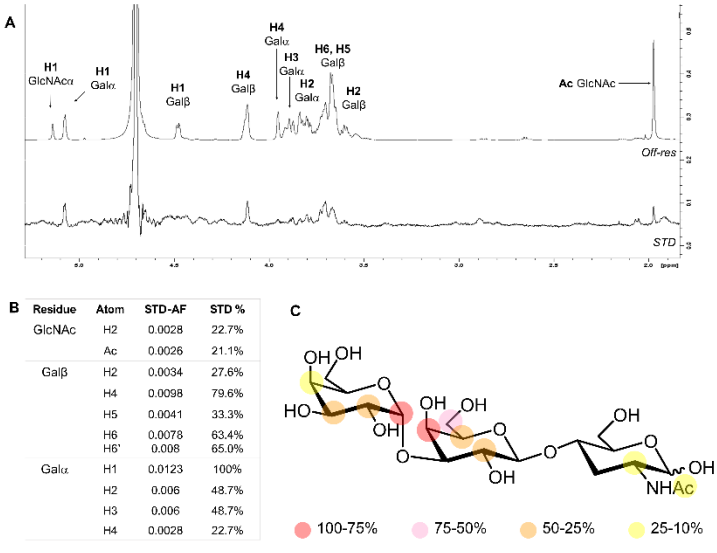
SUPPORTING INFORMATION

9.1 Supporting Information of Chapter 3



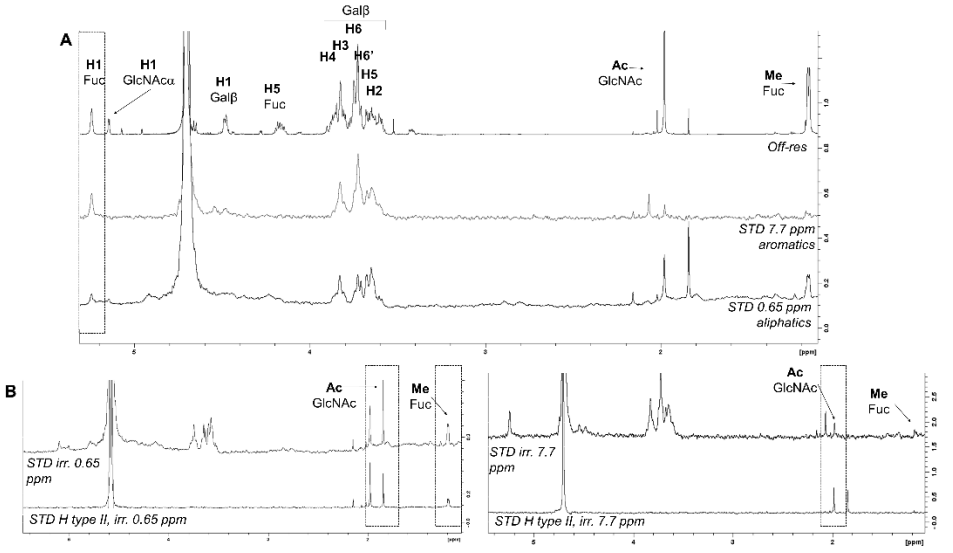
Residue	Atom	STD-AF	STD %
GlcNAc	Ac	0.0021	18.5%
Galβ	H2	0.006	53.09%
	H3	0.006	53.09%
	H4	0.0113	100%
	H5	0.0078	69.02%
	H6	0.0102	90.2%

Figure 9.1. STD of galectin-1 (50 μM) and **LacNAc** (2.5 mM). A) Off-resonance and STD spectra (irradiation at 0.83 ppm). B) Relative STD-AF intensities and STD percentage for the most intense and non-overlapped proton signals. Data calculated on the basis of the STD experiment. C) Representation of the epitope mapping obtained for aliphatic irradiation with the corresponding color legend.



Residue	Atom	STD-AF	STD %
GlcNAc	H2	0.0028	22.7%
	Ac	0.0026	21.1%
Galβ	H2	0.0034	27.6%
	H4	0.0098	79.6%
	H5	0.0041	33.3%
	H6	0.0078	63.4%
	H6'	0.008	65.0%
Galα	H1	0.0123	100%
	H2	0.006	48.7%
	H3	0.006	48.7%
	H4	0.0028	22.7%

Figure 9.2. STD of galectin-1 (50 μM) and **Galili** (2.5 mM). A) Off-resonance and STD spectra (irradiation at 0.83 ppm). B) Relative STD-AF intensities and STD percentage for the most intense and non-overlapped proton signals. Data calculated on the basis of the STD experiment. C) Representation of the epitope mapping obtained for aliphatic irradiation with the corresponding color legend.



C

Residue	Atom	STD-AF Aliphatics 0.65 ppm	STD % Aliphatics	STD-AF Aromatics 7.7 ppm	STD % Aromatics
GlcNAc	Ac	0.0027	32%	0.0005	7%
Galβ	H1	-	-	0.003	44%
	H2	0.003	35%	0.0037	49%
	H4	0.0036	42%	0.0044	49%
	H5	0.0085	100%	0.0062	90%
	H6	0.0023	27%	0.0046	67%
	H6'	0.0077	90%	0.006	87%
Fucα	H1	**0.0033**	**39%**	**0.0069**	**100%**
	Me	0.0023	27%	0.0003	4%

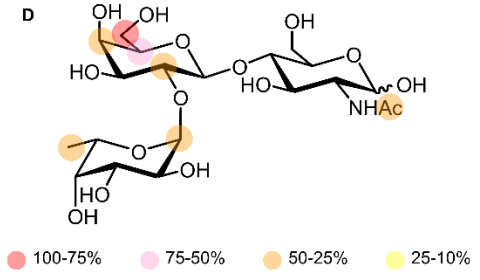
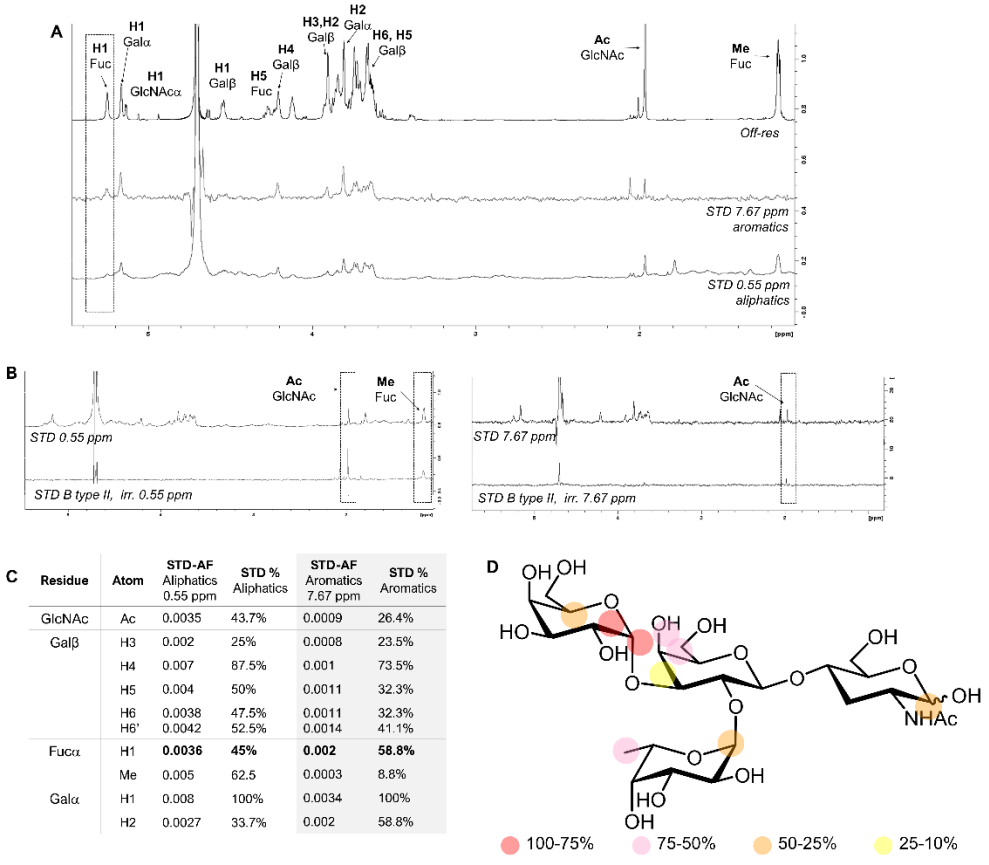


Figure 9.3. STD of galectin-1 (50 μM) and **H type-II** antigen (6.9 mM). A) Off-resonance (irradiation at 100 ppm) and STD spectra (irradiation at 7.76 ppm or 0.65 ppm). B) STD spectrum with irradiation at 0.65 ppm (on the left) and at 7.67 ppm (on the right) compared with the STD of the ligand alone in the same experimental conditions (**H type-II** at 6.9 mM). C) Relative STD-AF intensities and STD percentage for the most intense and non-overlapped proton signals. Data calculated based on the STD experiments. D) Representation of the epitope mapping obtained for aliphatic irradiation with the corresponding color legend.



C

Residue	Atom	STD-AF Aliphatics 0.55 ppm	STD % Aliphatics	STD-AF Aromatics 7.67 ppm	STD % Aromatics
GlcNAc	Ac	0.0035	43.7%	0.0009	26.4%
Galβ	H3	0.002	25%	0.0008	23.5%
	H4	0.007	87.5%	0.001	73.5%
	H5	0.004	50%	0.0011	32.3%
	H6	0.0038	47.5%	0.0011	32.3%
	H6'	0.0042	52.5%	0.0014	41.1%
Fucα	H1	**0.0036**	**45%**	**0.002**	**58.8%**
	Me	0.005	62.5	0.0003	8.8%
Galα	H1	0.008	100%	0.0034	100%
	H2	0.0027	33.7%	0.002	58.8%

Figure 9.4. STD of galectin-1 (50 μM) and **B type-II** antigen (5 mM). A) Off-resonance (irradiation at 100 ppm) and STD spectra (irradiation at 7.67 ppm and at 0.55 ppm). B) STD spectrum of with irradiation at 0.55 ppm (on the left) and at 7.67 ppm (on the right) compared with the STD of the ligand alone in the same experimental conditions (**B type-II** at 5 mM). C) Relative STD-AF intensities and STD percentage for the most intense and non-overlapped proton signals. Data calculated based on the STD experiments. D) Representation of the epitope mapping obtained for the aliphatic irradiation with the corresponding color legend.

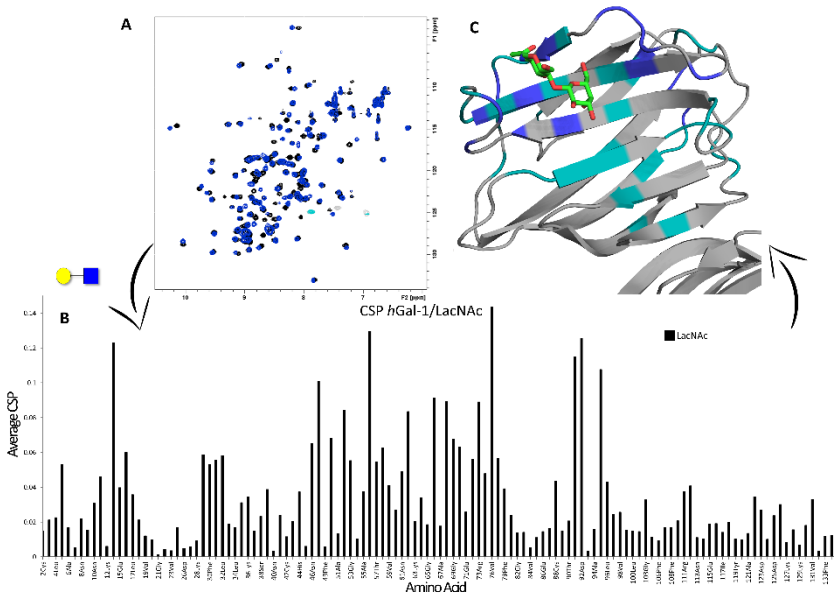


Figure 9.5. A) Superposition of the ^{1}H-^{15}N HSQC spectrum of galectin-1 in the *apo* form (black) and in presence of 10 equivalents of **LacNAc** (blue). B) CSP of backbone amides of galectin-1 upon addition of **LacNAc** (10 equivalents). C) Crystal structure of galectin-1 complexed with **LacNAc** (PDB ID: 1W6P). The most perturbed amino acids are highlighted on the protein structure using dark ($\Delta\delta > 2\sigma$) and light ($\Delta\delta = 1\text{-}2\sigma$) colors.

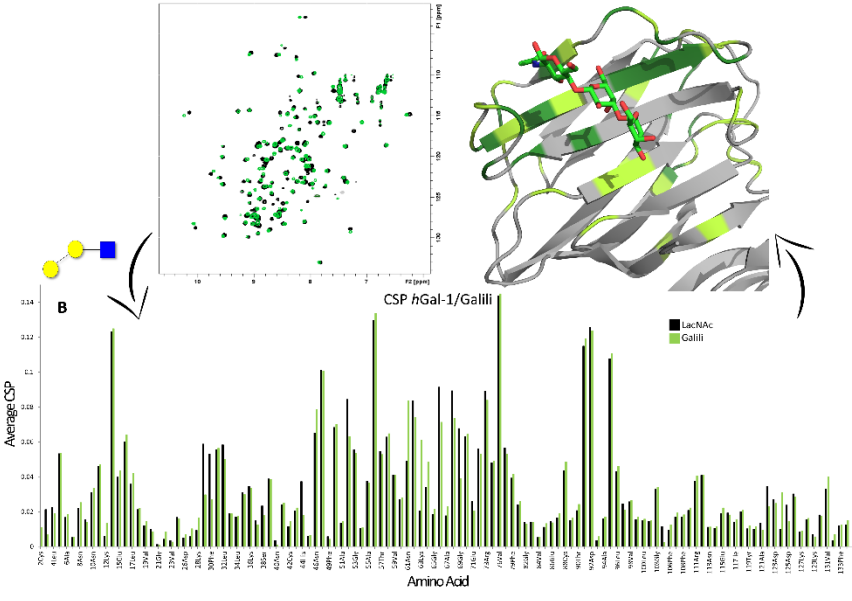


Figure 9.6. A) Superposition of the ^{1}H-^{15}N HSQC spectrum of galectin-1 in the *apo* form (black) and in presence of 10 equivalents of **Galili** (green). B) CSP (in green) of backbone amides of galectin-1 upon addition of **Galili** (10 equivalents). CSP values for LacNAc are reported in black as comparison. C) Crystal structure of galectin-1 (PDB ID: 1GZW) with **Galili** modeled in the binding site. The most perturbed amino acids were highlighted on the protein structure using dark ($\Delta\delta > 2\sigma$) and light ($\Delta\delta = 1\text{-}2\sigma$) colors.

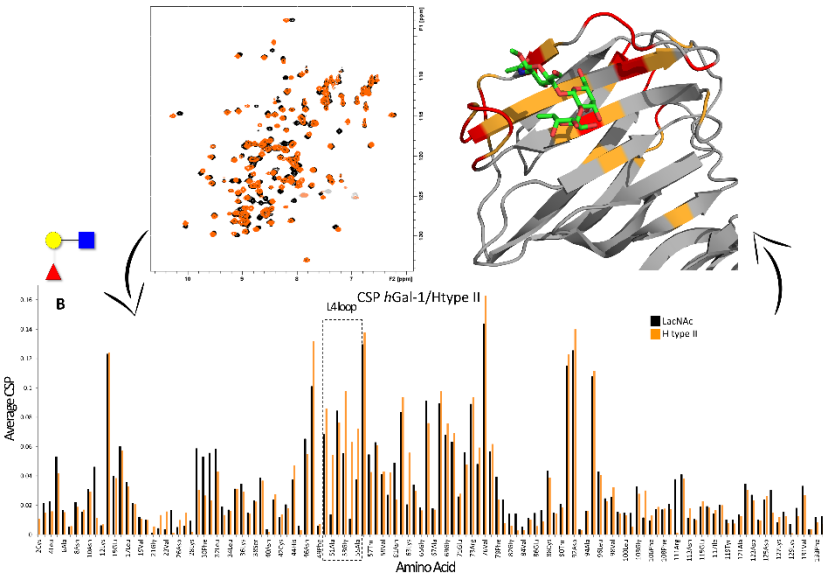


Figure 9.7. A) Superposition of the ^{1}H-^{15}N HSQC spectrum of galectin-1 in the *apo* form (black) and in presence of 10 equivalents of **H type-II** (orange). B) CSP (in orange) of backbone amides of galectin-1 upon addition of **H type-II** (10 equivalents). CSP values for LacNAc are reported in black as comparison. **C**) Crystal structure of galectin-1 (PDB ID: 1GZW) with **H type-II** modeled in the binding site. The most perturbed amino acids were highlighted on the protein structure using dark ($\Delta\delta > 2\sigma$) and light ($\Delta\delta = 1\text{-}2\sigma$) colors.

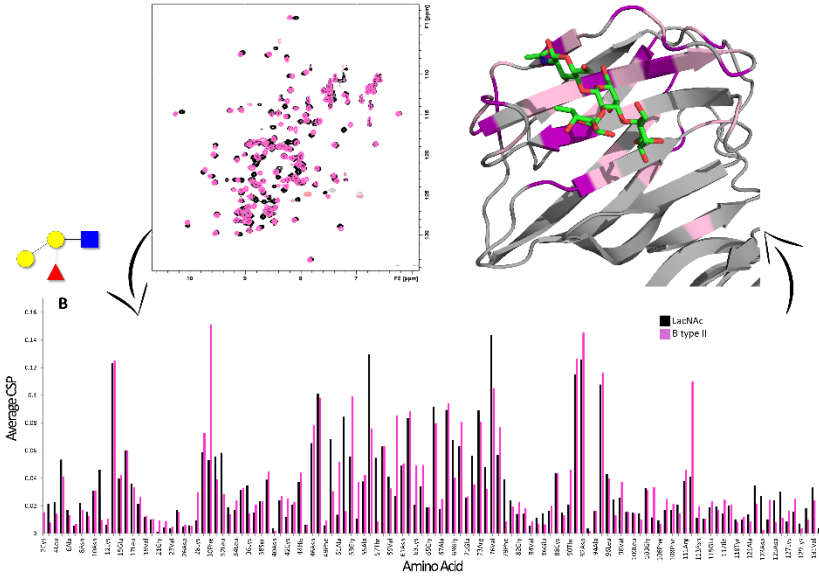
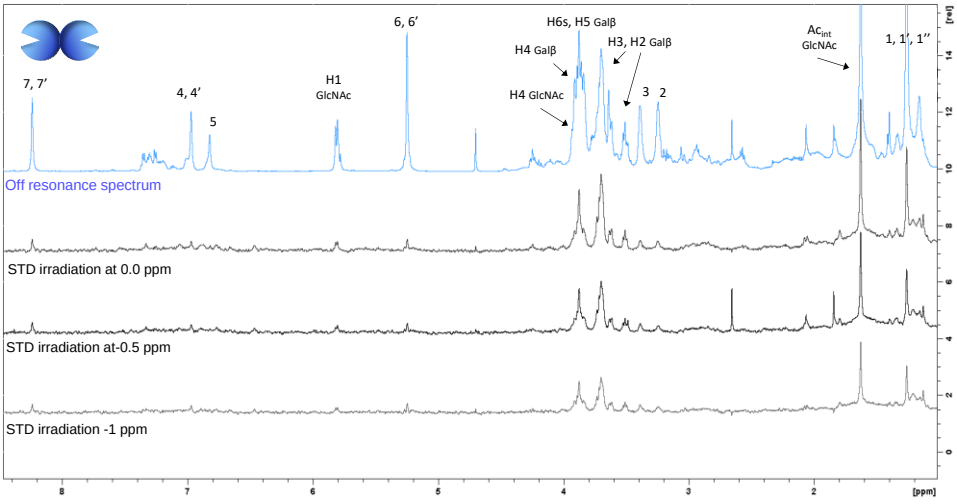


Figure 9.8. A) Superposition of the ^{1}H-^{15}N HSQC spectrum of galectin-1 in the *apo* form (black) and in presence of 10 equivalents of **B type-II** (magenta). B) CSP (in magenta) of backbone amides of galectin-1 upon addition of **B type-II** (15 equivalents). CSP values for LacNAc are reported in black as comparison. C) Crystal structure of galectin-1 (PDB ID: 1GZW) with **B type II** modeled in the binding site. The most perturbed amino acids were highlighted on the protein structure using dark ($\Delta\delta > 2\sigma$) and light ($\Delta\delta = 1\text{-}2\sigma$) colors.

9.2 Supporting Information of Chapter 4

A)



B)

0.0 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.014	30%
	H2	0.0232	49%
	H3	-	-
	H4	0.016	34%
	H5	-	-
	H6, H6'	-	-
	Ac	0.034	78%
Galβ	H1	-	-
	H2	0.03	63%
	H3	0.03	63%
	H4	0.038	80%
	H5	0.047	100%
	H6s	0.047	100%
scaffold	1, 1', 1''	0.02	42%
	2	0.01	21%
	3	0.012	25%
	4, 4'	0.012	25%
	5	0.011	23%
	6, 6'	0.006	12%
	7, 7'	0.012	25%

-0.5 ppm

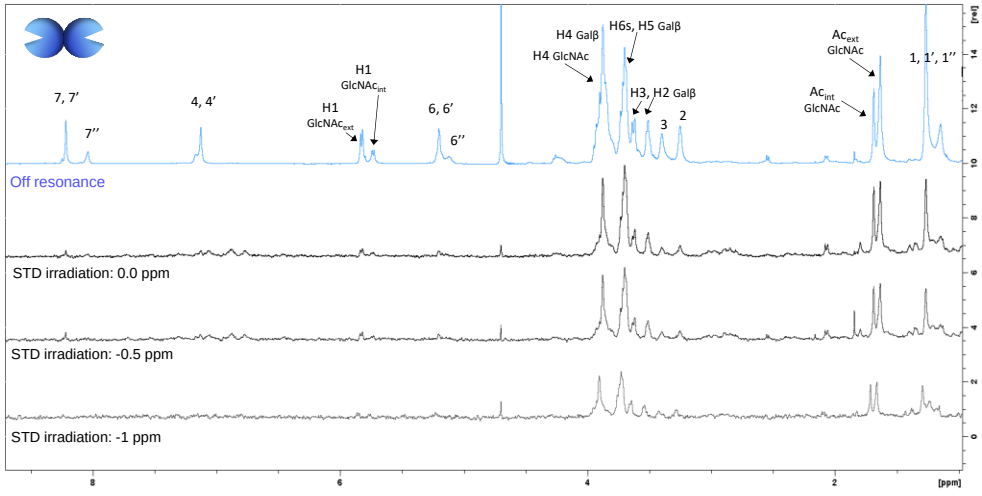
Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.0108	35%
	H2	0.0156	50%
	H3	-	-
	H4	0.0119	38%
	H5	-	-
	H6, H6'	-	-
	Ac	0.023	74%
Galβ	H1	-	-
	H2	0.022	71%
	H3	0.022	71%
	H4	0.024	77%
	H5	0.031	100%
	H6s	0.031	100%
scaffold	1, 1', 1''	0.012	38%
	2	0.008	25%
	3	0.01	32%
	4, 4'	0.01	32%
	5	0.008	25%
	6, 6'	0.005	16%
	7, 7'	0.01	32%

-1.0 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.0078	35%
	H2	0.011	50%
	H3	-	-
	H4	0.009	45%
	H5	-	-
	H6, H6'	-	-
	Ac	0.016	72%
Galβ	H1	-	-
	H2	0.016	72%
	H3	0.016	72%
	H4	0.017	77%
	H5	0.022	100%
	H6s	0.022	100%
scaffold	1, 1', 1''	0.008	36%
	2	0.007	31%
	3	0.009	40%
	4, 4'	0.009	45%
	5	0.007	31%
	6, 6'	0.005	22%
	7, 7'	0.009	45%

Figure 9.9. STD of galectin-1 and ligand **2**. **A**) Sample: galectin-1 100 μM, **2** 3 mM (ratio lectin:ligand = 1:30). Off-resonance (irradiation at 100 ppm) and STD spectra (irradiation at 0.0, -0.5 and -1 ppm, 50×). **B**) Relative STD-AF intensities and STD percentage for each proton signal calculated on the basis of the relative experiment.

A)



B)

0.0 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	$H1_{ext}$	0.012	25%
	$H1_{int}$	0.017	35%
	$H2_{ext}$	0.02	42%
	$H2_{int}$	0.025	52%
	H3	-	-
	H4	0.017	35%
	H5	-	-
	H6, H6'	-	-
	Ac_{ext}	0.036	75%
	Ac_{int}	0.048	100%
Galβ	H1	-	-
	H2	0.03	62%
	H3	0.03	62%
	H4	0.034	70%
	H5	0.042	87%
	H6s	0.042	87%
scaffold	1, 1', 1''	0.023	47%
	2	0.015	31%
	3	0.016	33%
	4, 4'	0.0085	17%
	6, 6'	0.009	18%
	6''	0.01	20%
	7, 7'	0.0073	15%
	7''	0.012	25%

-0.5 ppm

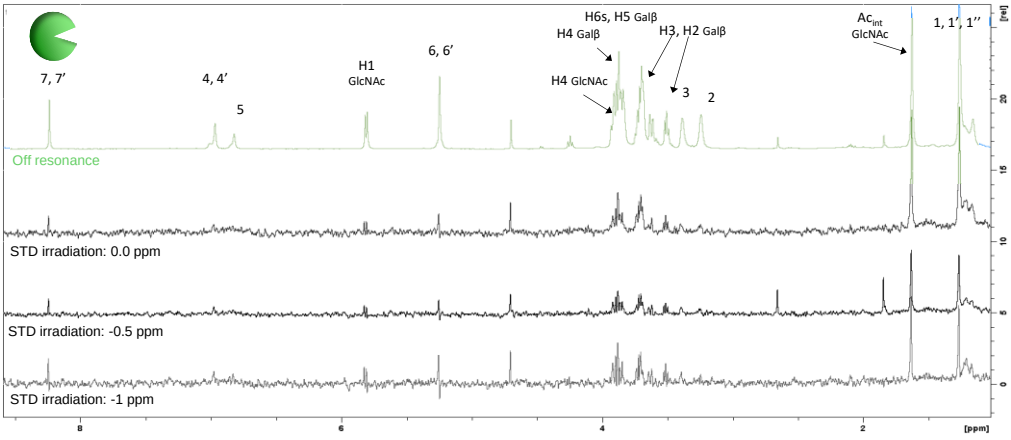
Residue	Atom	STD-AF	STD %
GlcNAc	$H1_{ext}$	0.01	31%
	$H1_{int}$	0.014	45%
	$H2_{ext}$	0.015	48%
	$H2_{int}$	0.016	52%
	H3	-	-
	H4	0.012	38%
	H5	-	-
	H6, H6'	-	-
	Ac_{ext}	0.023	74%
	Ac_{int}	0.031	100%
Galβ	H1	-	-
	H2	0.019	61%
	H3	0.019	61%
	H4	0.02	64%
	H5	0.027	87%
	H6s	0.027	87%
scaffold	1, 1', 1''	0.013	42%
	2	0.012	38%
	3	0.019	61%
	4, 4'	0.007	22%
	6, 6'	0.0075	24%
	6''	0.009	29%
	7, 7'	0.007	22%
	7''	0.009	29%

-1 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	$H1_{ext}$	0.005	31%
	$H1_{int}$	0.008	50%
	$H2_{ext}$	0.008	50%
	$H2_{int}$	0.009	56%
	H3	-	-
	H4	0.008	50%
	H5	-	-
	H6, H6'	-	-
	Ac_{ext}	0.011	68%
	Ac_{int}	0.016	100%
Galβ	H1	-	-
	H2	0.01	62%
	H3	0.01	62%
	H4	0.011	68%
	H5	0.015	93%
	H6s	0.015	93%
scaffold	1, 1', 1''	0.0066	41%
	2	0.007	43%
	3	0.008	50%
	4, 4'	0.0045	28%
	6, 6'	0.0045	28%
	6''	0.006	37%
	7, 7'	0.0025	15%
	7''	0.004	25%

Figure 9.10. STD galectin-1 and ligand **3**. A) Sample: galectin-1 100 μM, **3** 3 mM (ratio lectin:ligand = 1:30). Off-resonance (irradiation at 100 ppm) and STD spectra (irradiation at 0.0, -0.5 and -1 ppm, 50×). B) Relative STD-AF intensities and STD percentage for each proton signal calculated on the basis of the relative experiment.

A)



B)

0.0 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.008	42%
	H2	0.009	47%
	H3	-	-
	H4	0.007	36%
	H5	-	-
	H6, H6'	-	-
	Ac	0.019	100%
Galβ	H1	-	-
	H2	0.01	52%
	H3	0.009	47%
	H4	0.0095	50%
	H5	0.011	57%
	H6s	0.011	57%
scaffold	1, 1', 1''	0.015	78%
	2	0.006	31%
	3	0.007	36%
	4, 4'	0.008	42%
	5	0.01	52%
	6, 6'	0.006	31%
	7, 7'	0.008	42%

-0.5 ppm

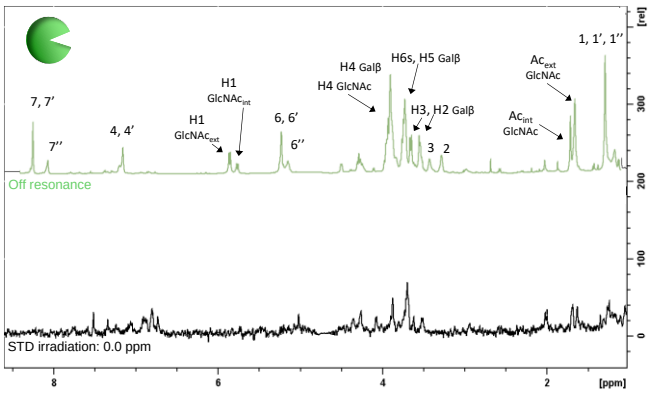
Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.0087	60%
	H2	0.009	62%
	H3	-	-
	H4	0.0069	48%
	H5	-	-
	H6, H6'	-	-
	Ac	0.0143	100%
Galβ	H1	-	-
	H2	0.0098	68%
	H3	0.0082	57%
	H4	0..0077	53%
	H5	0.0085	59%
	H6s	0.0085	59%
scaffold	1, 1', 1''	0.0105	73%
	2	0.0055	38%
	3	0.0078	54%
	4, 4'	0.01	69%
	5	0.011	76%
	6, 6'	0.0066	46%
	7, 7'	0.01	76%

-1.0 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	H1	0.009	75%
	H2	0.01	83%
	H3	-	-
	H4	0.0075	62%
	H5	-	-
	H6, H6'	-	-
	Ac	0.012	100%
Galβ	H1	-	-
	H2	0.01	83%
	H3	0.008	66%
	H4	0..008	66%
	H5	0.008	66%
	H6s	0.008	66%
scaffold	1, 1', 1''	0.0085	70%
	2	0.0050	41%
	3	0.0075	62%
	4, 4'	0.01	83%
	5	0.012	100%
	6, 6'	0.0077	64%
	7, 7'	0.01	83%

Figure 9.11. STD galectin-3 CRD and ligand **2**. A) Sample: galectin-3 CRD 50 μM, **2** 1.5 mM (ratio lectin:ligand = 1:30). Off-resonance (irradiation at 100 ppm) and STD spectra (irradiation at 0.0, -0.5 and -1 ppm, 50×). B) Relative STD-AF intensities and STD percentage for each proton signal calculated on the basis of the relative experiment.

A)



B)

-0.5 ppm

Residue	Atom	STD-AF	STD %
GlcNAc	$H1_{ext}$	0.004	44%
	$H1_{int}$	0.009	100%
	$H2_{ext}$	0.0066	73%
	$H2_{int}$	0.0066	73%
	H3	-	-
	H4	0.003	33%
	H5	-	-
	H6, H6'	-	-
	Ac_{ext}	0.004	44%
	Ac_{int}	0.0066	73%
Galβ	H1	-	-
	H2	0.006	66%
	H3	0.0066	73%
	H4	0.0045	50%
	H5	0.0085	94%
	H6s	0.0085	94%
scaffold	1, 1', 1''	0.0025	27%
	2	0.003	33%
	3	0.003	33%
	4, 4'	0.0028	31%
	6, 6'	0.0028	31%
	6''	0.005	55%
	7, 7'	0.0016	17%
	7''	0.004	44%

Figure 9.12. STD galectin-3 CRD and ligand **3**. A) Sample: galectin-3 CRD 50 µM, **3** 1.5 mM (ratio lectin:ligand = 1:30). Off-resonance (irradiation at 100 ppm) and STD spectrum (irradiation at -0.5, 50×). B) Relative STD-AF intensities and STD percentage for each proton signal calculated on the basis of the experiment.

9.3 Supporting Information of Chapter 5

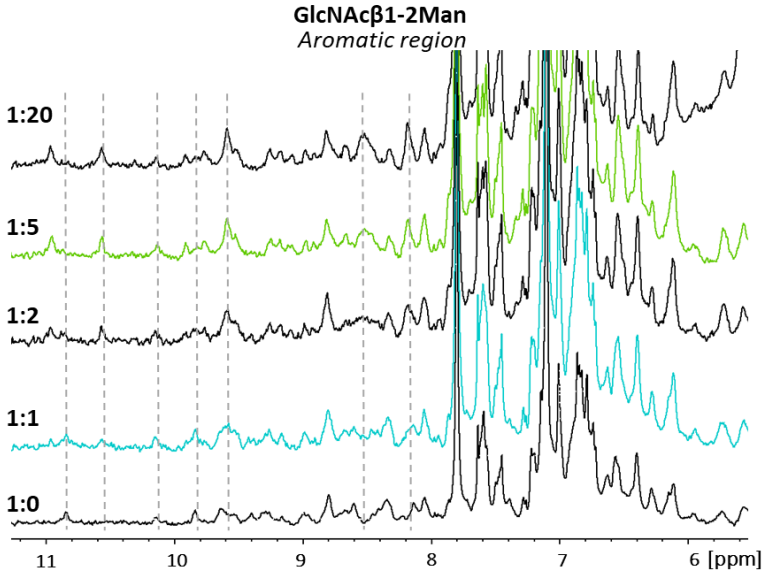


Figure 9.13. LSECtin:disaccharide titration. Stacked ^{1}H-NMR spectra acquired during the titration of disaccharide GlcNAcβ1-2Man to a sample containing LSECtin CRD (120 µM in deuterated buffer). The relative lectin:ligand ratios are reported above each spectrum. Expansion of the low field region of the spectra (aromatic region) with signal suffering strong changes during the titration highlighted.

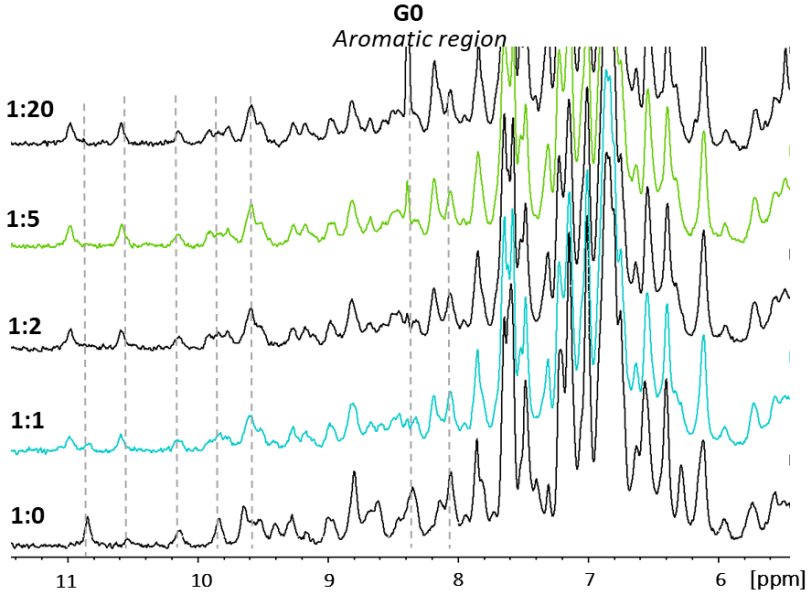


Figure 9.14. LSECtin:non-elongated N-glycan (**G0**) titration. Stacked ^{1}H-NMR spectra acquired during the titration of **G0** to a sample containing LSECtin CRD (120 µM in deuterated buffer). The relative lectin:ligand ratios are reported above each spectrum. Expansion of the low field region of the spectra (aromatic region) with signal suffering strong changes during the titration highlighted.

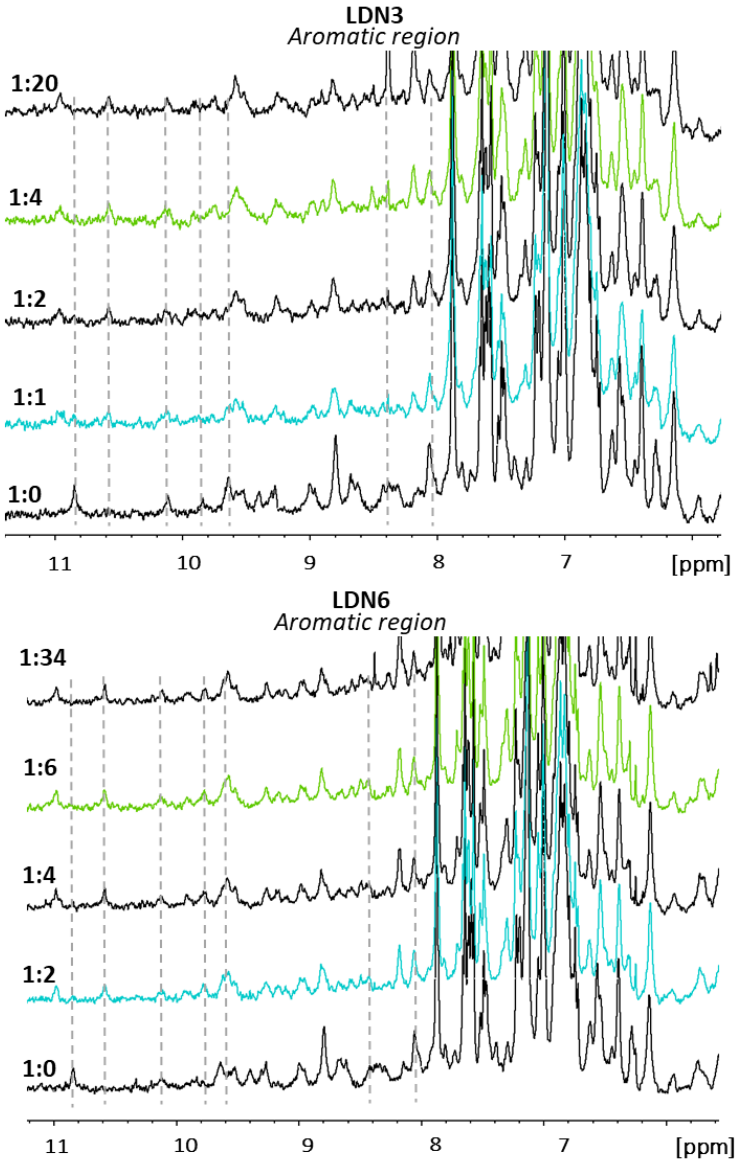


Figure 9.15. LSECtin:LDN3/LDN6 N-glycan titration. Stacked ^{1}H-NMR spectra acquired during the titration of asymmetric N-glycans (on the top **LDN3** and on the bottom **LDN6**; amounts of equivalents reported for each spectrum). LSECtin CRD was employed at a concentration of 82 µM in deuterated buffer. Expansion of the low field region (aromatic region) of the spectra with signal suffering strong changes during the titration highlighted.

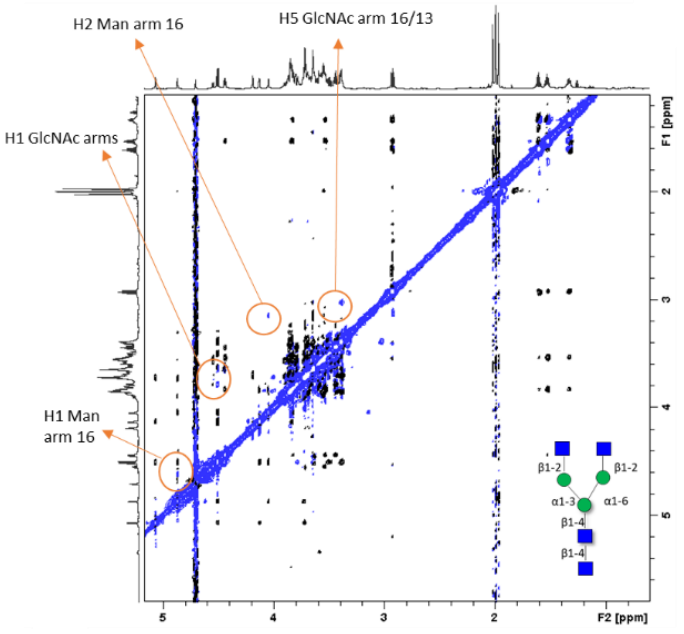
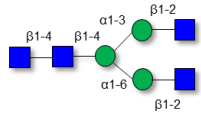


Figure 9.16. 2D ROESY NMR of LSECtin and **G0** glycan. Lectin:ligand molar ratio employed was 1:10 (being LSECtin CRD at a concentration of 120 μM). ROESY-NMR experiment acquired with 150 ms of mixing time. Chemical exchange-mediated crosspeaks of the ligand protons are pointed with arrows.

LSECtin and G0

EXSY calc	k_{off} (s^{-1})
H1 Man (arm 1-6)	0.211
H2 Man (arm 1-6)	0.140
H5 GlcNAc (16 or 13)	0.276
AVERAGE k_{off}	**0.209**

Extimated K_D (*Powlesland at al, 2007 – competition assay*) = 2.6 (μM)
k_{off} = 0.209 (s^{-1})
K_{on}= 80384 ($M^{-1} * s^{-1}$)



K_D from literature
k_{off} Determined from EXSY
k_{on} Calculated from K_D and k_{off} ($K_D = k_{off}/K_{on}$)

LSECtin and GlcNAcβ1-2Man

EXSY calc	k_{off} (s^{-1})
H1 GlcNAc	1.091
H1 Man (arm 1-6)	1.041
H2 Man (arm 1-6)	0.654
AVERAGE k_{off}	**0.928**

Extimated K_D (*Powlesland at al, 2007 – competition assay*) = 3.5 (μM)
k_{off} = 0.928 (s^{-1})
K_{on}= 265142 ($M^{-1} * s^{-1}$)

Kon and Koff values are faster for the disaccharide

Figure 9.17. EXSY Calculation for LSECtin/Disaccharide and LSECtin/**G0** systems. Both rates are faster for the disaccharide.