UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular I

TESIS DOCTORAL

Aplicación y nuevos desarrollos de la espectroscopía por RMN para el estudio de procesos de reconocimiento molecular entre carbohidratos y sus receptores

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Manuel Álvaro Berbís Moreno

Directores
Jesús Jiménez Barbero
Javier Cañada Vicinay

Madrid, 2014

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Químicas

Departamento de Bioquímica y Biología Molecular I

Aplicación y nuevos desarrollos de la espectroscopía por RMN para el estudio de procesos de reconocimiento molecular entre carbohidratos y sus receptores

Tesis Doctoral

Manuel Álvaro Berbís Moreno

Madrid, 2014

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Químicas

Departamento de Bioquímica y Biología Molecular I

Aplicación y nuevos desarrollos de la espectroscopía por RMN para el estudio de procesos de reconocimiento molecular entre carbohidratos y sus receptores

Memoria para optar el grado de doctor presentada por

Manuel Álvaro Berbís Moreno

Madrid, 2014

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

Bajo la codirección de

Dr. Jesús Jiménez Barbero

Dr. Javier Cañada Vicinay

ACKNOWLEDGEMENTS

This Thesis has been done at the Center for Biological Research of the Spanish National Research Council (CIB-CSIC), under the supervision of Jesús Jimenez-Barbero and Francisco Javier Cañada Vicinay, and thanks to an FPI Fellowship granted by the Spanish Ministry of Economy and Competitiveness (MINECO).

I wish to express my gratitude to my supervisors for giving me the opportunity to work in their team. Also, I would like to thank the rest of the team for helping me throughout these years.

Contents

ABBREVIATIONS

ADME: absorption, distribution, metabolism and excretion

BMRB: Biological Magnetic Resonance Data Bank

BSA: bovine serum albumin

CORCEMA: complete relaxation and conformational exchange matrix

COSY: correlation spectroscopy

CRD: Carbohydrate recognition domain

DMSO: dimethylsulfoxide

DNA: deoxyribonucleic acid

DOSY: Difussion-ordered spectroscopy

EDTA: ethylenediaminetetraacetic acid

FBDD: Fragment-based drug discovery

GAG: Glycosaminoglycan

Gal: galactose

Gal-1S: 1-β-thiogalactose

Gal-3: galectin-3

Gal-7: galectin-7

Glc: glucose

Glc-1-P: glucose-1-phosphate

GlcNAc: N-acetyl glucosamine

gp17: gene product 17 of phage T7

gp37: gene product 37 of phage T4

GPCR: G protein-coupled receptor

HMBC: heteronuclear multiple bond correlation

HOESY: Heteronuclear Overhauser enhancement spectroscopy

HSQC: Heteronuclear single-quantum coherence

HTS: high-throughput screening

Ig: immunoglobulin

ILOE: inter-ligand nuclear Overhauser enhancement

ITC: isothermal titration calorimetry

J: scalar coupling constant

K: desulfated keratan

K_D: dissociation constant

KS: keratan sulfate

LacNAc: N-acetyl lactosamine (Gal-β(1-4)GlcNAc)

LBT: lanthanide binding tag

LE: ligand efficiency

Ln^{3+}: lanthanide (III) ion

LPS: lipopolysaccharide

mAb: Monoclonal antibody

MD: molecular dynamics

MM: molecular mechanics

ND: N-terminal domain of gal-3

NMR: Nuclear magnetic resonance

NOE: Nuclear Overhauser enhancement

NOESY: NOE spectroscopy

PCS: Pseudo-contact shift

PDB: Protein Data Bank

PhDTA: phenylendiaminetetraacetic acid

PPBS: phenoxyphenyl-binding site of gal-7

PRE: Paramagnetic relaxation enhancement

RDC: residual dipolar coupling

RNA: ribonucleic acid

SLAPSTIC: spin labels attached to protein side chains as tool to identify interacting compounds

spUGP: UDP.glucose pyrophosphorylase of *Streptococcus pneumoniae*

STD: Saturation transfer difference

STDD: saturation transfer double difference

TAdV-3: turkey adenovirus 3

TB: tuberculosis

TFAM: trifluoroacetylcysteine methanethiosulfonate

TMS: tetramethylsilane

TOCSY: total correlation spectroscopy

trNOESY: Transferred NOE spectroscopy

TROSY: transverse relaxation optimized spectroscopy

TSP: trimethylsilylpropionate

UDA: *Urtica dioica* agglutinin

UDP-glucose:

UGP: UDP-glucose pyrophosphorylase

UTP:

WaterLOGSY: water-ligand observed via gradient spectroscopy

WGA: wheat germ agglutinin

α-Gal epitope: Galα-1,3-Galβ-1,4-GlcNAc-R

τ_c: correlation time

ABSTRACT

Carbohydrates are among the most ubiquitous and complex types of biomolecules in nature. Their interplay with a number of receptor families including lectins, antibodies, enzymes and viruses, triggers a variety of responses related to physiological and pathological events. Thus, knowledge on the forces that govern these interactions at atomic detail is essential for a thorough understanding of many vital processes and to open the possibility of their modulation.

At the same time, the high structural variability and functional versatility of glycans call for developing new tools to probe carbohydrate conformation and recognition features by receptors.

In this Thesis, we have made extensive use of NMR spectroscopy methods to gain structural insight into the interactions of a series of protein-carbohydrate systems of biomedical relevance. In addition, we present an innovative approach for conformational characterization of sugars and structural aspects of their recognition by receptors, based on the exploitation of paramagnetic effects arisen from lanthanide-binding carbohydrate conjugates.

RESUMEN

Los carbohidratos constituyen uno de los tipos de biomoléculas más complejos y ubicuos, cuyo reconocimiento por parte de receptores proteicos pertenecientes a diversas familias (lectinas, anticuerpos, enzimas y virus) determina una variedad de respuestas relacionadas con procesos fisiológicos y patológicos. El conocimiento, a nivel atómico, de la estructura de los carbohidratos y de sus interacciones con proteínas es esencial para una completa comprensión de muchos procesos vitales, así como abrir la puerta a su modulación.

Al mismo tiempo, la gran variabilidad estructural y versatilidad funcional de los glicanos impone la necesidad de desarrollar nuevas herramientas que permitan ampliar nuestro conocimiento sobre la estructura de estas moléculas y su reconocimiento por parte de receptores.

En esta Tesis se ha hecho un uso exhaustivo de métodos de espectroscopía por resonancia magnética nuclear (RMN) para obtener información estructural sobre procesos de reconocimiento molecular en una serie de sistemas carbohidrato-proteína de interés biomédico.

En distintos capítulos, esta Tesis aborda el estudio del reconocimiento molecular entre glicanos y los cuatro tipos principales de receptores de carbohidratos: lectinas, anticuerpos, enzimas y virus.

- Capítulo 1. Introducción

La introducción de esta Tesis centra su atención en la gran variabilidad estructural de los carbohidratos, que dota a estas moléculas de un elevado potencial para codificar información biológica, así como los distintos tipos de receptores proteicos de carbohidratos y sus características más notables.

Asimismo, se presenta una revisión comprensiva de los métodos de RMN más utilizados en la actualidad, con especial énfasis en su aplicación al estudio estructural de los carbohidratos y de sus interacciones con receptores. Dichos métodos incluyen experimentos basados en la observación de las señales de RMN de los ligandos, así como en la observación de las señales

de RMN de las proteínas. De entre los primeros, destacamos el experimento de diferencia de transferencia de saturación (STD, por sus siglas en inglés). Esta técnica se basa en la transferencia de magnetización entre el receptor, que es objeto de irradiación selectiva mediante pulsos de radiofrecuencia, y aquellos ligandos que, en exceso y en régimen de intercambio rápido, se unan al receptor. Entre los métodos basados en la observación de las señales de la proteína, debemos destacar los experimentos basados en la correlación heteronuclear de cuanto simple protón-nitrógeno-15 (^{1}H-^{15}N HSQC, por sus siglas en inglés).

También se profundiza en otros tipos de experimentos de RMN, tales como los destinados a la asignación de moléculas pequeñas y de proteínas, métodos basados en difusión traslacional o en la observación de núcleos distintos de protón, especialmente flúor-19. Asimismo, se presenta el concepto de descubrimiento de fármacos basado en fragmentos (FBDD, por sus siglas en inglés), así como aspectos prácticos del uso de RMN aplicado a este procedimiento.

- Capítulo 2. El reconocimiento entre determinantes de tipo carbohidrato y un conjunto de proteínas fibrilares víricas

En este capítulo se presenta el estudio del reconocimiento entre una serie de proteínas fibrilares de distintas especies víricas y sus ligandos sacarídicos, utilizando experimentos basados en la observación de las señales de ligando. En concreto, se han llevado a cabo experimentos de diferencia de transferencia de saturación (STD) para detectar la unión reversible entre estos receptores y sus ligandos, así como la obtención de información de epítopo desde el punto de vista del ligando.

En concreto, se ha estudiado las capacidades de interacción de tres de estos receptores virales: la proteína gp37 del fago T4, la proteína gp17 del fago T7, y el receptor fibrilar del adenovirus de pavo 3 (TAdV-3).

- Capítulo 3. Las propiedades de interacción de un anticuerpo monoclonal contra el α-glucano de la cápsula de *Mycobacterium*

En el Capítulo 3 se presenta el estudio de la interacción entre un anticuerpo monoclonal y el polisacárido glucógeno, así como una serie de oligosacáridos pequeños relacionados estructuralmente. Estos estudios han conducido a la identificación del epítopo del carbohidrato responsable de la unión al anticuerpo.

Asimismo, se han llevado a cabo experimentos de RMN en células de *Mycobacterium smegmatis* destinados a obtener información estructural de su α-glucano capsular.

- Capítulo 4. Estudios sobre la UDP-glucosa pirofosforilasa de *Streptococcus pneumoniae*, un candidato a diana terapéutica

El Capítulo 4 aglutina distintos estudios en torno a la UDP-glucosa pirofosforilasa (UGP) de *Streptococcus pneumoniae* (spUGP), una enzima que cataliza la transformación reversible de UTP y glucosa-1-P en UDP-glucosa y pirofosfato. En primer lugar, se ha desarrollado una serie de experimentos destinados a adquirir información de carácter mecanístico sobre esta enzima. Mediante experimentos de STD, se ha obtenido información respecto al modo secuencial en que los sustratos se unen a la enzima: en primer lugar se produce la unión de UTP al centro activo, seguido de glucosa-1-P, siendo el Mg^{2+} esencial para la unión de los sustratos.

También se ha desarrollado un ensayo de actividad enzimática basado en RMN, mediante el seguimiento de la interconversión entre el H1 de glucosa-1-P y el H1'' de UDP-glucosa. Este ensayo representa una estrategia directa para el seguimiento de la actividad UGP, y es ventajoso sobre el método más utilizado actualmente, basado en la medición indirecta de la actividad UGP mediante su acoplamiento con la oxidación de la UDP-glucosa seguida por procedimientos espectrofotométricos.

Finalmente, se llevó a cabo un cribado de fragmentos preliminar contra spUGP, encontrándose una serie de compuestos estructuralmente relacionados que se unían con moderada afinidad al subsitio de unión de UTP de spUGP.

Capítulo 5: Galectinas

El Capítulo 5 presenta diferentes estudios relacionados con galectinas, una familia de lectinas animales con capacidad de unión a β-galactósidos. En términos estructurales, se distingue entre tres tipos de galectinas: (i) prototipo, que constan de un dominio de reconocimiento de carbohidratos (CRD) que usualmente dimeriza; (ii) en tándem, que constan de dos CRDs distintos en la misma cadena polipeptídica; y (iii) quimera, que constan de un CRD y una porción no-CRD. Hasta la fecha, la única galectina quimera identificada es galectina-3 (gal-3). El Capítulo centra su atención en gal-3 y en galectina-7 (gal-7), una galectina prototipo.

Los estudios sobre gal-7 se iniciaron con la asignación completa de sus señales de RMN mediante experimentos bi y tridimensionales. Dicha asignación reveló la existencia de dos confórmeros de la proteína en disolución, en proporciones aproximadas 55:45, en equilibrio de intercambio lento en la escala de tiempo del desplazamiento químico, que posteriores estudios atribuyeron a la isomerización *cis*-*trans* del enlace peptídico V3-P4. También se llevó a cabo un cribado de fragmentos, hallándose una serie de compuestos pequeños que se unían a gal-7 en dos epítopos distintos: uno, previamente desconocido, homólogo estructural del sitio de galectina-1 de unión a farnesil-Ras, y un segundo en la interfaz de dimerización de gal-7. La unión de fragmentos en la interfaz de dimerización mostró capacidad para inhibir la autoasociación de gal-7, a la vez que contribuyó a demostrar el efecto estabilizador del dímero de gal-7 promovido por su unión a lactosa.

Asimismo, los estudios sobre gal-3 se iniciaron por la asignación de las señales de RMN de la proteína completa. Se llevó a cabo estudios de interacción de pequeños oligosacáridos, así como polisacáridos (keratan sulfato y keratan desulfatado) con el CRD de gal-3. Finalmente, se estudió la capacidad de interacción entre el CRD de gal-3 y péptidos derivados del dominio N-terminal de gal-3, sin y con fosforilación en posiciones fisiológicamente relevantes. Los resultados mostraron la capacidad de algunos de estos péptidos de interaccionar con el CRD de manera dependiente de su estado de fosforilación.

- Capítulo 6. Nuevos desarrollos: aplicación del paramagnetismo para la estudiar la conformación de azúcares y su interacción con receptores

Finalmente, el Capítulo 6 presenta el desarrollo de nuevas metodologías relacionadas con el aprovechamiento de efectos paramagnéticos en el campo de los carbohidratos. Dichas nuevas metodologías están destinadas a la obtención de información de tipo estructural sobre la conformación de oligosacáridos y aspectos estructurales de su unión a receptores.

CHAPTER 1

INTRODUCTION

1.1 Carbohydrates and Glycoscience

1.1.1 General aspects

All cell surfaces and most proteins are coated by carbohydrates that intrinsically contain structural information used in recognition processes relevant in health and disease. These are secondary gene products that are entirely missed by the current screening techniques relying solely on genomics. Glycans (carbohydrates, sugars) are expressed on glycoproteins, glycolipids, proteoglycans and free complex carbohydrates and the entire category of these structures, known as the glycome, is probably the least studied and most complicated and biologically important molecular classifications in humans.

In particular, carbohydrates are among the most ubiquitous and complex types of biomolecules in nature. They are composed of smaller units, called monosaccharides, which are linked in the form of linear sequences, with or without sophisticated branches [1-3], allowing for a high potential for overall structural diversity. In addition, the introduction of substituents at the hydroxyl groups, such as phosphate, sulfate [4] and/or acetyl [5] groups, greatly increases carbohydrates' potential for complexity. This structural diversity allows carbohydrates to be specifically recognized by receptors with exquisite selectivity, making them very suitable for encoding biological information [6].

Carbohydrates, especially when conjugated as glycoproteins or glycolipids at the periphery of the cell, thus play a very important role in molecular recognition processes by establishing specific interactions that trigger a variety of intra-, extra-cellular physiological and pathological events (fig. 1). Thus, modifications of cell-surface glycosylation may lead to or be involved in pathological phenomena, such as inflammation and cancer [7-9]. In this context, knowledge at

the maximum resolution possible on the 3D structure of carbohydrates, and characterization of their recognition features by different entities, such as lectins, enzymes, viruses, and antibodies, are essential for a thorough understanding of many vital processes and to open the possibility of their modulation. Therefore, glycoscience, a discipline lying at the interface between chemistry, biology, and biomedicine is today a challenging and active field of research. Glycoscience represents a fruitful and innovative source of novel therapeutics that is relatively untapped. In fact, there is a current need to produce innovative therapeutic solutions based on the exploitation of the emerging opportunities that exist in this emerging field.

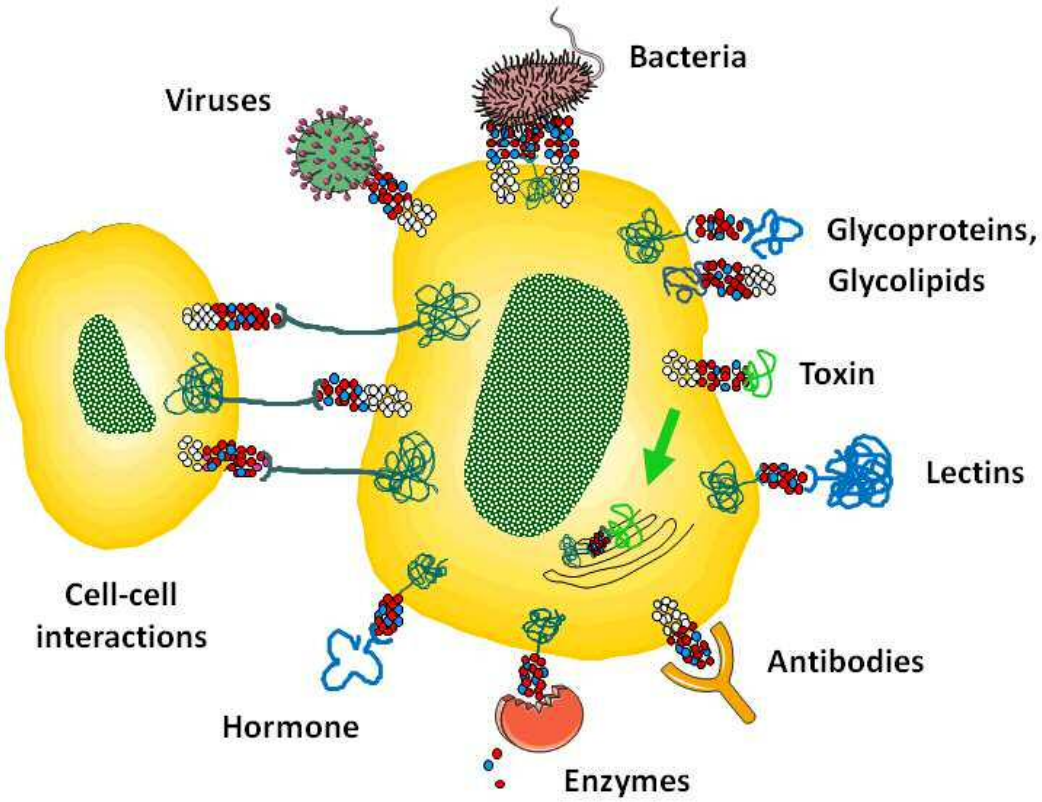


Figure 1: The importance of protein-carbohydrate interactions. Cell-surface carbohydrates are recognized by a diversity of protein families, including lectins, enzymes, antibodies and viruses, triggering a variety of physiological and pathological responses.

1.1.2 Structure of carbohydrates

1.1.2.1 Monosaccharides

Carbohydrate molecules are built from small building blocks, called monosaccharides. Many different monosaccharides exist in Nature. They may have different number of carbon atoms, usually from five to seven, although other possibilities also exist.

The high potential for diversity displayed by carbohydrates is originated, at the monosaccharide level, by a number of sources of structural variability. First, due to the presence of different substituents for every carbon atom, these become stereogenic centers (fig. 2). The absolute configuration at each chiral center of the sugar backbone defines the type of sugar (e.g. glucose, galactose, mannose). In addition, the orientation of the substituent in the anomeric position, i.e. axial or equatorial in respect to the plane containing the sugar ring, defines the α or β anomers, respectively.

Finally, the sugar ring conformation introduces another source of structural variability [10] which is of special relevance for the molecular recognition of sugars by receptors, since it determines the orientation of functional groups in the three-dimensional space. For the common pyranose rings, the regular geometry is one of the possible chair conformations, which minimizes steric congestion and accounts for the anomeric effect. However, monosaccharide rings may populate a number of additional, distinct three-dimensional shapes, at relative ratios that depend on the free energy of the different conformers. This, in turn, depends heavily on the orientation, nature and number of substituents at the sugar ring. For certain sugars in solution, rapid interconversion may occur between the major chair conformations and some other minor geometries [11], which may include envelope, boat and skew ring puckers, as well as alternative chairs (fig. 2).

1.1.2.2 The anomeric effect

The anomeric effect was first described in the 1950s by J. T. Edward and R. U. Lemieux [12]. In carbohydrate chemistry, the anomeric effect describes the tendency of the electronegative substituents at the anomeric carbon (C1) of a pyranose ring to adopt the axial rather than the equatorial orientation. This is in contrast to what would be expected based solely on terms of steric factors. Normally, for D-sugars, the α isomer is more stable than expected (fig. 2).

The origin of the anomeric effect has been a matter of historical controversy. Over the last decades, modeling approaches including *ab initio* calculations and molecular mechanics studies [13-14] have permitted to gain insight into all factors that come into play to produce this phenomenon. On the basis of these studies, the anomeric effect is now understood as the result of multiple steric, electrostatic and stereoelectronic interactions. Although the anomeric effect is a well-known phenomenon for carbohydrate chemists, it is not exclusive of the sugar field, appearing also in many other types of molecules.

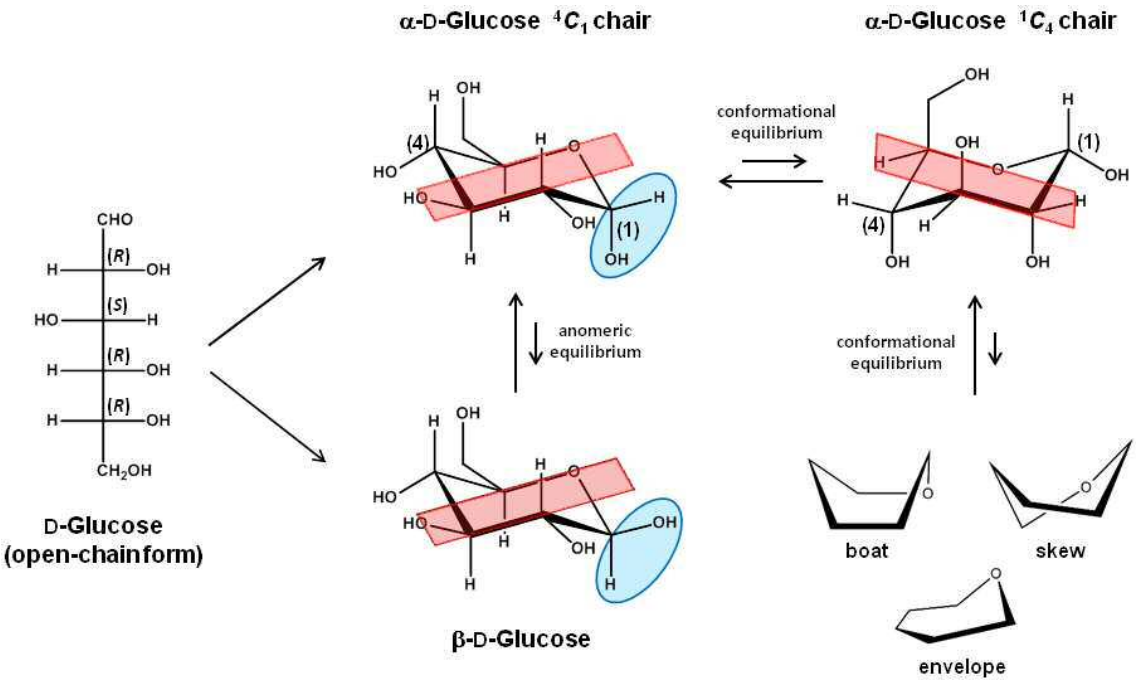


Figure 2: Structural variability of a monosaccharide: glucose. Schematic representation of D-glucose in its open-chain form. The absolute configuration is labeled with either the (*R*) or (*S*) descriptors for each stereocenter. The substituents at each carbon atom can be at axial or equatorial disposition with respect to the imaginary plane containing the C2, C3, C5 and O5 atoms. In β-D-glucose, the orientation of the hydroxyl group in C1 is equatorial, while in the α anomer it is axial (shadowed in blue). Both anomers can coexist, but one of them is usually preferred. The energetically privileged chair conformations of the cyclic sugar ring are labeled as $^4C_1$ or $^1C_4$, depending on the position of carbons 1 and 4, i.e. above or below the sugar ring plane (shadowed in red), respectively. Many other ring puckers can coexist in conformational equilibrium, populating boat, skewed and envelope forms.

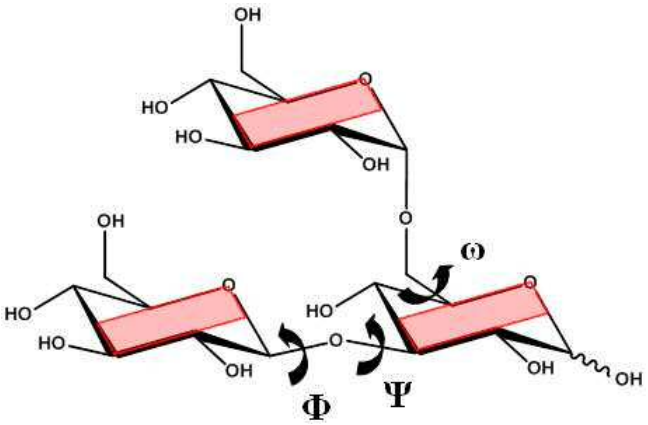


Figure 3: Torsion angles Φ, Ψ and ω in a model oligosaccharide, which determine the relative orientation of the different sugar ring planes (shadowed in red). All *O*-glycosidic linkages are defined by Φ and Ψ. In addition, 1-6 linkages demand to determine an additional angle (ω) around C5-C6 to provide the relative orientation of the two sugar planes.

1.1.2.3 Disaccharides. The glycosidic linkage

One given monosaccharide is attached to other one through a glycosidic linkage. The anomeric oxygen of one particular monosaccharide is attached to a hydroxyl group of the second sugar moiety to build one disaccharide. The conformations around the glycosidic linkage are defined by the torsion angles Φ and Ψ (fig. 3). Obviously, not every possible combination of these angles is permitted from the energetic viewpoint, as certain geometries would give rise to steric clashes, while others provide stabilizing inter-residue van der Waals interactions. The glycosidic torsion angle Φ is restricted by the exo-anomeric effect, which favors a value of Φ of ca. +50° in the case of β-D- and α-L-hexopyranosides, and -50° in α-D- and β-L-hexopyranosides [15]. The origin and implications of the exo-anomeric effect to define the geometry of oligosaccharides and their mimetics remain a matter of investigation [16] Finally, 1-6 linkages demand the determination of an additional angle (ω) around C5-C6 to provide the relative orientation of the two sugar planes (fig. 3).

1.1.2.4 Increasing complexity: oligosaccharides to polysaccharides

As the number of sugar units increase, so does carbohydrate complexity. Depending on the nature and the position of the sugar substituents, different oligosaccharides may adopt a variety of shapes and display distinct dynamic features. For instance, they may resemble chain-like structures, as in chitooligosaccharides, which are linear β(1-4) linked *N*-acetylglucosamine (GlcNAc) oligomers (fig. 4A) [17], or branched structures, with special mention to the blood

group determinant oligosaccharides (the ABH-system), and the GM1 oligosaccharide (fig. 4B), the sugar part of the corresponding ganglioside. Their conformational properties, adaptability and recognition features have been and still are deeply analyzed [15].

Further increase in carbohydrate size leads to polysaccharides. They are one of the most important classes of biological polymers, adopting a variety of functions, from energy storage, as in glycogen or starch, to structural composition of plant (cellulose) or fungal cells (chitin).

The physical properties of polysaccharides may differ substantially from those of their monosaccharide building blocks. Often, they are amorphous and insoluble in water, or may display viscoelastic properties [18]. They may also differ in chemical complexity, as they can be composed of one single repeating unit (e.g. amylose) or incorporate many different types of sugar building blocks, and even aglyconic parts, such as the bacterial lipopolysaccharides (LPS) (fig. 4C, D).

Another important class of oligo- and polysaccharides are glycosaminoglycans (GAGs), a term which encompasses a variety of sulfated amino-containing glycans. They include heparin, heparan sulfate, hyaluronic acid and keratan sulfate, which perform important structural functions, and whose recognition by different receptors triggers a plethora of biological responses, such as inflammation, cell adhesion, and regulation of cell growth and proliferation [19-20].

In summary, carbohydrates are structurally complex molecules, whose diversity emerges from a number of sources of variability at different levels of hierarchy. Carbohydrates' potential for structural diversity surpasses those of proteins and nucleic acids [21-22]. This endows them with the capacity to store biological information. Through molecular recognition, carbohydrate messages are interpreted by sugar receptors, which translate them into biological function.

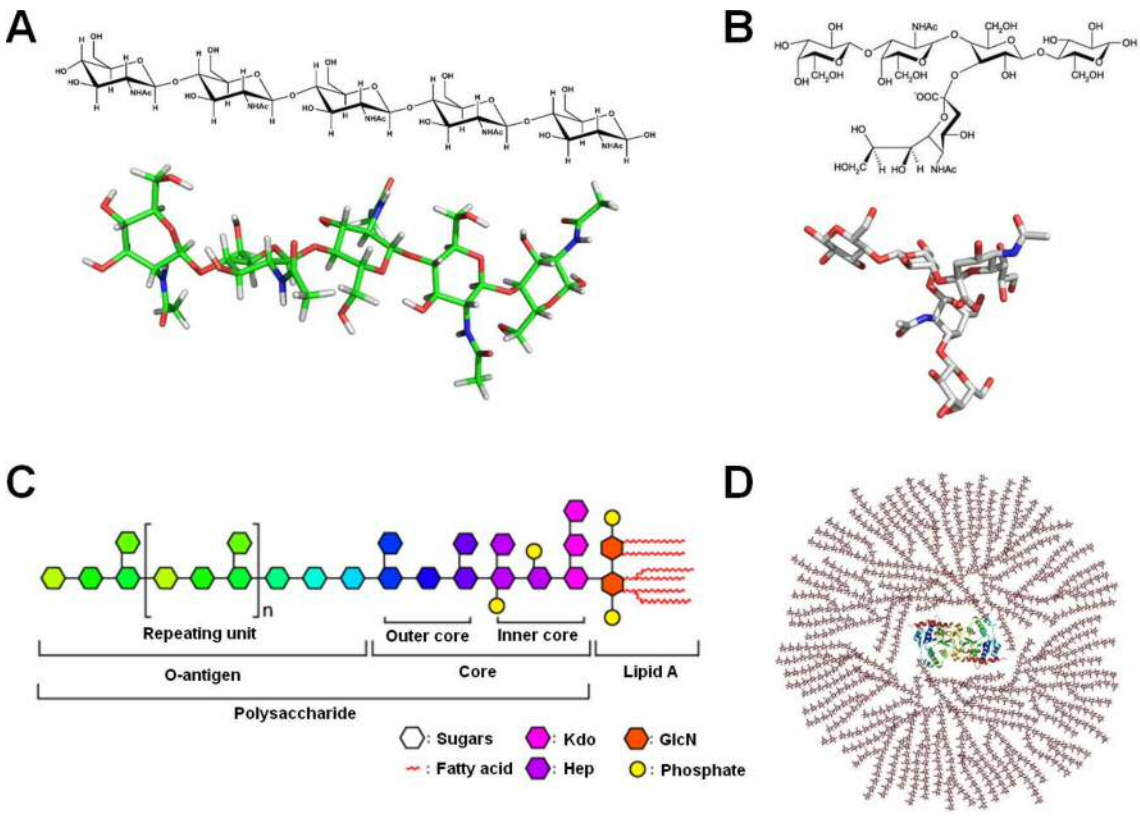


Figure 4: Oligo- and polysaccharides. A: Formula and 3D structure of chitopentaose as a linear oligosaccharide. B: Formula and 3D structure of the GM1 pentasaccharide as a branched oligosaccharide. C: Schematic structure of lypopolysaccharide. D: Schematic 2D cross-sectional view of glycogen, including a core molecule of glycogenin. The whole glycogen granule may contain more than 30,000 units of glucose [23].

1.2 Molecular recognition of carbohydrate and their receptors

The notion of carbohydrates being specifically recognized by proteins dates back to the 'lock-and-key' principle used by Emil Fisher as early as 1894 [24] to illustrate the complementarity between a glucoside and an enzyme that permits the one to fit into the other. Today, knowledge on the intrinsic flexibility and variability of carbohydrates has reformulated the older paradigm as a 'bunch of keys', [25] each of which may be selected by one or more among a 'bunch of locks'. This concept unifies the existing possibilities of induced-fit- or conformational-selection-based molecular recognition processes.

The advent of modern techniques devoted to the structural elucidation of biomolecules have permitted to gain detailed insight into the mechanisms governing the phenomenon of

molecular recognition, providing a thorough understanding of the physicochemical bases underlying the receptor-ligand interaction.

1.2.1 The Nature of the Carbohydrate-Protein Interaction

In general, glycans bind at domains defined by shallow pockets on the hydrophilic surface of lectins and other receptors. As a consequence, affinity regimes of monovalent interactions between lectins and their ligands tend to be rather weak. Overall, sugar binding to proteins is made possible via a number of attractive forces including polar and hydrophobic interactions. Despite the phylogenetic and structural diversity of carbohydrate-binding proteins, the signature of the protein-carbohydrate interaction can be seen in different protein families and all throughout the evolutionary tree.

1.2.1.1 Hydrogen bonding

The establishment of hydrogen bonds between sugars and their receptors constitutes perhaps the most evident protein-carbohydrate bonding potential due to the existence of numerous –OH groups in all saccharides (fig. 5A). Furthermore, hydrogen bonding is made possible not only thanks to the hydroxyl groups, but also to the presence of amine and carboxyl groups of many substituted carbohydrates. Hydroxyl groups may participate in hydrogen bonds both as donors and as acceptors, by means of their lone electron pairs. In some instances, the same –OH group simultaneously acts as donor and acceptor, a characteristic phenomenon in sugar-protein interactions, known as cooperative hydrogen bonding. Sugar hydroxyl groups establish hydrogen bond contacts with the side chains of polar residues, most often aspartic and glutamic acid, asparagine, glutamine, arginine and serine, as well as backbone amine and carbonyl groups. A characteristic feature of sugar-protein interactions commonly seen in lectins and enzymes are bidentate hydrogen bonds established between two adjacent hydroxyl groups of a sugar and both carboxylate oxygens of either aspartic or glutamic acids [26]. Of course, hydrogen bonding not only contributes to affinity but also to selectivity, being, in many cases, the characteristic stereochemical arrangement of hydroxyl groups what gives a protein its specificity towards a given sugar type. Our group has studied the implications of hydrogen bonding on selectivity by approaching the phenomenon from both perspectives, i.e. those of the proteins and of the sugars, by studying the effect of mutation of key residues of the

receptors [27], and using engineered sugars 'deleterious' for hydroxyl groups via the synthesis of monodeoxy [28] and fluoro-monodeoxy derivatives [29-30] of natural lectin ligands.

1.2.1.2 Apolar interactions

In some sugars, the clustering of three or more adjacent C–H groups caused by the characteristic steric disposition of hydroxyl groups creates hydrophobic patches on the sugar surface that can establish apolar interactions with hydrophobic epitopes in proteins, most notably the aromatic rings of Trp, Tyr and Phe residues (fig. 5B). Although the fine details of the nature of such interactions are currently subject to investigation, it is thought that attractive forces are due to both an entropic contribution, arising from the mutual shielding of both apolar surfaces from the bulk water, and the enthalpic contribution of non-conventional hydrogen bonds established between the partially positively charged C–H groups and the quadrupole created by the π-system of the aromatic ring [31]. Our group has contributed to the understanding of CH-π interactions using a multidisciplinary approach combining NMR spectroscopy and molecular mechanics calculations. Using simple models involving monosaccharides and small aromatic compounds, such as free aromatic amino acids, it was possible to detect and quantify the magnitude of their interaction in water by simple NMR experiments [32]. In addition, experimental evidence for the hydrophobic component to the interactions was provided, finding that they were absent in aprotic solvents [33]. Theoretical results obtained in our laboratory also indicated that the carbohydrate-aromatic interaction is enthalpically stabilized by weak electro-attractive forces between the sugar hydrogens and the aromatic ring [34], a hypothesis that was confirmed experimentally [35].

1.2.1.3 Other interactions

Further forces involved in the recognition of some carbohydrates by their protein receptors include electrostatic interactions between charged residues of some saccharides, such as sialic acid, and protein residues of opposite charge [36] (fig. 5C). Another type of interaction involves the coordination of a divalent cation bridging certain sugar hydroxyls and negatively charged aspartates or glutamates. Such interaction is displayed by proteins belonging to the C-type family of lectins, which require the presence of Ca^{2+} ions to bind their carbohydrate ligands [21] (fig. 5D).

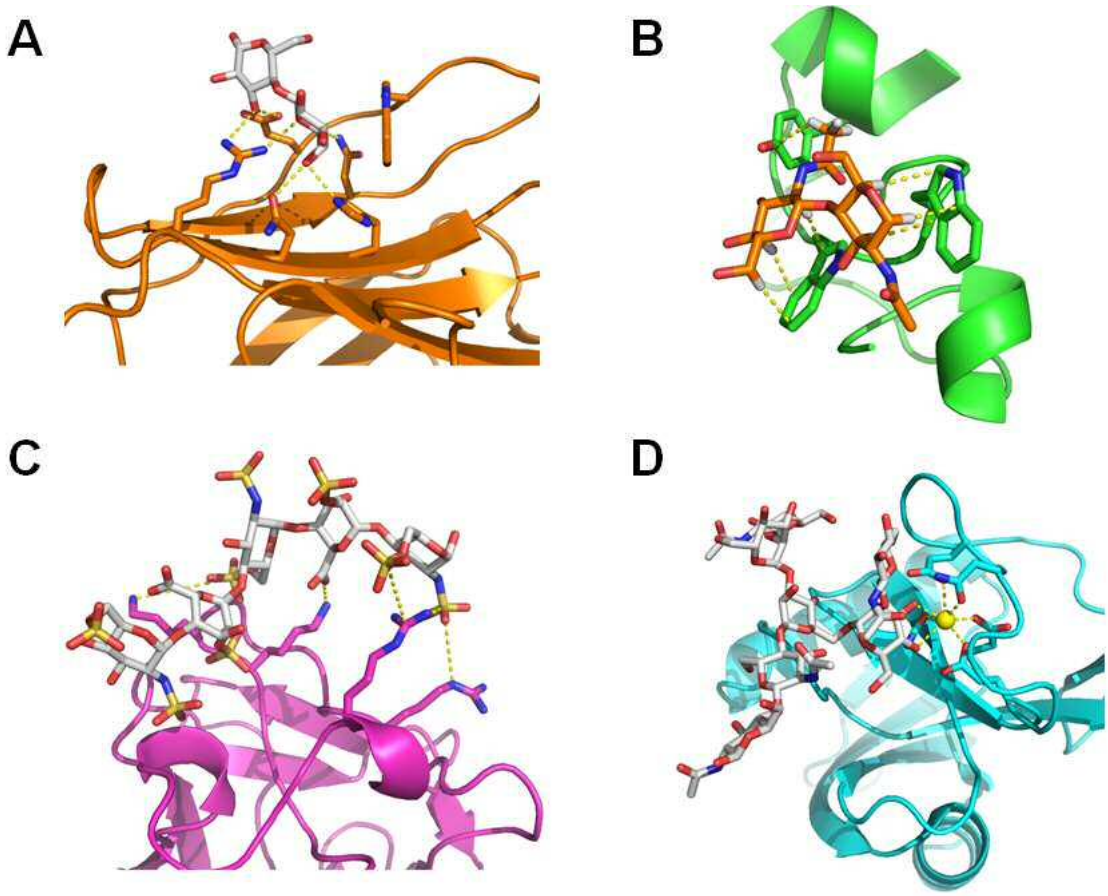


Figure 5: Different types of protein-carbohydrate interactions (shown in yellow-dashed lines). A: Hydrogen bonds. Crystallographic structure of lactose-bound human galectin-7 (PDB: 4GAL) [37]. B: CH-π interactions. NMR structure of hevein in complex with chitobiose (PDB: 1T0W) [38]. C: Electrostatic interactions. Crystallographic structure of the fibroblast growth factor in complex with a heparin pentasaccharide (PDB: 2AXM) [36]. D: Ca^{2+}-mediated. Crystallographic structure of codakine, a C-type lectin, in complex with a biantennary complex-type *N*-glycan (PDB: 2VUZ) [39].

1.2.2 Lectins and other carbohydrate-receptor proteins

Carbohydrates interact with partners belonging to a number of protein families including lectins, antibodies, viruses and enzymes. The latter catalyze the synthesis of activated sugar donors [40], as well as the transfer of sugar residues from these donors to other carbohydrates or aglycans [41], and thus they are responsible for the genesis of the whole cellular glycan repertoire. For their part, anti-glycan antibodies are generated by the immune system and play a role in the defense against pathogens and cancer cells [42]. In contrast to antibodies, lectins

are not products of the immune system and they display structural diversity [43-44]. Furthermore, and unlike enzymes, they lack catalytic activity. Finally, viruses are non-living, infectious particles which specifically recognize surface protein and glycan determinants of their hosts through fibrous multi-domain receptors [45].

Etymologically from the Latin *legere* [8], meaning, among other things, 'to select', lectins recognize and bind selectively and specifically to carbohydrate epitopes of glycoproteins, glycolipids and free oligosaccharides without modifying them. First discovered in plants, lectins are now known in all domains of life.

Lectins act as central mediators of the transfer of information in biological systems through their interactions with glycans, both as free carbohydrates in solution and presented by glycolipids or glycoproteins [8, 44]. These interactions mediate a number of important functions, including host-pathogen recognition [46], protein trafficking [47], lymphocyte homing [48], fertilization and early development [49]. The key involvement of lectins in these processes offers the potential of exploiting them for therapeutic use, both as pharmacological targets [21], or as carriers to deliver drugs to their site of action [50].

Lectins are diverse proteins according to structure, specificity and function. Although no absolute classification of lectins has been made, it is useful, in some contexts, to classify lectins according to their origin, hence distinguishing mainly between animal and plant lectins.

1.2.2.1 Plant lectins

Many plant lectins are involved in defense against parasites or predators, being sometimes secreted in large quantities as highly toxic agents, such as ricin [51]. Thanks to the easy disposal and purification of legume lectins, many of which are available in a commercial basis, plant lectins find themselves among the best characterized systems.

Plant lectins are often found in the oligomeric state, e.g. as dimers or tetramers, which confers them the ability to precipitate multivalent carbohydrates and agglutinate cells. This makes of lectins a family of biologically relevant study systems, not only from the perspective of protein/sugar, but also of protein/protein interactions.

1.2.2.1.1 Hevein domains

Hevein domains are common to plant lectins of different origins. They are specific for β(1-4)GlcNAc-containing glycans, notably chitin, a polysaccharide found in fungal cell walls and in the exoskeleton of insects and other invertebrate parasites. Lectins containing hevein domains are secreted by many plant species as a defense mechanism.

The hevein domain is a structural motif containing 40-50 amino acids, characterized by the presence of 6 or 8 cysteine residues, at conserved positions, establishing three or four intramolecular disulfide bonds that highly stabilize the structure. Hevein domains are found in many legume lectins, such as hevein itself, a small protein of 43 amino acids isolated from the rubber tree (*Hevea brasiliensis*).

The main driving force of the recognition of chito-oligosaccharides by hevein domains are CH-π interactions (fig. 5B). Thanks to the very convenient properties of hevein, which are related to its small size and high stability, our group has profusely employed this lectin as a model system to study protein-carbohydrate interactions in solution [31], with a focus on the nature of the CH-π interaction. Using NMR spectroscopy, our group has obtained detailed information on affinity and binding geometry for several hevein/$(GlcNAc)_n$ complexes, and provided the first structure of a lectin/ligand complex solved by NMR [52].

Other proteins bearing hevein domains include the stinging nettle lectin (*Urtica dioica* agglutinin, UDA), which is composed of two hevein domains [53], and the wheat germ agglutinin (WGA), a dimeric protein in which each subunit is an assembly of four hevein domains [54].

1.2.2.2 Animal lectins. Galectins

Animal lectins are expressed in a variety of tissues, where they perform disparate functions. They are generally divided into two main groups, according to their function and cellular location. The first one comprises intracellular lectins, such as the calnexin family and M-, L- and P-type lectins. Acting in different compartments along the cellular secretory pathway, they are mainly involved in glycoprotein sorting, trafficking and quality control [55-57] The second group includes extracellular lectins, such as C- and R-type lectins, siglecs and galectins, which may be membrane-bound or secreted to the extracellular medium, playing important roles in cell adhesion, cell signaling and recognition of pathogens [58].

Galectins [59] are a family of animal lectins defined by a shared consensus of amino acid sequences and a carbohydrate recognition domain (CRD) of around 130 amino acids, with affinity for β-galactose (Gal) containing oligosaccharides. The more than 30 solved 3D structures of many galectin family members show a highly-conserved β-sandwich fold (fig. 6), slightly bent forming a concave groove on the side opposite to the sugar-binding site.

Although, in some instances, galectins have been recognized to have intracellular functions, they are best known for their extracellular activities, being found on the cell surface or deposited in the extracellular matrix. There, galectins specifically recognize glycan moieties of glycolipids and glycoproteins on the surface of various cell types, triggering a number of cell responses both in developmental stages and in the adult tissue. In general, galectins have been associated with cell-cell and/or cell-matrix adhesion and migration, signal transduction, cell growth and proliferation and apoptosis [60]. Up to now, 15 members of the galectin family have been identified in mammals (Table I), although this number might increase [61]. Galectins are also known to other animal groups, such as birds [62] and invertebrates [63-64].

Galectins are classified in three groups according to their structure (fig. 7). Prototype galectins are composed of a single CRD, which can be monomeric or, more often, self-associate to form homodimers (gal-1, -2, -5, -7, -10, -11, -13, -14 and -15). Tandem-repeat galectins consist of two distinct CRDs connected by a linker peptide domain (gal-4, -6, -8, -9 and -12). Finally, chimera-type galectins are composed of one CRD attached to a non-lectin part. At present, the only chimera-type galectin identified is gal-3, which consists of a CRD and an N-terminal domain (ND), including several collagen-type repeats rich in glycine and proline, and an N-terminal peptide region, which can be phosphorylated [65]. The ND of gal-3 is involved in pentamerization of the lectin [66]. Of note, *in vitro* and *in vivo* enzymic cleavage of full length gal-3 by extracellular metalloproteases yields a truncated, monomeric version of gal-3 consisting of only its CRD [66].

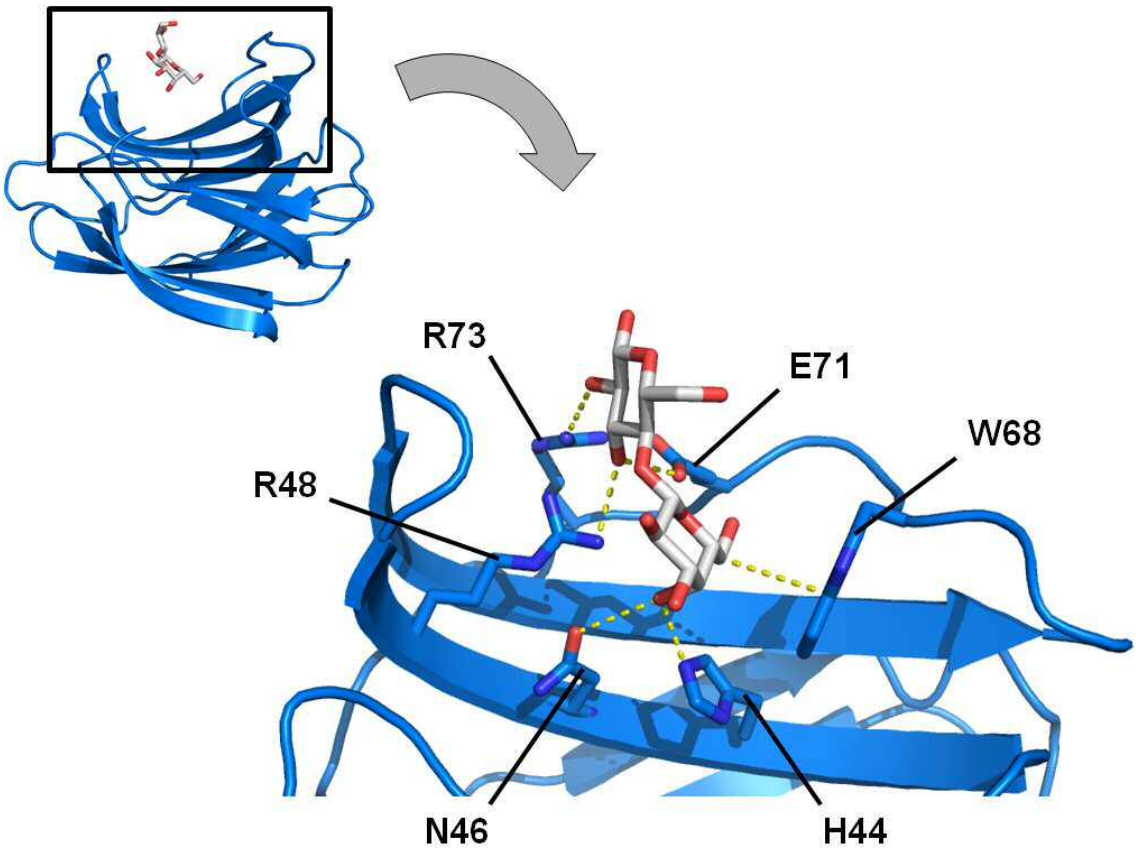


Figure 6: Galectin fold and interactions. Crystallographic structure of a lactose-bound human gal-1 (PDB: 1GZW) [67] monomer, showing the characteristic galectin β-sandwich fold. The expanded view shows the network of contacts with lactose, including hydrogen bonds between sugar hydroxyl groups and the side-chains of conserved residues H44, N46, R48, E71 and R73, and a CH-π interaction established between W68 and the apolar face of the galactose reside.

1.2.2.2.1 Recognition features of galectins

The ability of galectins to bind β-galactoside-containing glycans has been well documented both *in vitro* and *in vivo* [68]. Galectin family members do not appear to have specific individual receptors, but rather they can bind to a set of cell-surface or extracellular matrix glycoproteins containing suitable oligosaccharides [68]. The minimal saccharide unit recognized by galectins is the galactose residue, linked to another monosaccharide in the β-configuration. Although many structural aspects of the galectin/galactoside interaction have been explored by using simple sugars down to the level of disaccharides (see, for example, [69-70]), galectins perform their biologically relevant functions by binding to β-galactosides incorporated in larger, more complex ligands, such as *N*-acetyl lactosamine (Gal-β(1-4)GlcNAc) residues found in *N*-linked "complex type" glycans and *O*-linked glycans on glycoproteins [71].

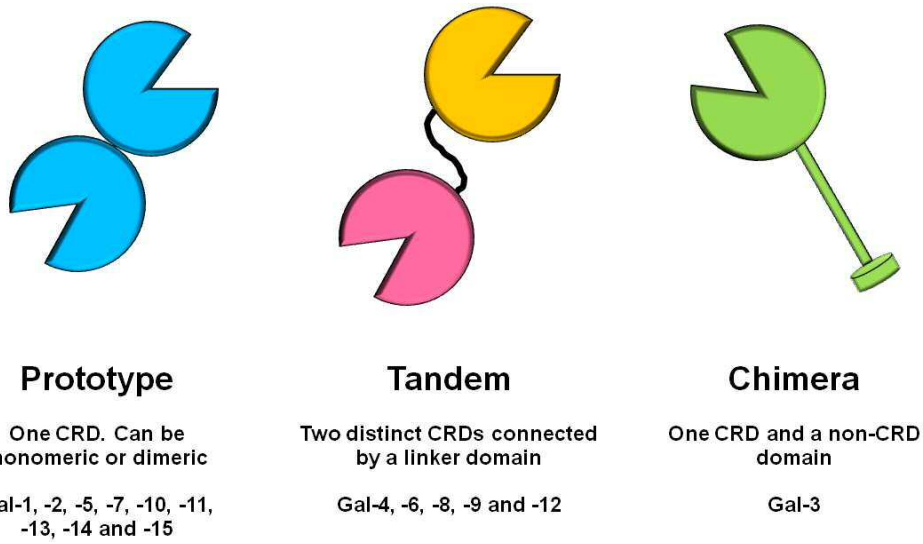


Figure 7: Structural types of galectins: prototype, tandem-repeat and chimeric, with examples.

Galectin	Type	Location	Function
Gal-1	Prototype	Ubiquitous	Induction of apoptosis in activated T cells, cell growth, mRNA splicing, regeneration of nerve axon [72].
Gal-2	Prototype	Gastrointestinal tract	Induction of apoptosis in T cells.
Gal-3	Chimera	Ubiquitous	Pro- or anti-apoptotic, cell adhesion, mRNA splicing, macrophage chemotactic factor, regulation of genes including JNK1 [73].
Gal-4	Tandem	Gastrointestinal tract	Activation of intestinal CD4+ cells
Gal-7	Prototype	Keratinocytes (stratified epithelium)	Pro-apoptotic effector, differentiation of keratinocytes, wound healing [74].

Gal-8	Tandem	Ubiquitous	Cell adhesion through binding to integrins, regulation of neutrophil function.
Gal-9	Tandem	Immune cells, lung, kidney gastrointestinal tract	Induction of apoptosis in activated T cells and cancer cells, eosinophil chemotaxis, urate transporter in kidney [75].
Gal-10	Prototype	Eosinophils, basophils	Suppression of T cell proliferation.
Gal-12	Tandem	Adipose tissue	Induction of apoptosis in adipocytes.
Gal-13	Prototype	Placenta	Lysophospholypase activity [76].

Table I: Structure, location and function of the principal human galectins.

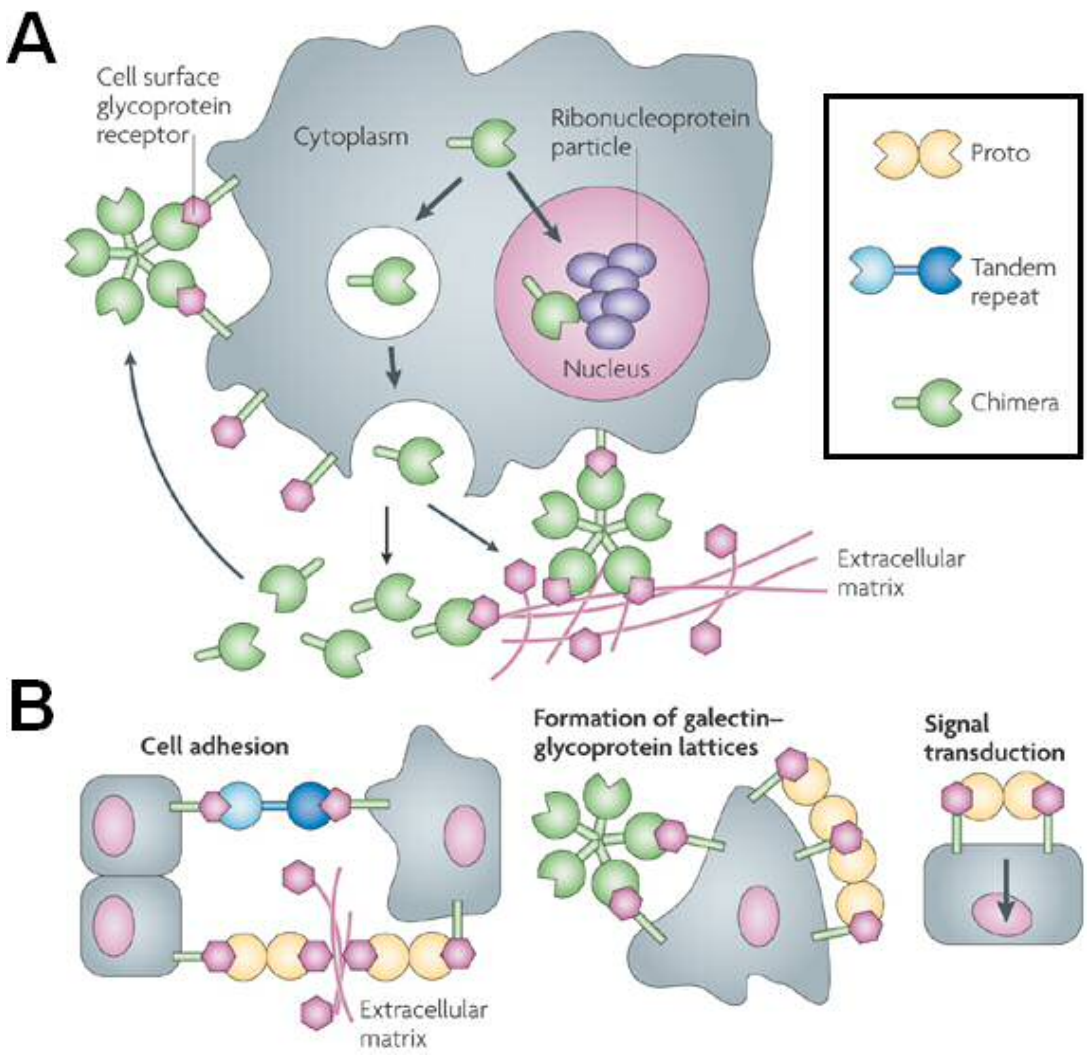


Figure 8: Intra- and extra-cellular functions of galectins. A: Subcellular localization and secretion. From the cytosol, galectins can be translocated into the nucleus or vesicles, or stay in the cytosol. In addition, and although galectins lack a canonical secretion signal peptide, they can be exported across the plasma membrane by non-classical mechanisms through vesicle transport. Once in the extracellular space, they can bind to cell surface glycans and cross-link them to glycans of the extracellular matrix. B: Extracellular roles of galectins. By cross-linking glycans on neighboring cells, either directly or via the extracellular matrix, galectins can mediate cell adhesion. Also, by binding to cell surface glycans in *cis*, they can form homogenous lattices on cell surfaces, which can trigger the activation of signaling pathways of functional relevance. Adapted from [60].

As sugar receptors, the affinity of galectins toward simple saccharides is rather low. For example, the K_D of gal-3 for lactose, the pan-galectin ligand, is around 170 µM [77]. Binding affinities can be modulated by substitutions at the β-galactoside core, resulting from subtle, but significant differences in the architecture of the sugar binding site of each galectin member. For example, *N*-acetylation of galactose significantly reduces the affinity of gal-1 for this residue, whereas it increases that of gal-3 [78]. In addition, α(2-6)-sialylation notably

decreases affinity for gal-1 and gal-3. This modification has been shown to be of functional relevance, being crucial for the gal-1-dependent skewing of T cells toward a T-helper cytokine profile, both in mice and in humans [79]. For its part, α(2-3)-sialylation greatly boosts binding affinity to some galectins, notably the tandem-repeat type gal-8 [80], down to the nM range. The main driving force responsible of this strong interaction, hardly found in other monovalent lectin-sugar systems, is an electrostatic contact with the Arg59 side-chain of the gal-8 N-CRD. In contrast, the gal-8 C-CRD does not interact with α(2-3)-sialylated ligands, showing preference toward LacNac-containing oligomers instead [81]. Thus, gal-8 is an example of how tandem-repeat galectins may display different recognition features at each individual sugar-binding site.

In any case, monovalent interactions of galectin CRDs with their glycan ligands are lower than typical protein-protein interactions [78]. In order to increase their avidity toward their ligands, most galectins, as other lectins, self-associate to form higher-order, multivalent glycoclusters. For example, truncated gal-3, which lacks the ND needed for oligomerization, also lacks the ability to tightly bind to the glycocalix of endothelial cells, in contrast to the full-length protein [82]. Similarly, dimeric gal-1 binds with considerably higher affinity to extended glycans containing terminal *N*-acetyllactosamine (LacNAc) sequences than a monomeric gal-1 mutant [83]. Thus, higher avidities of soluble galectins are achieved by oligomerization of monovalent galectin monomers, which is essential for these proteins to perform their functions.

1.2.2.2.2 Galectins in protein-protein interactions

A number of studies have shown that some galectins can play a role inside the cell, participating in protein-protein interactions in a manner independent of their sugar-binding capabilities. For instance, gal-1 and -3 are known pre-mRNA-splicing factors, and gal-3, -7 and -12 have been shown to regulate cell-growth and apoptosis intracellularly [84]. In some cases, the protein interacting partners of galectins mediating these intracellular processes are known. For instance, physical interaction has been demonstrated between gal-7 and at least two non-glycosylated partners: the apoptotic regulator Bcl-2 [85] and the transcription factor Smad2/3 [86]. Of note, the gal-7/Bcl-2 interaction was shown to be independent of carbohydrate binding. Furthermore, non-glycosylated protein partners are also known to other galectins, such as Bcl-2 itself for gal-3 [87] and farnesyl-Ras for gal-1 [88]. In the latter case, the binding epitope on gal-1 was delineated by the construction of a mutant (gal-1 L11A) which retained normal carbohydrate-binding abilities, but inhibited Ras-GTP, giving rise to an attenuated pro-

oncogenic activity of H-Ras(G12V) [89]. Taken together, all these data evidence the interacting capabilities of galectins beyond the carbohydrate recognition site.

1.2.2.2.3 Role of galectins in the immune response

Galectins are widely expressed in mammalian tissues, including most cells responsible for both the innate (dendritic cells, macrophages, mast cells, natural killer cells and B1 cells) and adaptive (activated B and T cells) immune responses [60].

Galectins are involved in the adaptive response through regulation of immune cell development and homeostasis at different levels. For example, galectins can modulate B cell maturation and differentiation [90]. Galectins are also bivalent modulators of apoptosis. Gal-1 is pro-apoptotic on T cells depending on the developmental stage and activation status of the cell, as well as on the microenvironment [79]. On the other hand, the effect of gal-3 on T cell survival varies depending on its location: it is anti-apoptotic when intracellular, and pro-apoptotic when outside the cell [91]. Also, galectins are known to exert important regulatory functions in T cell homeostasis. Signaling cascades triggered by their binding and lattice formation at the surface of T cells have implications in a range of downstream events modulating their differentiation, activation, and production of pro- and anti-inflammatory cytokines [68, 92].

Recently, galectins have been discovered to also play a role in innate immunity, by binding glycans at the surface of different types of pathogens [60, 93]. For instance, gal-1 is able to inhibit the fusion of the Nipah virus with endothelial cells by binding to *N*-linked oligosaccharides of the virion envelope [94]. Also, gal-3 has been reported to be a receptor for bacterial LPS in a species-dependent manner, with at least two independent binding sites for LPS [95]: its CRD binds to β-galactoside-containing groups in the glycan domain of LPS, and its ND binds the lipid A moiety of LPS, which is essential for LPS-mediated neutrophil activation [96-97]. GalNAc-β(1-4)GlcNAc (LacdiNAc)-containing glycans constitute a parasite pattern for gal-3-mediated immune recognition of various invertebrates, such as helminths [98]. Also, gal-3 can bind to Gal-β(1-3)-Gal, a *Leishmania* virulence factor [99-100], the Gal-α(1-3)-Gal-β(1-4)-GlcNAc-R xenoantigen [101-102], and A/B-histo-blood group tetrasaccharides [103]. In addition, the recent demonstration that galectins 4 and 8 are able to kill *E. coli* bacteria expressing blood group B antigen [104] has reinforced the idea that innate immunity functions are widespread among the galectin family.

1.3 Nuclear Magnetic Resonance Spectroscopy (NMR)

Over the last years, numerous investigations have looked deeply into the fine structural details of carbohydrates and their recognition by receptors. Advances in different biophysical and spectroscopic techniques, such as X-ray crystallography and Nuclear Magnetic Resonance spectroscopy (NMR), complemented by quantitative data obtained through isothermal titration microcalorimetry (ITC), [105] have allowed for the gathering of information on protein-carbohydrate complexes [106] at atomic resolution, and permitted to deduce the role of the different forces involved in their interactions.

Among these, NMR is one of the most widely used techniques to characterize molecular recognition events, [107] thanks to its versatility, and a spectacular progress that it has experienced in recent times, with advances in distinct areas to generate an improvement in the resolution and sensitivity of NMR experiments. Also, the developments of procedures for structural determination, as well as the study of the conformation and dynamics of complex oligosaccharides have permitted key advances in the expansion of NMR [108]. Also of note, NMR techniques provide the possibility to work in solution, i.e. in conditions akin to those physiological.

Nuclear magnetic resonance [109-110] is an intrinsic property of atomic nuclei having a non-zero spin quantum number, *I*. Such nuclei possess an associated magnetic moment, μ via the relation $\mu = \gamma I$, where γ is the gyromagnetic ratio, which is characteristic for each nucleus. The angular momentum associated with nuclear spin is quantized, e.g. both the magnitude and orientation of the angular momentum can only adopt a limited number of values. For a given nucleus with spin *I*, there exist 2*I*+1 possible angular momentum states, taking values from -*I* to +*I*. Thus, for nuclei with spin *I* = ½, such as ^{1}H, ^{13}C or ^{15}N, there are two possible sublevels, or spin states: *m* = +½ and *m* = - ½. In the absence of an external magnetic field, these levels are degenerate, which means that both states have the same energy, and the number of nuclei in both states will be approximately equal. However, if an external magnetic field (B_0) is applied, the degeneracy is broken, being the difference between both states as shown in eq. 1. As a result, a slightly larger fraction of the nuclei is present in the lower energy spin state, following the Boltzmann distribution (eq. 2).

$$\Delta E = \gamma \hbar B_0 \qquad (1) \qquad \frac{N_{low}}{N_{upp}} = e^{\frac{-\Delta E}{k \cdot T}} \qquad (2)$$

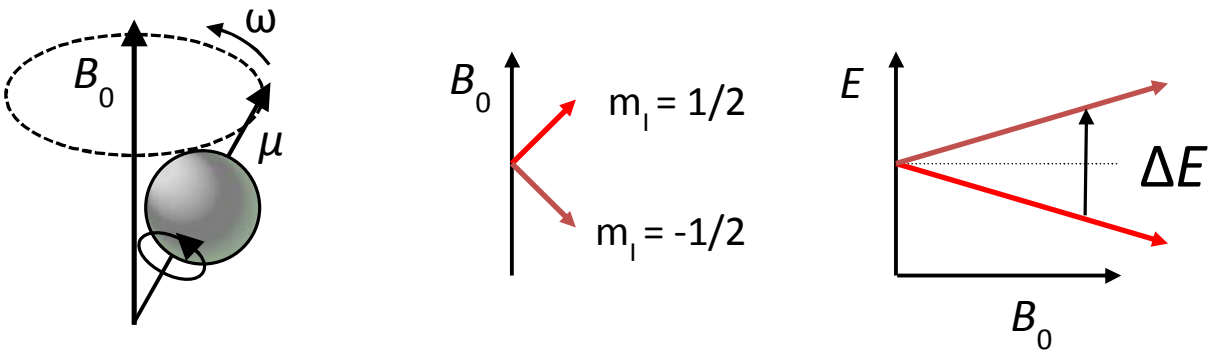


Figure 9: Nuclear magnetic resonance. Left: precession of a m = ½ nucleus with an associated magnetic moment μ around the external magnetic field B_0. Center: two possible momentum states of a nucleus with spin I = ½ under an external magnetic field B_0. Right: energy difference (ΔE) between both momentum states and its dependence on the external magnetic field strength.

where k is the Boltzmann's constant. The signal in NMR spectroscopy results from the difference between the energy absorbed by the spins which make a transition from the lower energy state to the higher energy state, and the energy emitted by the spins which simultaneously make a transition from the higher energy state to the lower energy state. The NMR signal, therefore, is proportional to the population difference between both states. Noteworthy is that, given the small energy difference between both spin states even at very high magnetic fields, the population bias towards the lower energy state is always tiny, making NMR a rather insensitive spectroscopy.

The magnetic moment of each nucleus precesses around B_0 (fig. 9). The frequency of this precession is called the Larmor frequency (ω_0), which is the frequency of resonance for each nucleus, and is equivalent to the energy difference between the two levels:

$$\omega_0 = \gamma B_0 \Rightarrow \omega_0 = \frac{\Delta E}{\hbar} \qquad (3)$$

The Larmor frequency for each nucleus depends thus on the gyromagnetic ratio and the strength of the magnetic field, as shown in eq. 3. However, there exist local perturbations of the magnetic field experienced by each nucleus, mainly due to the shielding exerted by nearby

electrons. As a result, nuclei within different chemical environments precess at slightly different frequencies (eq. 4), which permits the use of NMR as a spectroscopy technique.

$$\omega_0 = \gamma(B_0 - B_i) \tag{4}$$

For practical reasons, NMR spectroscopists prefer the use of chemical shifts (δ, expressed in ppm) instead of Larmor frequencies. The relation between both is shown in eq. 5.

$$\delta = \left(\frac{\omega - \omega_{ref}}{\omega_{ref}} \right) 10^6 \tag{5},$$

where ω is the Larmor frequency for a given nucleus, and ω_{ref}, is the Larmor frequency of a nucleus in a reference compound, usually tetramethylsilane (TMS) or trimethylsilylpropionate (TSP) in proton NMR.

1.3.1 NMR spectroscopy of carbohydrates

For decades, NMR spectroscopy has been the preferred biophysical technique for structural characterization of carbohydrates. Reasons for this include the intrinsic flexibility of glycans, which makes them difficult to crystallize, hampering its study by X-ray diffraction, and the possibility to work in solution. In addition, NMR spectroscopy offers a great variety of methods, which permit the gathering of very useful information regarding the fine structure of carbohydrates, their conformational and dynamic properties, and their recognition features by receptors.

Among several experimental restraints that can be gathered for structural determination of carbohydrates, vicinal ^{1}H-^{1}H coupling constants ($^3J_{HH}$) are particularly useful, as they permit to determine the relative configuration of the sugar nuclei. Additionally, nuclear Overhauser enhancement (NOE) spectroscopy (NOESY) is employed for conformational analyses around the glycosidic linkage [107, 111].

1.3.1.1 Coupling constants

In the late 1950's, Nobel Laureate Martin Karplus described a relationship between the dihedral torsion angle established between vicinal hydrogens and the $^3J_{HH}$ coupling constant, by using a theoretical approach [112]:

$$^3J_{HH'} = A + B\cos\Phi + C\cos^2\Phi \quad (6)$$

Where Φ is the dihedral angle between the two protons. According to eq. 6, for dihedral angles of 0º or 180º the observed coupling constant is large; while, conversely, dihedral angles around 90°, lead to $\cos^2\Phi$ values close to zero, and thus the observed coupling constant is small or null. For its part, coefficients *A*, *B* and *C* are parameterized for each particular kind of molecules, atoms and substituents. In this regard, and for instance, eight new Karplus relationships have recently been refined for use in conformational studies of saccharide *N*-acetyl side-chains [113].

A practical example of the usefulness of the Karplus equation to determine the relative configuration of the chiral centers is provided by the differentiation between two monosaccharides, such as β-glucose and β-mannose, which are epimers at C2. The dihedral angle between H_1 and H_2 is around 180º in β-glucose and 60º in β-mannose. Subsequently, the $^3J_{HH}$ value is large in glucose, of around 10-15 Hz, and small in mannose, of around 3-5 Hz.

Also, the absolute configuration of the anomeric function in glycosides can be obtained with the same approach: in α-glucose, the $^3JH_1H_2$ is small due to the small angle established between these two protons, while their *trans*-type di-axial disposition in β-glucose entails a larger *J*-coupling value.

1.3.1.2 The Nuclear Overhauser Effect (NOE). NOE spectroscopy (NOESY)

The 3D molecular geometry of glycans becomes more intrincated as the number of saccharide units increases. Knowledge on the spatial relationships among nuclei belonging to different units can be provided by both nuclear Overhauser enhancement (NOE) spectroscopy (NOESY) studies. This source of structural information is based on its dependency on the interproton distance through space.

The NOE is defined as the change in intensity of a nucleus' signal when the spin transitions of another nucleus cause a perturbation of its equilibrium populations. The two nuclei do not share a scalar (through bond) coupling; instead, they are sufficiently close in space to share a dipolar coupling. Thus, the NOE originates from dipolar cross-relaxation between pairs of protons, and depends on the proton-proton distance and on the molecular motion of the interproton vector:

$$I_{\mathrm{NOE}} \approx <1/\mathrm{r}^6> f(\tau_c) \qquad (7)$$

where I_{NOE} is the NOE intensity, r is the proton-proton distance, and f is a function that depends, among others factors, on the correlation time (τ_c) that describes the motion of the interproton vector. The inverse-sixth relationship in eq. 7 implies that the NOE only develops between nuclei that are close in space, usually within 5 Å. In 2002, Kurt Wüthrich was awarded the Nobel Prize in Chemistry for demonstrating that two-dimensional NOE spectroscopy could be used to determine the 3D structures of biological macromolecules in solution [114-115]. In 2D NOESY experiments, a map of the NOE relationships of every hydrogen nucleus is obtained, thus allowing for a thorough structural characterization of the molecule.

Of course, not only homonuclear NOESY, but also heteronuclear studies are possible. For instance, ^{19}F-^{1}H heteronuclear NOE spectroscopy (HOESY) may be employed to estimate fluorine-proton distances [116].

1.3.1.3 Spectral assignment of carbohydrates

Among all biophysical techniques, NMR is unique in that it permits the monitoring of individual nuclei, allowing for complete structural elucidation of biomolecules, as well as epitope-fine mapping of their interactions at atomic resolution. To initiate these studies, the first step is thus to obtain the complete or partial spectral assignment of the molecules of interest. A palette of NMR experiments can provide such information, usually when used in conjunction.

Correlation spectroscopy (COSY) is a 2D homonuclear experiment used to identify spins which are directly *J*-coupled to each other, thus allowing for the identification of neighboring protons. Total correlation spectroscopy (TOCSY) experiments are conceptually similar to COSY-type methods. However, they allow for the identification of nuclei that are not only directly

coupled, but also indirectly coupled through a chain of *J*-couplings, i.e. belonging to the same spin system. Thus, in carbohydrates, TOCSY experiments allow for the identification of spins belonging to the same sugar residue. For their part, NOESY experiments allow for the establishment of inter-residue connectivities.

Finally, heteronuclear experiments are useful for achieving the splitting of signals that are overlapping in the ^{1}H dimension, something not unusual in carbohydrates due to their poor spectral dispersion. The most commonly used is ^{1}H-^{13}C heteronuclear single-quantum coherence (HSQC), which provides a 2D mapping of pairs of directly bound C-H nuclei.

NMR spectral elucidation can often be aided by assignments reported in the literature or stored in databases (notably, the Biological Magnetic Resonance Data Bank BMRB [117]) or computationally simulated, e.g. with CASPER [118].

1.3.2 NMR methods for the study of protein-carbohydrate interactions

In addition to those suitable for the study of isolated carbohydrate molecules, a variety of NMR techniques are also available for studying protein-ligand interactions. These methods, usually assisted by computational protocols, permit to elucidate the structural features of the binding entities, as well as to characterize the kinetic and thermodynamic parameters of the free-bound equilibrium.

NMR techniques devoted to the study of molecular recognition processes encompass both ligand- and receptor-detected experiments. Receptor-based methods are based on the monitoring and comparison of the receptor signals in the presence and absence of compounds of interest. By identifying perturbations of specific protein resonances, their binding sites can be delineated. These methods demand physicochemical properties of the protein target that are often challenging. First, they are usually applicable to relatively small receptors (< 40 kDa). Also, milligram quantities of soluble, non-aggregated protein must be expressed and purified. Thus, suitable expression hosts must be found that permit isotope enrichment (e.g. ^{13}C, ^{15}N, and sometimes also ^{2}H) needed for resonance assignment of the heteronuclear NMR spectrum of the macromolecule. In addition, it must be ensured that samples are stable for the time required for sequential resonance assignment (weeks).

For their part, ligand-detected methods are based on the monitoring of the NMR parameters of ligands in the presence and absence of the receptor. Saturation transfer difference (STD),

WaterLOGSY and transferred NOE spectroscopy (trNOESY) are used for detecting binding based on either changes in the motional properties of the saccharides upon binding to large macromolecular receptors, or the transfer of magnetization via their binding to the receptors. In contrast to protein-detected methods, these approaches render the size of the protein not an issue. In fact, most ligand-based experiments become more sensitive when working with larger and multiple site receptors. Additionally, ligand observation bypasses the need to produce milligram quantities of isotopically labeled protein.

An obvious setback of ligand-based approaches is the impossibility to identify their binding site on the receptor surface. Also, ligand-based approaches rely on the exchange-mediated transfer of bound state information to the free state. This requirement biases ligand-detected methods toward the study of weakly binding ligands (i.e. in fast exchange) and the use of large ligand molar excesses.

In short, both receptor- and ligand-based approaches have distinct advantages and disadvantages, but they can be used together to provide a thorough vision on all the structural aspects that govern the sugar-protein interaction.

1.3.2.1 Ligand-detected methods

With certainty, the most powerful ligand-detected NMR techniques for probing binding activities are those based on the transfer of magnetization, such as saturation transfer difference (STD) and WaterLOGSY. The common mechanism underlying both techniques is the detection of dipole-dipole interactions between the spins of the ligands and those of the receptor, being their main difference the target of the irradiation: in STD experiments, the magnetization is transferred from the irradiated protein to the binding ligands, while in the WaterLOGSY experiment, the bulk water is selectively excited, and the magnetization is transferred from the water molecules lying at the receptor-ligand interface to the ligand nuclei.

1.3.2.1.1 Saturation transfer difference (STD)

The STD experiment [119] involves the comparison of two ^{1}H-NMR spectra of the same sample, acquired either with (on-resonance spectrum) or without (off-resonance spectrum)

saturation of the protein. In the on-resonance experiment, a subset of the protein protons are selectively irradiated by a train of pulses directed towards frequencies of the ^{1}H-NMR spectrum devoid of ligand signals. Typically, the irradiation frequency is set between -1 and 1 ppm, thus being side-chain protons of Ile, Val and Leu residues the target of irradiation. Alternatively, if the ligand lacks signals in the aromatic region, irradiation can be set at 6-8 ppm. The magnetization then quickly propagates, via ^{1}H-^{1}H cross-relaxation pathways, from the irradiated protons to the rest of the protein, and from there to any ligands in direct contact to the protein. This will lead to a spectrum where the intensity of the signals belonging to saturated molecules (both the receptor and binding ligands) is decreased, but leaving signals of non-binding molecules unaffected.

In parallel, an off-resonance spectrum, equivalent to a reference ^{1}H-NMR spectrum of the complex, is obtained by irradiating the sample at spectral regions devoid of either receptor and ligand resonances (e.g. 100 ppm). The difference between the reference and the on-resonance spectra yields a saturation spectrum, in which the signals of the saturated nuclei are represented, thus allowing the discrimination of ligands with binding activities in a compound mixture (fig. 10).

Typically, the sample is irradiated with a total saturation time of between 0.5 and 4 s. During this time, the ligand molecules can enter the binding site, become saturated, and leave it. If the ligand is in excess versus the receptor, this process can take place for many times for different individual ligand molecules, provided that their binding exchange kinetics is fast enough. Thanks to the slow longitudinal relaxation of the ligands, the saturated state largely persists when ligands dissociate, giving rise to a pool of unbound saturated ligands, which are ultimately responsible for the STD signal. On the basis of this, two considerations can be made. First, the STD experiment is most sensitive to ligands binding strongly enough to become effectively saturated, but weakly enough to be in fast exchange between the bound and the unbound states so that a large population of unbound, saturated molecules is generated. Put into numbers, the optimal range of dissociation constants is about 10^{-3}-10^{-8} M [120]. Second, the method enables a large excess of ligand to be used, typically 20:1 to 100:1. Thanks to the efficient saturation of proteins, the concentration of receptor can be kept as low as 1 µM in favorable cases, which means that concentrations between 50 µM and 200 µM of ligands can be used. Of course, this is also highly dependent on the water solubility of the screened compounds.

The small amounts of protein needed for the STD experiment, added to the simplicity of its read-out and the possibility to screen several compounds at a time, make of it the preferred NMR-based method for screening a large number of putative ligands. Moreover, in its more than 10 years of history, the STD has proven versatile and useful for the pharmaceutical industry far beyond its role as a screening technique, being today routinely utilized for the characterization of the ligand epitopes and determination of affinity constants.

The STD experiment was originally designed to screen a library of carbohydrate compounds against the wheat germ agglutinin (WGA) [119]. In the same study, however, it was realized that this technique was not only useful to detect binding events, but also to map the parts of the ligand making a major contribution to the binding. Indeed, in complex molecules, the ligand epitopes establishing tighter contacts with the protein receive a higher amount of magnetization, which is translated into stronger signals in the STD spectrum. This information facilitates the discrimination between parts of the molecule with major importance for binding (pharmacophore), and those with less relevance, which can be subject to chemical development in order to achieve better affinities or improved properties for the overall drug.

In addition, STD experiments allow to measure binding affinities, thus enabling the ranking of bioactive compounds. For this purpose, STD-titration curves are built in the presence of variable amounts of the ligand. Special care must be taken to keep the protein concentration constant, as it is an important factor affecting the intensity of the STD signal. Measuring of K_Ds by STD experiments must be executed carefully, as there are a number of factors which can introduce biases in the estimations, such as rebinding events and the longitudinal relaxation of the ligands. To overcome these sources of error, K_Ds are determined by STD titration curves at zero saturation time, since, at these conditions, neither rebinding nor relaxation occurs. Curves at zero saturation time are extrapolated from titration experiments performed at different saturation times. Affinity data estimated this way have proven to be in agreement with those obtained by other biophysical techniques, such as calorimetric studies [121].

STD experiments have been successfully applied to whole, living cells, enabling the study of ligand binding to membrane proteins, such as G-protein coupled receptors (GPCRs), which are often difficult to isolate and manipulate. A typical problem arising when STD experiments are performed on whole cells is the observation of crowded STD spectra due to the saturated cell components, which can preclude the identification and epitope mapping of the ligand signals. In such cases, a variant technique known as saturation transfer double difference (STDD) is applied. According to this method, the STD spectrum of a control performed in the presence of

the cells is subtracted from the STD spectrum performed in the presence of the cells and the ligand, leading to the exclusive observation of signals corresponding to the ligand STD effects in the STDD spectrum [122].

1.3.2.1.2 WaterLOGSY

In the second magnetization transfer experiment, the magnetization of the bulk water is selectively transferred, via the protein-ligand complex, to the ligand in the unbound state.

In essence, the WaterLOGSY (Water-Ligand Observed via Gradient SpectroscopY) experiment [123] can be understood as a NOESY experiment, in which the magnetization of the water molecules is selectively inverted, and during a long mixing time of up to several seconds, the magnetization is transferred, via ^{1}H-^{1}H cross-relaxation, to the ligand spins at the protein-ligand interface. There are essentially two pathways by which this transfer of the magnetization occurs: (i) via direct cross-relaxation from the water molecules tightly bound at the protein-ligand interface, and (ii) via chemical exchange between the water protons and the labile protons of amine and hydroxyl groups in the protein, which in turn cross-relax with the protons of the bound ligand. In either case, an inversion of the NOE sign takes place during the mixing time, since both the protein protons and those of the water molecules bound to the protein adopt the motional and NMR relaxation properties of the receptor, with a slow tumbling rate and therefore, a long correlation time. For their part, non-interacting ligands receive the magnetization only in the free state, via water molecules involved in their solvation sphere, and the sign of the NOE stays unaltered. As a result, the WaterLOGSY experiment yields a spectrum in which ligands with binding activity are manifested by positive signals, whereas non-binding ligands are represented by negative signals (fig. 11).

In practical terms, the same considerations regarding the use of a ligand excess and low protein requirements can be made with respect to STD experiments. The sensitivity and range of applicability of both methods is comparable, and the WaterLOGSY can also be used to obtain affinity data [124]. For K_D determination by WaterLOGSY, a parallel series of control experiments in the absence of the receptor must be performed. Since each signal in the WaterLOGSY experiment is a weighted combination of the negative NOEs of the hydrated unbound ligand and the positive NOEs of the ligand in the bound state, the WaterLOGSY signals must be normalized by subtracting the signals of the reference spectrum from those of the WaterLOGSY spectrum. Titration curves are then built by plotting the resulting intensities against the respective ligand concentration values.

In addition, it has been shown that WaterLOGSY experiments can also provide information on the ligand orientation in the bound state, given the different degrees of bulk water accessibility to parts of the ligand facing the protein (giving rise to positive NOEs) and those protruding from the protein (giving rise to negative NOEs) [125].

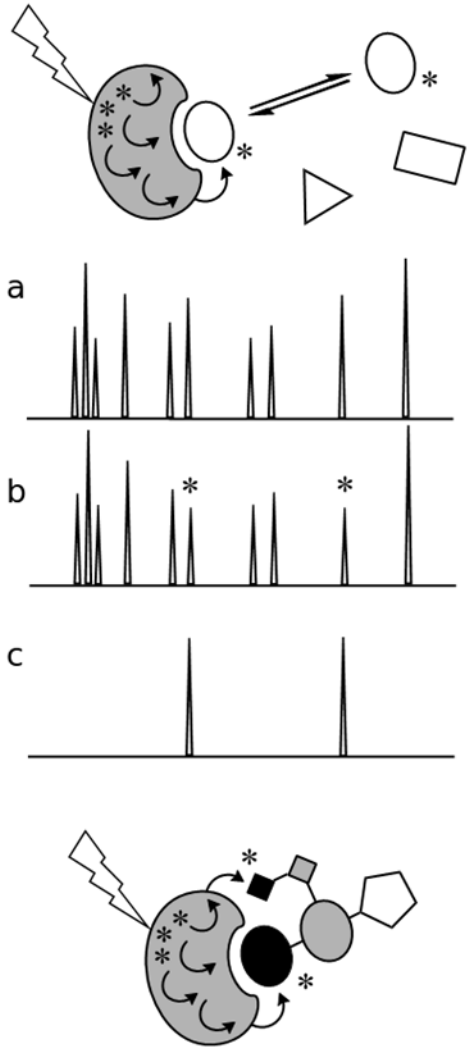


Figure 10: The STD experiment. Top panel: schematic representation of the STD experiment, with arrows representing the transfer of magnetization from the irradiated protons in the macromolecule to the ligand protons. Middle panel: Simulated STD experiment. (a) Reference (off-resonance) spectrum; (b) On-resonance spectrum; (c) Difference spectrum (up-scaled). Bottom panel: illustration of the STD epitope mapping effect

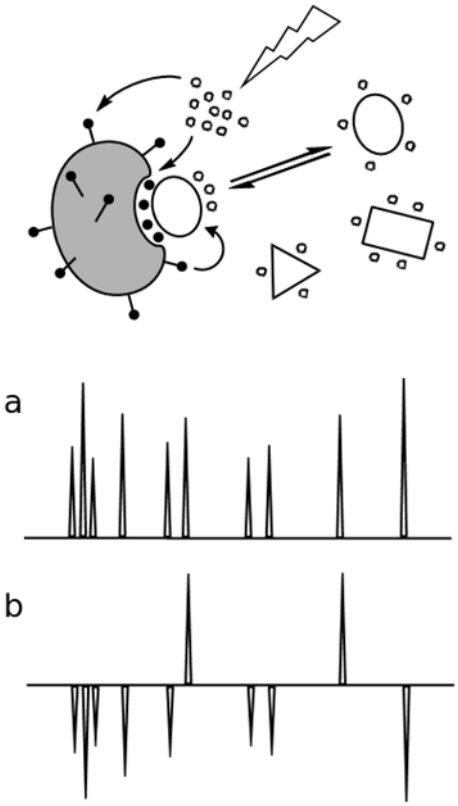


Figure 11: The WaterLOGSY experiment. Top panel: schematic representation of the WaterLOGSY experiment. White circles represent water molecules causing negative NOEs, and filled circles represent water molecules causing positive NOEs or labile protons in the protein exchanged from the irradiated water. Bottom panel: Simulated WaterLOGSY experiment. (a) Reference spectrum; (b) WaterLOGSY spectrum.

A practical drawback of the WaterLOGSY experiment is that samples must be prepared in protonated H_2O as a solvent, which requires the introduction of an efficient means to suppress the water signal. On the other hand, its lesser dependence on spin diffusion makes WaterLOGSY the preferred option when dealing with receptors of low proton densities, such as nucleic acids [126].

1.3.2.1.3 Transferred NOESY (trNOESY)

Since its introduction by Bothner-By in the early 1970s, trNOE spectroscopy [127-128] has been thoroughly used to study receptor-ligand interactions. This technique focuses on the change of sign in the ligand NOE peaks upon binding to the receptor. This is due to the dependence of the NOE sign on the correlation time of the molecule, which in turn is related with its size. As a result, fast moving small molecules show positive NOEs, while slow tumbling macromolecules exhibit negative NOEs.

The trNOESY experiment is a regular 2D NOESY experiment in which the cross-peaks of the ligand change their NOE sign by virtue of the fact that, in the bound state, they adopt the motional properties of the macromolecule. As a result, non-binders show no NOE sign change upon the presence of receptor and their NOEs remain positive, while binders show negative NOE cross-peaks. Of note, the observed NOE is a contribution of both the free (positive NOE) and the bound ligand (negative NOE), but the cross-relaxation rate of the bound ligand is much larger than that in the free state. Thus, the observed NOEs show a predominance of the contribution arisen by the bound form [129]. As a result, the trNOESY experiment not only permits to detect interactions, but also provides information on the conformation of the ligand in the bound state.

1.3.2.1.4 Employing other nuclei (^{19}F)

The hydrogen-1 nucleus (^{1}H) accounts for the overwhelming majority of all NMR observations of (bio)organic compounds. This privileged position of the proton is due to a sum of factors that made the observation of this nucleus very convenient. With a natural isotopic abundance close to 100% and the highest gyromagnetic ratio among natural nuclei, ^{1}H is the most sensitive nucleus of the periodic table. It has a spin I = ½, which enormously simplifies its NMR signals, and is present in virtually all organic molecules.

The virtues of the proton are in contrast with a majority of the other elements of ubiquitous presence in organic compounds. Carbon and oxygen have the disadvantage that their most abundant isotopes, ^{12}C and ^{16}O, are null-spin nuclei, and hence are NMR inactive, as is the case for ^{32}S. The majority nitrogen isotope, ^{14}N, has a spin $I = 1$, which causes its NMR signals to be significantly broadened by quadrupolar interactions, to the extent that they become often unobservable in high-resolution NMR spectrometers. For their part, ^{13}C and ^{15}N are stable, magnetically active nuclei with a spin $I = ½$, but their low natural isotopic abundance (1.1% and 0.37%, respectively) and poor sensitivity in comparison to ^{1}H preclude their use for routine NMR studies such as compound screening. However, their use in receptor-detected methods, especially in isotopically enriched samples, is of great usefulness in the drug discovery process, as will be discussed later.

Phosphorous-31 has a spin I = ½ and a 100% isotopic abundance. Depending on its oxidation state, phosphorous nuclei can adopt a broad range of chemical shifts, being the typical ^{31}P chemical shift dispersion in organic compounds of about 100 ppm. In addition to the many important enzymatic reactions involving phosphorous-containing reactants, phosphorous plays a role in many protease inhibitors by mimicking the tetrahedral intermediate of a peptide bond hydrolysis. With these considerations, ^{31}P-NMR has been regarded as a tool for compound screening [130], but the otherwise limited presence of phosphorous in drug candidates has restricted the use of ^{31}P-NMR mainly to clinical or in vivo applications, e.g. to assess metabolites such as phospholipids, ATP and inorganic phosphate.

Fluorine-19, on the other side, exhibits a number of properties that make the NMR spectroscopy of this nucleus very attractive for drug research, while at the same time being suitable for a wide range of compounds of pharmaceutical interest. Indeed, around 20% of the currently marketed drugs contain one or more fluorine atoms, a number which is steadily increasing [131]. Fluorinated analogues have long been regarded in the pharmaceutical industry for their high metabolic stability and increased membrane permeability. At the same time, fluorinated drugs are very amenable to trace in order to determine its absorption, distribution, metabolism and excretion (ADME) properties, given the virtual absence of endogenous fluorine in living organisms.

From the NMR viewpoint, ^{19}F is a $I = ½$ nucleus with a 100% isotopic abundance and a gyromagnetic ratio comparable to that of proton, being the relative sensitivity of ^{19}F 83% with respect to ^{1}H. Moreover, ^{19}F displays a much broader span of chemical shifts than ^{1}H does, ranging from -272 ppm in the CH_3F molecule to over +85 ppm in pentafluorosulfanyl (SF_5)

substituents [132], which minimizes the probability of signal overlapping and, added to the absence of fluorine in biological molecules and in most solvents and buffer components, accounts for very clean ^{19}F-NMR spectra.

Furthermore, and owing to its large chemical shift anisotropy, ^{19}F signals are highly perturbed by local changes, such as binding to a receptor. Indeed, thanks to the exquisite sensitivity of the fluorine nuclei to changes in their microenvironment, and to the potentially large difference between the ^{19}F chemical shifts in the free and bound states, even weak binding between fluorine-containing compounds and a target of interest can be easily detected in simple 1D experiments, by monitoring chemical shift perturbation of the ^{19}F signals after the addition of the receptor. In addition, appreciable changes in transverse relaxation rates (T2) upon binding to a macromolecule enables binding of fluorinated compounds to be assessed by substantial broadening of the ^{19}F signals or by changes in their intensity when a relaxation filter is applied.

The possibility to straightforwardly detect protein-ligand interactions in ^{19}F-NMR 1D experiments makes it a very efficient tool vis-à-vis ^{1}H-NMR for the high-throughput screening of chemical libraries. In addition, other usual screening procedures well-established for ^{1}H-NMR, such as STD, are also feasible by detecting ^{19}F [133]. Of course, the use of NMR spectrometry of fluorine-19 also enables structural characterization of fluorochemicals by a wide span of multidimensional experiments, both homonuclear (^{19}F-^{19}F) and heteronuclear (^{19}F-^{1}H and e.g. ^{19}F-^{13}C), such as COSY, TOCSY, NOESY, HOESY, HMBC or 3D triple resonance HCF experiments [134].

Obviously, a requisite is that the studied compounds contain at least one atom of fluorine. While the increasing interest towards ^{19}F-NMR over the recent years has prompted the release of chemical libraries composed entirely of fluorine-containing compounds, there are methodologies that allow exploiting the advantages of fluorine NMR for the screen and binding assessment of non-fluorinated compounds, e.g. through the use of fluorinated spy molecules [135]. According to this strategy, a reference experiment is first performed in the presence of the spy molecule and the target receptor. Next, the spy molecule is made to compete with a compound or mixture of compounds, and changes in the spectral properties of the spy molecule (i.e. ^{19}F chemical shift or transverse relaxation) are monitored. If the affinity data for the spy molecule are known, which can be reliably determined through ^{19}F-NMR titration experiments, competition methods also allow measuring binding constants for the screened compounds.

1.3.2.2 Receptor-detected methods: ^{1}H-^{15}N HSQC

The above mentioned techniques are useful to study the phenomenon of molecular recognition from the ligand's perspective, to detect binding events, and to providing insights into the structure of the bound saccharide. When it comes to study binding from the viewpoint of the other partner involved, i.e. the macromolecule, another set of methodologies focusing on monitoring the protein signals can provide information on the receptor binding epitopes. Given the high spectral overlap of proteins in the proton dimension, most protein-detected methods are 2D experiments (notably ^{1}H-^{15}N heteronuclear single-quantum coherence, HSQC), among which heteronuclear experiments require isotopic labeling, and thus protein expression in suitable systems, which poses a limitation to the size of the receptors studied, usually not exceeding a size of 40 kDa [129], although current advances are pushing this limit further. In addition, detailed study on receptor/ligand interactions focusing on the protein signals require a full or partial assignment of the protein ^{1}H, ^{13}C and ^{15}N resonances, which is often very time- and material-consuming. However, these setbacks are compensated by the amount of information that these techniques provide.

In its simple version, the interaction between ligands and peptides, very small proteins or domains thereof can be studied from the receptor's perspective by monitoring of their 1H signals in mono- or two-dimensional (e.g. NOESY) experiments with unlabeled receptor. In such cases, the assignment of the receptor's ^{1}H-NMR signals can be achieved with the standard strategies used for spectral assignment of small ligands. While these instances constitute only an exiguous minority of all sugar/protein systems, these mono-nuclear strategies have long been applied, for example, to study the interaction between hevein and β(1–4) linked *N*-acetylglucosamine oligosaccharides, a model system of protein-carbohydrate interactions. [21, 31]

For a majority of the sugar-protein systems, however, 2D heteronuclear experiments are required. By far, the most widely used protein-detected technique is ^{1}H-^{15}N HSQC. Heteronuclear coherence experiments are based on the transfer of magnetization between ^{1}H and a bonded heteronucleus via scalar couplings. The result of a ^{1}H-^{15}N HSQC experiment is a two-dimensional spectrum [136] showing a number of cross-peaks close to the number of residues. Prolines, which lack an amide proton, are not represented in the spectrum, while side-chain amines of nitrogen-containing residues (i.e. asparagine, glutamine, arginine, lysine and histidine) provide additional cross-peaks. If the cross-peaks are assigned to their respective residues, ^{1}H-^{15}N HSQC experiments can provide valuable information on the protein

epitopes for binding ligands, since cross-peaks belonging to residues located at the binding site become shifted upon the addition of the ligand (fig. 12).

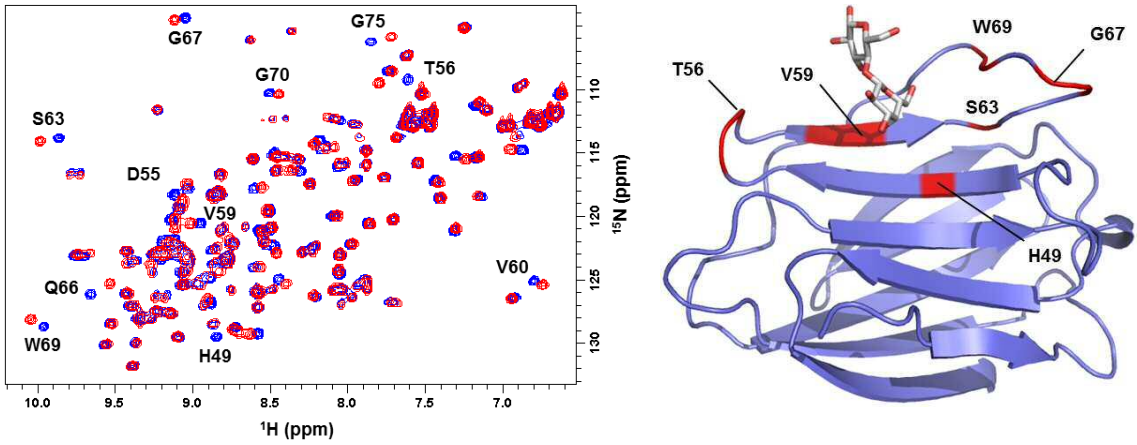


Figure 12: Left: ^{1}H-^{15}N HSQC spectrum of human galectin-7 in the absence (blue) and presence (red) of lactose, with labels indicating some of the most shifted peaks, according to its full spectral assignment [137]. Right: crystallographic structure of lactose-bound galectin-7 (PDB: 4GAL [37]), with labels indicating some of the residues with the most perturbed cross-peaks upon addition of the ligand in the ^{1}H-^{15}N HSQC experiment.

Fast relaxation of the protein ^{1}H-^{15}N signals becomes an issue when dealing with high molecular weight macromolecules (> 35 kDa). However, technological advances in hardware and novel developments in pulse sequences, such as transverse relaxation optimized spectroscopy (TROSY), [138] are pushing this limit further. TROSY-edited experiments permit the obtaining of high-resolution spectra (e.g. ^{1}H-^{15}N TROSY-HSQC) of proteins up to 100 kDa, [139] by selectively recording the most slowly relaxing signal component of the ^{15}N-^{1}H correlation.

Lastly, a variety of experiments that can be collectively referred to as selective labeling methods, are also utilized to gather insight into the ligand/protein binding from the perspective of the receptor. They all have in common the simplification of the protein NMR spectrum via selective spin-labeling of only a fraction of its nuclei, and are especially useful when they constitute an important part of the binding site. For instance, selective ^{13}C-labeling of Val, Ile and Leu side-chain methyl groups can be achieved by bacterial culturing in the presence of labeled versions of their biochemical precursors, [140] and information is obtained through ^{1}H-^{13}C-HSQC experiments. Also, the virtues of fluorine-19 NMR [141] can be exploited

by ^{19}F-labeling of proteins using fluorine-containing analogs of natural amino acids, such as fluorophenylalanine [142], and information is obtained through clean, mono-dimensional ^{19}F-NMR spectra. Finally, other strategies for selective labeling that do not require manipulation at the pre-translational level have been reported. Such is the case of solvent exposed-amines (SEA)-HSQC [143] or SEA-TROSY, [144] which allow for the selective detection of solvent-exposed NHs, and are useful for screening of interactions, since binding mainly occurs on the surface of a protein.

1.3.2.2.1 Protein NMR assignment

As already mentioned, obtaining ^{1}H-^{15}N HSQC data requires the use of ^{15}N isotopically-enriched protein and the spectral assignment of at least most ^{15}NH resonances, which is achieved by sequential assignment of the protein backbone in 3D triple ^{1}H/^{13}C/^{15}N resonance NMR experiments. For this purpose, proteins are usually expressed in bacterial hosts cultured in minimum media containing ^{13}C-glucose and/or ^{15}N-ammonium as sole carbon and nitrogen sources, respectively. Also, deuterated water can be included in culture media for proteins to be also labeled with ^{2}H at non-exchangeable hydrogens. [145] While this usually leads to a decrease in bacterial growth and protein harvest efficiency, it is used for assigning large proteins, as deuterium enrichment of the protein side-chains helps to minimize ^{1}H-mediated relaxation pathways that produce signal broadening and loss of intensity [146].

Full NMR assignment of proteins is carried out as a two-step strategy involving assignment of the backbone and, optionally, of the side-chains.

Assignment of the protein backbone is usually approached through a combination of 3D NMR experiments correlating intra- and inter-residue ^{1}H, ^{15}N, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ resonances. For example, the HN(CO)CACB experiment correlates the ^{15}NH signal of an amino acid (i) with the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ resonances of the preceding residue (i-1) [147], while the HNCACB experiment establishes the intra-residue ^{15}NH-$^{13}C_{\alpha}$-$^{13}C_{\beta}$ connectivities [148] (fig. 13). Several other possibilities exist, including CBCANH, i.e. the reversed version of HNCACB; HNCO [149] and HN(CA)CO experiments, to name a few.

Then, additional experiments are performed to assign side-chain resonances. ^{1}H-^{13}C HSQC experiments provide the ^{1}H-NMR resonances of the protons attached to carbons $^{13}C_{\alpha}$ and $^{13}C_{\beta}$, and other experiments (e.g. ^{1}H-^{1}H NOESY and H^{13}C^{13}CH-TOCSY [150]) allow for assignment of side-chain nuclei beyond these.

Besides, homo- and heteronuclear 2D experiments can be used alone for full assignment of small peptides or protein fragments, similarly to carbohydrates.

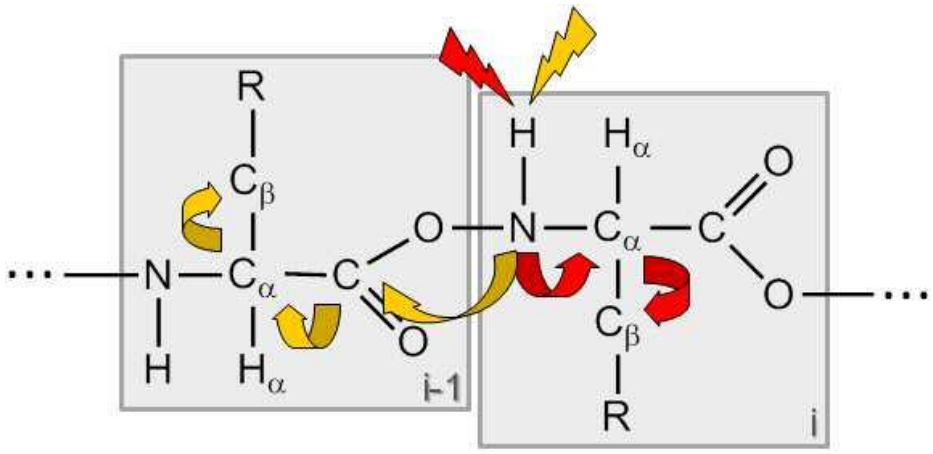


Figure 13: Examples of backbone spectral assignment experiments. Red and yellow arrows show the magnetization transfer pathway in HNCACB and HN(CO)CACB experiments, respectively. In the HNCACB experiment, the magnetization is transferred from H_i to $^{15}N_i$, and then via the ^{15}N-$^{13}C_\alpha$ *J*-coupling to the $^{13}C_{\alpha,i}$, and then from $^{13}C_{\alpha,i}$ to $^{13}C_{\beta,i}$. Then, the magnetization is transferred back to $^{15}NH_i$ for detection. Usually, transfer from $^{13}C_{\alpha,i}$ also occurs to $^{15}N_{i-1}$, becoming visible in the spectra with lower intensity than that of the intra-residue connectivity. In the HN(CO)CACB experiment, the magnetization is transferred from H_i to $^{15}N_i$, then to $^{13}CO_{i-1}$, and then to $^{13}C_{\alpha,i-1}$ and $^{13}C_{\beta,i-1}$ and their bound protons, for detection.

1.3.2.3 Paramagnetism-based NMR methods

Paramagnetism arises from the presence of unpaired electrons in atomic or molecular orbitals of certain chemical species, notably metal ions and organic radicals. An emerging family of methods taking advantage of paramagnetic effects, including Residual Dipolar Couplings (RDCs), Pseudo Contact Shifts (PCSs), and Paramagnetic Relaxation Enhancements (PREs) [151] are currently starting to be applied for structural characterization of glycans and their interactions with receptors.

RDCs, which originate from dipolar interactions between a pair of nuclei (e.g. ^1H–^{13}C, ^1H–^1H), causing a splitting in the signal of each nucleus involved, permit to correlate the relative geometrical orientation between structural fragments (e.g. sugar rings), independently of how apart they are in the molecule [152-155]. RDCs are particularly useful to deduce the geometry of the glycosidic linkages, which has a major influence on the conformation of glycans and

their overall shape [156]. RDC measurements require the partial alignment of the molecule under study. [157-158] Several alignment media have been tested for carbohydrates, using both external alignment systems [159-160] and paramagnetic tags [161].

Pseudo Contact Shifts (PCSs) [162-163] also arise from a dipolar interaction, in this case between the unpaired electrons of a paramagnetic entity and the nuclei in its vicinity. PCS-derived information is analogous to NOE restraints in the sense that it is distance-dependent. The main advantage of PCS effects is their longer-range effect, due to the favorable r^{-3} distance dependency, instead of r^{-6} in the case of NOEs.

Finally, PREs are due to enhancements of dipolar relaxation of nuclei caused by the presence of a paramagnetic species in its vicinity. PREs also show the typical r^{-6} distance dependence, and may be employed to obtain information on the distance between the paramagnetic center and the monitored nuclear spins.

In the context of carbohydrate-protein interactions, it is worth mentioning the application of nitroxide spin-labels for mapping the sugar binding sites on protein surfaces, thanks to PREs arisen from the nitroxide moiety. Perturbations in the ^{1}H-^{15}N HSQC spectrum of galectin-3 were monitored and used to identify protein residues proximate to the binding site for N-acetyllactosamine [151].

RDCs have also been successfully employed for the study of protein-carbohydrate complexes. In this regard, RDCs have been applied to determine the glycan conformation in the bound state [164], the location of the sugar binding site and the structure of the complex, in combination with PCS-derived information [165].

1.3.2.3 Diffusion-based methods: DOSY

Diffusion-ordered spectroscopy (DOSY) is based on a pulse-field gradient spin-echo NMR experiment, in which the different components of a sample mixture experience diffusion. As a consequence, the signals of each component decay at different rates as the gradient strength varies. Thus, it is possible to obtain a 2D NMR spectrum, in which the first dimension is the regular NMR chemical shift, and the additional dimension is related to the diffusion coefficient. This allows for the obtaining of virtually deconvoluted NMR spectra, with many potential applications, such as identifying individual components and impurities in complex mixtures.

Importantly, self-diffusion of a molecule is related to its size, as it is deduced from the Stokes-Einstein equation:

$$D = \frac{k_B T}{f} \quad (8)$$

where D is the diffusion coefficient, k_B is the Boltzmann constant, T is the temperature, and f is the friction coefficient. If the molecule is considered as a spherical particle with an effective hydrodynamic radius (i.e. the Stokes radius) r_S in a solution of viscosity η, then the friction coefficient is given by:

$$f = 6 \quad _S \quad (9)$$

Of course, the hydrodynamic radius of a molecule is related to its molecular weight [166]. This can be exploited for estimating the size of chemical species through DOSY experiments, most commonly by extrapolating obtained D values from calibrated data [167].

1.3.3 Assisting NMR observations with computational protocols

During the past decades, the development of molecular modeling techniques has allowed for successful prediction of the structure of ligands, as well as their interactions with receptors. Indeed, the award of the Nobel Prize in Chemistry 2013 to M. Karplus, M. Levitt, and A. Warshel "for the development of multiscale models for complex chemical systems" shows the increasing importance that computer models have gained through the years.

The combination of computational protocols with NMR experimental data has proven extremely useful to deduce the conformational and dynamic properties of free and bound carbohydrate molecules [168]. Molecular modeling of saccharides can be performed at different levels of complexity. The most employed protocols are based on force fields, which are used for performing molecular mechanics (MM) and dynamics (MD) calculations. In glycoscience, GLYCAM parameters [169] are the most commonly employed for reproducing dynamic and electronic properties of sugars. Nevertheless, more sophisticated computational chemistry methods based on quantum chemistry may also be employed. *Ab initio* calculations

are used to solve the Schrödinger equation through a series of approximations, and have also been employed to model a variety of systems in the carbohydrate field [168].

In addition, the combination of NMR experiments with docking and MD protocols can be used for obtaining the 3D structure of carbohydrate-protein complexes [170]. Ideally, experimental information from ligand-based and receptor-based NMR experiments is available, thus providing full empirical information on the recognition process that can be used as input for computational studies. Docking protocols are used to explore the possible interactions of a given carbohydrate ligand with a particular receptor (fig. 14A). These programs (e.g. AutoDock [171]) evaluate the free energy of the interaction for different possible binding modes, in which the conformation of the sugar and/or the receptor may change. Further refinement of the proposed docking poses may be accomplished through molecular dynamics simulations (fig. 14B), and may take into account the available experimental information as knowledge-based restrictions (i.e., through NMR).

Other computing methods have been conceived to predict or interpret specific NMR phenomena, such as the simulation of STD events by the CORCEMA (Complete relaxation and conformational exchange matrix) package, which permits to assess the agreement of the predicted docking based on trNOESY [172] and/or STD [173] experimental data.

Finally, for carbohydrate-processing enzymes, methods combining quantum mechanics with molecular dynamics (QM/MM) have been developed [174]. They permit to define the region of the catalytic site (the ligand and protein residues of interest) to be studied through *ab initio* calculations, leaving the rest of the protein subject to standard molecular mechanics. Catalytic processes or reaction pathways can thus be modeled [175-176].

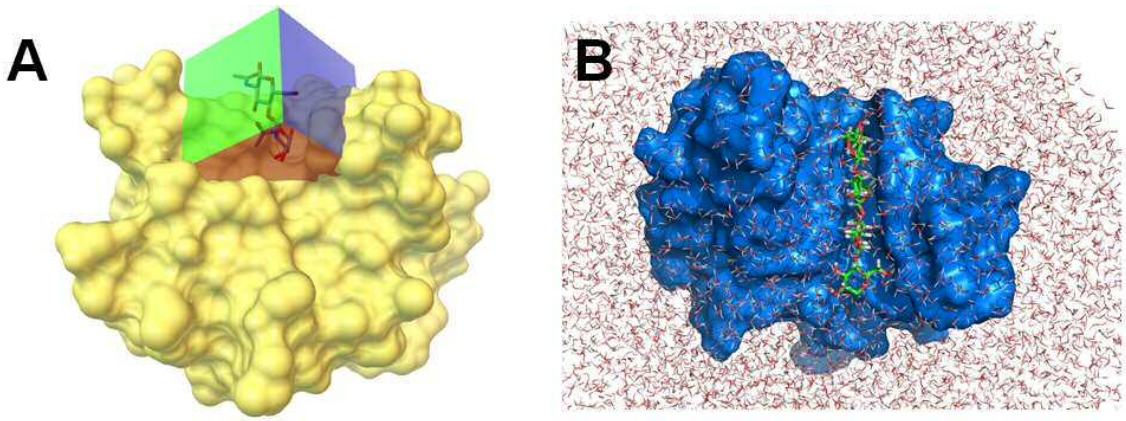


Figure 14: Computational protocols. A: AutoDock [171] calculation using the crystallographic structure of human galectin-3 (PDB: 1A3K) and the structure of *N*,*N*'-diacetyllactosamine, treated as a flexible ligand, allowed to explore a 34 Å × 34 Å × 34 Å cubic box around the sugar binding site. B: MD calculation, in explicit solvent (water), on the binding of cellotetraose to the carbohydrate-binding module of a *Clostridium* cellulosic enzyme [177].

1.3.4 Screening. Fragment-based drug discovery

The screening of a battery of compounds against a macromolecular target of biomedical interest is the fuel that feeds the drug discovery engine. Regardless of the many biochemical and biophysical techniques used for the detection and evaluation of binding hits, two different philosophies of compound screening coexist in the current pharmaceutical landscape: high-throughput screening (HTS) and fragment-based drug discovery (FBDD). The aim of HTS, which has dominated the drug discovery activity for much of its history, is the evaluation of large numbers of heavy, drug-like molecules. Typically, hundreds of thousands to several millions of compounds are screened for each target, and high potency inhibitors are sought. For its part, FBDD is based on the screening of small libraries of low molecular weight compounds, called fragments, in pursuit of hits with low, but efficient binding activity, which are then combined or developed into more potent inhibitors via different strategies, as illustrated in fig. 15. Over the last decade or so, FBDD has gained increasing acceptance both in the industry and in the academia. The success of this new drug discovery paradigm is illustrated by the recent approval of the first fragment-based drug (Vemurafenib) [178] by the FDA.

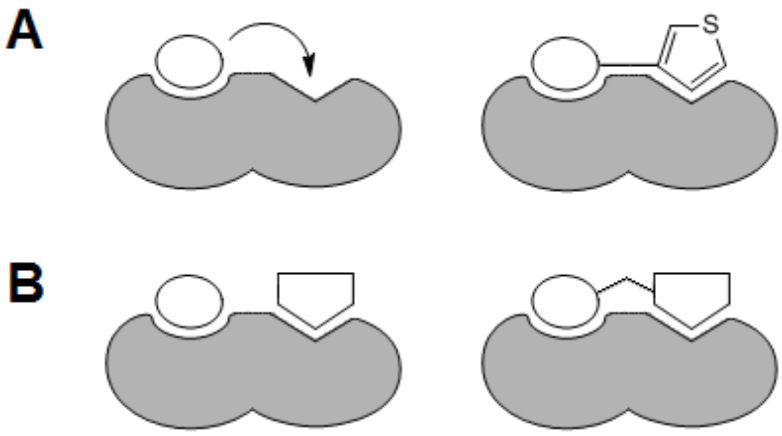


Figure 15: Two strategies for lead generation following a fragment-based approach: (a) Fragment growth; (b) Fragment linking.

From the beginning, NMR has represented by far the preferred biophysical technique in FBDD programs, due to its reliability and versatility. Indeed, thanks to the wide variety of NMR experiments available, the use of this spectroscopy can guide the whole process (fig. 16), from the initial screening of compounds, to the ranking of the hits according to their affinity or efficiency and the characterization of the binding modes, both from the ligand and the receptor viewpoints [179].

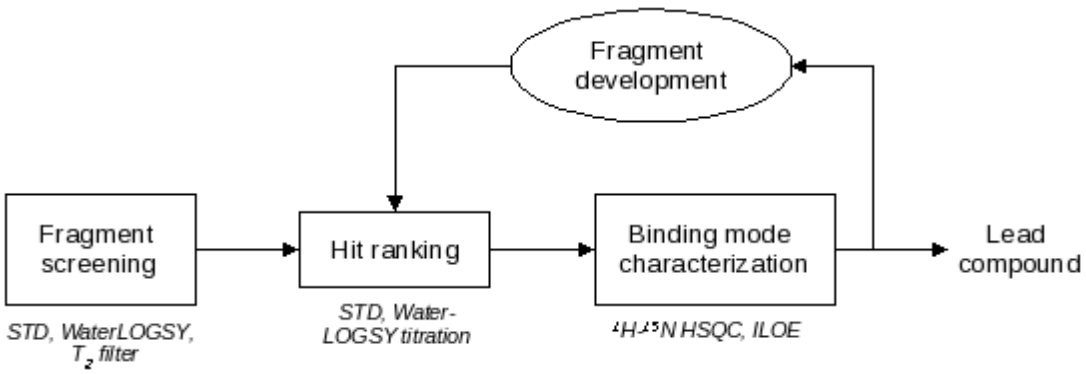


Figure 16: Flow scheme for an NMR-driven fragment-based screening.

1.3.4.1 Fragment screening and ranking

1.3.4.1.2 Magnetization transfer experiments: STD and WaterLOGSY

Most usually, fragments are screened by using a magnetization transfer-based technique, namely STD or WaterLOGSY. These methods present the advantage that their detection range is suitable to the weak affinities expected for fragment-receptor interactions. To speed up the process and minimize protein consumption, experiments are performed on cocktails of fragments, in a number such that signal overlapping does not preclude the identification of the compounds and their solubility is not compromised. Normally, 6 to 10 fragments are assayed in each experiment. The composition of the mixtures might be entirely random, although it is more advisable to bring together fragments as chemically different as possible, in order to easily identify hits by inspection of the ^{1}H-NMR spectrum without need to deconvolute the mixture. Alternatively, mixtures can be made up of similar fragments, if one wishes to identify structural features or scaffolds with binding activities toward the target.

Ranking of the fragment hits can be undertaken by determining the dissociation constants with either STD or WaterLOGSY titration experiments, as described in previous units. Sometimes, fragment hits are ranked by their ligand efficiencies (LE), rather than their dissociation constants. The LE (eq. 10) is defined as the free energy of binding per heavy atom (*N*), and its measurement is useful for comparing ligands with different potencies and different sizes, as it normalizes the affinity of a given compound to its size [180]:

$$\mathrm{LE} = \frac{-\Delta G_{bind}}{N} = \frac{-RT \cdot \ln K_D}{N} \qquad (10)$$

1.3.4.1.3 Transverse relaxation rates

Another popular and fast technique for fragment screening is based on monitoring the changes in T2 relaxation of the fragment signals upon binding to the target. Ligands that bind to a macromolecular receptor experience an enhancement of their T2 relaxation, which is translated into a broadening of their signals. In the practice, however, rather than directly measuring the line-width of the ligand signals, a relaxation filter is applied by setting a spin-lock between the 90º excitation pulse and the signal acquisition, and a comparison of spectra

recorded in the presence and in the absence of the target will reveal a change in the intensity of signals corresponding to binding compounds. Depending on the duration of the spin-lock filter, a detection cut-off can be tuned in order to scope for stronger or weaker-binding fragments, with longer spin-lock times enabling to detect weaker interactions. This possibility also introduces a way to qualitatively rank fragment hits, i.e. by comparing series of experiments performed with different spin-lock times.

1.3.4.1.4 Paramagnetic labeling of proteins: SLAPSTIC

Paramagnetic tagging of proteins has been also used for compound screening. According to this method, dubbed SLAPSTIC (Spin Labels Attached to Protein Side chains as Tool to identify Interacting Compounds) [181], a paramagnetic label is covalently attached to a specific site of the target (e.g. a Cys thiol group), and hits are detected by the transmission of paramagnetic effects (mainly PREs) to the binding fragments, most usually through a transverse relaxation filter. The paramagnetic label should be judiciously placed, so that it lies relatively close to the active site, as paramagnetic effects decrease with the distance, but its attachment does not alter the structure and recognition properties of the protein. Therefore, it is required not only a previous knowledge of the structure of the target, but also to perform genetic engineering in the case that a suitable tagging site does not exist in the natural protein.

A modification of this strategy, involving the use of a paramagnetic derivative of a fragment with demonstrated binding activity toward the target, can be used to discover fragments binding to different sites [182]. At saturating concentrations of this fragment, a screening is performed looking for second binders, which will be detected by the transmission of paramagnetic effects from the first fragment. This event only is possible if both molecules are bound at the same time to the protein. This method presents notorious advantages in comparison to other procedures, which are essentially based on performing any of the mentioned NMR screening experiments (e.g., STD) in the presence of saturating concentrations of a first binder, since, for weakly-binding compounds, such as fragments, it is often difficult to achieve saturating concentrations before reaching their solubility limit, and false positives can arise.

1.3.4.1.5 Competition approaches

Finally, for compound screening, indirect approaches, in the shape of competition experiments involving a spy molecule, can be employed. According to this strategy, the NMR properties of a known ligand for a given target are monitored before and after the addition of the screened compounds, using any of the NMR experiments presented here. If the spy molecule is fluorinated, ^{19}F-NMR experiments can be carried out.

Competition approaches present the advantage that, if the binding parameters of the spy molecule are well-determined, those of the competitors can be easily derived applying mathematical models. For example, assuming a competitive inhibition model between a fragment hit and a spy molecule for binding to a target, the dissociation constant (K_D) for the spy molecule (*F*) is shifted to an apparent value (K_D^{ap}) in the presence of non-zero concentrations of a fragment hit (F) displaying an inhibition constant K_I.

$$K_D^{ap} = K_D\left(1+\frac{F}{K_I}\right) \tag{11}$$

Therefore, the relation between the observed values for an NMR property of the spy molecule S (e.g. the intensity of a signal applying a T2 relaxation filter) after (*I*) and before (I_0) the addition of the competitor fragment can be represented as:

$$\frac{I}{I_0} = \frac{\dfrac{I_{\max}\cdot[\mathrm{S}]}{K_D\left(1+\dfrac{[\mathrm{F}]}{K_I}\right)+[\mathrm{S}]}}{\dfrac{I_{\max}\cdot[\mathrm{S}]}{K_D+[\mathrm{S}]}} = \frac{K_I\cdot K_D + K_I\cdot[\mathrm{S}]}{K_I\cdot K_D + K_D\cdot[\mathrm{F}] + K_I\cdot[\mathrm{S}]} \tag{12}$$

And, for selected experimental conditions where [S] = K_D,

$$\frac{I}{I_0} = \frac{2\cdot K_I}{2\cdot K_I + F} \tag{13}$$

The principal drawbacks associated with competition methods are the impossibility to identify the active compounds if screened within mixtures containing a very high affinity ligand, and

the occurrence of false-negatives when fragments bind to sites different from that of the spy molecule.

1.3.4.3 Binding mode characterization

When it comes to get a detailed picture on the geometry of the ligand-receptor complex, the versatility of NMR offers the scientist a wide variety of experiments, both ligand- and receptor-detected, providing valuable information from different points of view.

1.3.4.3.1 Protein-detected techniques: ^{1}H-^{15}N HSQC experiments

Competition experiments themselves represent a first insight into the characterization of the binding mode, since, taking allosteric effects aside, if the binding site of the spy molecule is known, so is that of the studied fragment.

In order to delineate in more detail the binding epitope of the protein, changes in the resonances of the latter are monitored upon the addition of the fragment, usually in ^{1}H-^{15}N HSQC experiments. By doing so, amide groups closest to the ligand are identified by the strong perturbation of the chemical shifts of their respective cross-peaks in the spectrum. This information is given as input to computational modeling programs in order to get a 3D view of the binding. Understanding the mode of binding of the hits in the three-dimensional context of the protein is of great importance to design potentially stronger binding compounds (fig. 15A).

1.3.4.3.2 Ligand-detected techniques: Inter-ligand NOEs, STD, trNOESY

When two or more fragments bind to different, but adjacent sites on the target surface, knowledge on their relative orientation and distance becomes essential in order to attempt the connection of the fragments (fig. 15B) via a suitable linker to create a larger, more potent compound. In this regard, NOESY experiments can provide valuable information in the shape of inter-ligand NOEs (iLOEs) [183] given rise to by pairs of protons belonging to different fragments, but located close in space when docked at the binding site.

It has been already explained how one can take advantage of STD and trNOESY experiments to obtain information on the ligand epitopes more exposed to the protein surface and the

conformation of the bound ligand, respectively. While for fragments, as relatively simple and small molecules, such experiments might not provide decisive information, they can be of usefulness for the study of the bound structure of lead compounds in more advanced stages.

1.4 Objectives

The general objectives of this Thesis are the study of interaction processes between carbohydrates and a series of protein receptors of different types, and the development of novel NMR procedures which permit to advance in the knowledge of carbohydrate structure and recognition features.

This Thesis is divided into different Chapters, which have the following specific objectives:

Chapter 2:

- Explore, at atomic detail, the interaction features of an array of viral fiber receptors toward small oligosaccharides, using ligand-detected NMR experiments.

Chapter 3:

- Characterization of the interaction between glycogen and related oligosaccharides and a monoclonal antibody through ligand-detected experiments.

- Structural characterization of the α-glucan envelope of *Mycobacterium smegmatis* through 2D heteronuclear NMR experiments.

Chapter 4:

- Exploration of the catalytic mechanism of the UDP:glucose pyrophosphorylase (UGP) of *Streptococcus pneumoniae*.

- Development of a direct, NMR-based assay to measure UGP activity.

- Targeting of the UGP of *Streptococcus pneumoniae* through fragment-based drug discovery.

Chapter 5:

- Full NMR assignment of galectin-3 (gal-3) and galectin-7 (gal-7).

- Understanding of the conformational duality of gal-7 and structural characterization of the minor conformer thereof.

- Targeting of gal-7 through fragment-based drug discovery.

- Characterization of the interaction between gal-3 and a series of glycan ligands, using both ligand- and protein-detected NMR methods, assisted by molecular modeling.

- Provide insight into the interacting capabilities of the N-terminal domain of gal-3 toward the gal-3 carbohydrate recognition domain.

Chapter 6:

- Developing novel methods for the application of paramagnetic effects to the study of carbohydrate conformation and recognition features by receptors

1.5 References

[1] Enriquez-Navas, P. M., Marradi, M., Padro, D., Angulo, J., and Penades, S. (2011), *Chemistry* **17**, 1547-1560.

[2] Silipo, A., Larsbrink, J., Marchetti, R., Lanzetta, R., Brumer, H., and Molinaro, A. (2012), *Chemistry* **18**, 13395-13404.

[3] Silipo, A., Sturiale, L., De Castro, C., Lanzetta, R., Parrilli, M., Garozzo, D., and Molinaro, A. (2012), *Carbohydr Res* **357**, 75-82.

[4] Nieto, L., Canales, A., Gimenez-Gallego, G., Nieto, P. M., and Jimenez-Barbero, J. (2011), *Chemistry* **17**, 11204-11209.

[5] Perepelov, A. V., Shekht, M. E., Liu, B., Shevelev, S. D., Ledov, V. A., Senchenkova, S. N., L'Vov V, L., Shashkov, A. S., Feng, L., Aparin, P. G., Wang, L., and Knirel, Y. A. (2012), *FEMS Immunol Med Microbiol* **66**, 201-210.

[6] Gabius, H. (2009), *Wiley-VCH* **28**.

[7] Dwek, R. A. (1996), *Chem Rev* **96**, 683-720.

[8] Gabius, H. J., Siebert, H. C., Andre, S., Jimenez-Barbero, J., and Rudiger, H. (2004), *Chembiochem* **5**, 740-764.

[9] Varki, A. (1993), *Glycobiology* **3**, 97-130.

[10] Angyal, S. J. (1969), *Angewandte Chemie International Edition in English* **8**, 157-166.

[11] Sattelle, B. M., and Almond, A. (2011), *Glycobiology* **21**, 1651-1662.

[12] Juaristi, E., and Cuevas, G. (1992), *Tetrahedron* **48**, 5019-5087.

[13] Lii, J. H., Chen, K. H., and Allinger, N. L. (2003), *J Comput Chem* **24**, 1504-1513.

[14] Lii, J. H., Chen, K. H., Durkin, K. A., and Allinger, N. L. (2003), *J Comput Chem* **24**, 1473-1489.

[15] Widmalm, G. (2013), *Carbohydr Res* **378**, 123-132.

[16] Xu, B., Unione, L., Sardinha, J., Wu, S., Etheve-Quelquejeu, M., Pilar Rauter, A., Bleriot, Y., Zhang, Y., Martin-Santamaria, S., Diaz, D., Jimenez-Barbero, J., and Sollogoub, M. (2014), *Angew Chem Int Ed Engl* **53**, 9597-9602.

[17] Franca, E. F., Lins, R. D., Freitas, L. C. G., and Straatsma, T. P. (2008), *Journal of Chemical Theory and Computation* **4**, 2141-2149.

[18] Rah, M. J. (2011), *Optometry* **82**, 38-43.

[19] Esko, J. D., Kimata, K., and Lindahl, U. (2009).
[20] Sasisekharan, R., Raman, R., and Prabhakar, V. (2006), *Annu Rev Biomed Eng* **8**, 181-231.
[21] del Carmen Fernandez-Alonso, M., Diaz, D., Berbis, M. A., Marcelo, F., Canada, J., and Jimenez-Barbero, J. (2012), *Curr Protein Pept Sci* **13**, 816-830.
[22] Gabius, H. J. (2000), *Naturwissenschaften* **87**, 108-121.
[23] McArdle, W. D., Katch, F. I., and Katch, V. L. (2007), Lippincott Williams & Wilkins.
[24] Fischer, E. (1894), *Berichte der deutschen chemischen Gesellschaft* **27**, 2985-2993.
[25] Hardy, B. J. (1997), *Journal of Molecular Structure: THEOCHEM* **395–396**, 187-200.
[26] Piotukh, K., Serra, V., Borriss, R., and Planas, A. (1999), *Biochemistry* **38**, 16092-16104.
[27] Chavez, M. I., Vila-Perello, M., Canada, F. J., Andreu, D., and Jimenez-Barbero, J. (2010), *Carbohydr Res* **345**, 1461-1468.
[28] Rivera-Sagredo, A., Solis, D., Diaz-Maurino, T., Jimenez-Barbero, J., and Martin-Lomas, M. (1991), *Eur J Biochem* **197**, 217-228.
[29] Solis, D., Fernandez, P., Diaz-Maurino, T., Jimenez-Barbero, J., and Martin-Lomas, M. (1993), *Eur J Biochem* **214**, 677-683.
[30] Solis, D., Jimenez-Barbero, J., Martin-Lomas, M., and Diaz-Maurino, T. (1994), *Eur J Biochem* **223**, 107-114.
[31] Jimenez-Barbero, J., Javier Canada, F., Asensio, J. L., Aboitiz, N., Vidal, P., Canales, A., Groves, P., Gabius, H. J., and Siebert, H. C. (2006), *Adv Carbohydr Chem Biochem* **60**, 303-354.
[32] Vandenbussche, S., Diaz, D., Fernandez-Alonso, M. C., Pan, W., Vincent, S. P., Cuevas, G., Canada, F. J., Jimenez-Barbero, J., and Bartik, K. (2008), *Chemistry* **14**, 7570-7578.
[33] Ramirez-Gualito, K., Alonso-Rios, R., Quiroz-Garcia, B., Rojas-Aguilar, A., Diaz, D., Jimenez-Barbero, J., and Cuevas, G. (2009), *J Am Chem Soc* **131**, 18129-18138.
[34] del Carmen Fernandez-Alonso, M., Canada, F. J., Jimenez-Barbero, J., and Cuevas, G. (2005), *J Am Chem Soc* **127**, 7379-7386.
[35] Chavez, M. I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Groves, P., Asensio, J. L., Asensio, G., Muraki, M., Canada, F. J., and Jimenez-Barbero, J. (2005), *Chemistry* **11**, 7060-7074.
[36] DiGabriele, A. D., Lax, I., Chen, D. I., Svahn, C. M., Jaye, M., Schlessinger, J., and Hendrickson, W. A. (1998), *Nature* **393**, 812-817.
[37] Leonidas, D. D., Vatzaki, E. H., Vorum, H., Celis, J. E., Madsen, P., and Acharya, K. R. (1998), *Biochemistry* **37**, 13930-13940.
[38] Asensio, J. L., Canada, F. J., Bruix, M., Gonzalez, C., Khiar, N., Rodriguez-Romero, A., and Jimenez-Barbero, J. (1998), *Glycobiology* **8**, 569-577.
[39] Gourdine, J. P., Cioci, G., Miguet, L., Unverzagt, C., Silva, D. V., Varrot, A., Gautier, C., Smith-Ravin, E. J., and Imberty, A. (2008), *J Biol Chem* **283**, 30112-30120.
[40] Ginsburg, V. (1978), *Prog Clin Biol Res* **23**, 595-600.
[41] Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008), *Annu Rev Biochem* **77**, 521-555.
[42] Oberli, M. A., Tamborrini, M., Tsai, Y. H., Werz, D. B., Horlacher, T., Adibekian, A., Gauss, D., Moller, H. M., Pluschke, G., and Seeberger, P. H. (2010), *J Am Chem Soc* **132**, 10239-10241.
[43] Gabius, H. J., Andre, S., Kaltner, H., and Siebert, H. C. (2002), *Biochim Biophys Acta* **1572**, 165-177.
[44] Gabius, H. J., Andre, S., Jimenez-Barbero, J., Romero, A., and Solis, D. (2011), *Trends Biochem Sci* **36**, 298-313.
[45] van Raaij, M. J., and Mitraki, A. (2013), In *Proteins in Solution and at Interfaces*, pp 219-232, John Wiley & Sons, Inc.
[46] Karlsson, K. A. (2001), *Adv Exp Med Biol* **491**, 431-443.
[47] Delacour, D., Koch, A., and Jacob, R. (2009), *Traffic* **10**, 1405-1413.

[48] Imai, Y., True, D. D., Singer, M. S., and Rosen, S. D. (1990), *J Cell Biol* **111**, 1225-1232.
[49] Poirier, F., and Kimber, S. (1997), *Mol Hum Reprod* **3**, 907-918.
[50] Bies, C., Lehr, C. M., and Woodley, J. F. (2004), *Adv Drug Deliv Rev* **56**, 425-435.
[51] Griffiths, G. D. (2011), *Toxins (Basel)* **3**, 1373-1392.
[52] Asensio, J. L., Canada, F. J., Bruix, M., Rodriguez-Romero, A., and Jimenez-Barbero, J. (1995), *Eur J Biochem* **230**, 621-633.
[53] Shibuya, N., Goldstein, I. J., Shafer, J. A., Peumans, W. J., and Broekaert, W. F. (1986), *Arch Biochem Biophys* **249**, 215-224.
[54] Nagata, Y., and Burger, M. M. (1974), *J Biol Chem* **249**, 3116-3122.
[55] Dodd, R. B., and Drickamer, K. (2001), *Glycobiology* **11**, 71R-79R.
[56] Braakman, I. (2001), *EMBO Rep* **2**, 666-668.
[57] Dahms, N. M., Lobel, P., and Kornfeld, S. (1989), *J Biol Chem* **264**, 12115-12118.
[58] Kilpatrick, D. C. (2002), *Biochim Biophys Acta* **1572**, 187-197.
[59] Cummings, R. D., and Liu, F. T. (2009).
[60] Vasta, G. R. (2009), *Nat Rev Microbiol* **7**, 424-438.
[61] Cooper, D. N., and Barondes, S. H. (1999), *Glycobiology* **9**, 979-984.
[62] Kaltner, H., and Gabius, H. J. (2012), *Histol Histopathol* **27**, 397-416.
[63] Hirabayashi, J., Satoh, M., and Kasai, K. (1992), *J Biol Chem* **267**, 15485-15490.
[64] Pfeifer, K., Haasemann, M., Gamulin, V., Bretting, H., Fahrenholz, F., and Muller, W. E. (1993), *Glycobiology* **3**, 179-184.
[65] Berbis, M. A., Andre, S., Canada, F. J., Pipkorn, R., Ippel, H., Mayo, K. H., Kubler, D., Gabius, H. J., and Jimenez-Barbero, J. (2014), *Biochem Biophys Res Commun* **443**, 126-131.
[66] Ahmad, N., Gabius, H. J., Andre, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B., Macaluso, F., and Brewer, C. F. (2004), *J Biol Chem* **279**, 10841-10847.
[67] Lopez-Lucendo, M. F., Solis, D., Andre, S., Hirabayashi, J., Kasai, K., Kaltner, H., Gabius, H. J., and Romero, A. (2004), *J Mol Biol* **343**, 957-970.
[68] Yang, R. Y., Rabinovich, G. A., and Liu, F. T. (2008), *Expert Rev Mol Med* **10**, e17.
[69] Miller, M. C., Ribeiro, J. P., Roldos, V., Martin-Santamaria, S., Canada, F. J., Nesmelova, I. A., Andre, S., Pang, M., Klyosov, A. A., Baum, L. G., Jimenez-Barbero, J., Gabius, H. J., and Mayo, K. H. (2011), *Glycobiology* **21**, 1627-1641.
[70] Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998), *J Biol Chem* **273**, 13047-13052.
[71] Varki, A., and Chrispeels, M. J. (1999), Cold Spring Harbor Laboratory Press.
[72] Camby, I., Le Mercier, M., Lefranc, F., and Kiss, R. (2006), *Glycobiology* **16**, 137R-157R.
[73] Chen, H. Y., Sharma, B. B., Yu, L., Zuberi, R., Weng, I. C., Kawakami, Y., Kawakami, T., Hsu, D. K., and Liu, F. T. (2006), *J Immunol* **177**, 4991-4997.
[74] Saussez, S., and Kiss, R. (2006), *Cell Mol Life Sci* **63**, 686-697.
[75] Graessler, J., Spitzenberger, F., Graessler, A., Parpart, B., Kuhlisch, E., Kopprasch, S., and Schroeder, H. E. (2000), *Adv Exp Med Biol* **486**, 179-183.
[76] Than, N. G., Pick, E., Bellyei, S., Szigeti, A., Burger, O., Berente, Z., Janaky, T., Boronkai, A., Kliman, H., Meiri, H., Bohn, H., Than, G. N., and Sumegi, B. (2004), *Eur J Biochem* **271**, 1065-1078.
[77] Canales, A., Mallagaray, A., Berbis, M. A., Navarro-Vazquez, A., Dominguez, G., Canada, F. J., Andre, S., Gabius, H. J., Perez-Castells, J., and Jimenez-Barbero, J. (2014), *J Am Chem Soc* **136**, 8011-8017.
[78] Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W. E., Yagi, F., and Kasai, K. (2002), *Biochim Biophys Acta* **1572**, 232-254.
[79] Toscano, M. A., Bianco, G. A., Ilarregui, J. M., Croci, D. O., Correale, J., Hernandez, J. D., Zwirner, N. W., Poirier, F., Riley, E. M., Baum, L. G., and Rabinovich, G. A. (2007), *Nat Immunol* **8**, 825-834.

[80] Carlsson, S., Oberg, C. T., Carlsson, M. C., Sundin, A., Nilsson, U. J., Smith, D., Cummings, R. D., Almkvist, J., Karlsson, A., and Leffler, H. (2007), *Glycobiology* **17**, 663-676.
[81] Vokhmyanina, O. A., Rapoport, E. M., Ryzhov, I. M., Korchagina, E. Y., Pazynina, G. V., Severov, V. V., Kaltner, H., Andre, S., Gabius, H. J., and Bovin, N. V. (2011), *Biochemistry (Mosc)* **76**, 1185-1192.
[82] Nieminen, J., Kuno, A., Hirabayashi, J., and Sato, S. (2007), *J Biol Chem* **282**, 1374-1383.
[83] Leppanen, A., Stowell, S., Blixt, O., and Cummings, R. D. (2005), *J Biol Chem* **280**, 5549-5562.
[84] Liu, F. T., Patterson, R. J., and Wang, J. L. (2002), *Biochim Biophys Acta* **1572**, 263-273.
[85] Villeneuve, C., Baricault, L., Canelle, L., Barboule, N., Racca, C., Monsarrat, B., Magnaldo, T., and Larminat, F. (2011), *Mol Biol Cell* **22**, 999-1013.
[86] Inagaki, Y., Higashi, K., Kushida, M., Hong, Y. Y., Nakao, S., Higashiyama, R., Moro, T., Itoh, J., Mikami, T., Kimura, T., Shiota, G., Kuwabara, I., and Okazaki, I. (2008), *Gastroenterology* **134**, 1180-1190.
[87] Yang, R. Y., Hsu, D. K., and Liu, F. T. (1996), *Proc Natl Acad Sci U S A* **93**, 6737-6742.
[88] Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E., and Kloog, Y. (2001), *Oncogene* **20**, 7486-7493.
[89] Rotblat, B., Niv, H., Andre, S., Kaltner, H., Gabius, H. J., and Kloog, Y. (2004), *Cancer Res* **64**, 3112-3118.
[90] Rossi, B., Espeli, M., Schiff, C., and Gauthier, L. (2006), *J Immunol* **177**, 796-803.
[91] Liu, F. T., and Hsu, D. K. (2007), *Drug News Perspect* **20**, 455-460.
[92] Rabinovich, G. A., Toscano, M. A., Jackson, S. S., and Vasta, G. R. (2007), *Curr Opin Struct Biol* **17**, 513-520.
[93] Dam, T. K., and Brewer, C. F. (2010), *Glycobiology* **20**, 270-279.
[94] Levroney, E. L., Aguilar, H. C., Fulcher, J. A., Kohatsu, L., Pace, K. E., Pang, M., Gurney, K. B., Baum, L. G., and Lee, B. (2005), *J Immunol* **175**, 413-420.
[95] Fermino, M. L., Polli, C. D., Toledo, K. A., Liu, F. T., Hsu, D. K., Roque-Barreira, M. C., Pereira-da-Silva, G., Bernardes, E. S., and Halbwachs-Mecarelli, L. (2011), *PLoS One* **6**, e26004.
[96] Mey, A., Leffler, H., Hmama, Z., Normier, G., and Revillard, J. P. (1996), *J Immunol* **156**, 1572-1577.
[97] Li, Y., Komai-Koma, M., Gilchrist, D. S., Hsu, D. K., Liu, F. T., Springall, T., and Xu, D. (2008), *J Immunol* **181**, 2781-2789.
[98] van den Berg, T. K., Honing, H., Franke, N., van Remoortere, A., Schiphorst, W. E., Liu, F. T., Deelder, A. M., Cummings, R. D., Hokke, C. H., and van Die, I. (2004), *J Immunol* **173**, 1902-1907.
[99] Pelletier, I., and Sato, S. (2002), *J Biol Chem* **277**, 17663-17670.
[100] Pelletier, I., Hashidate, T., Urashima, T., Nishi, N., Nakamura, T., Futai, M., Arata, Y., Kasai, K., Hirashima, M., Hirabayashi, J., and Sato, S. (2003), *J Biol Chem* **278**, 22223-22230.
[101] Jin, R., Greenwald, A., Peterson, M. D., and Waddell, T. K. (2006), *J Immunol* **177**, 1289-1295.
[102] Greenwald, A. G., Jin, R., and Waddell, T. K. (2009), *Transplantation* **87**, 44-51.
[103] Krzeminski, M., Singh, T., Andre, S., Lensch, M., Wu, A. M., Bonvin, A. M., and Gabius, H. J. (2011), *Biochim Biophys Acta* **1810**, 150-161.
[104] Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., Ju, T., Molinaro, R. J., Rivera-Marrero, C., Xia, B., Smith, D. F., and Cummings, R. D. (2010), *Nat Med* **16**, 295-301.
[105] Toone, E. (1994), *Current Opinion in Structural Biology* **4**, 719-728.
[106] Weis, W. I., and Drickamer, K. (1996), *Annu Rev Biochem* **65**, 441-473.
[107] Roldós, V., Cañada, F. J., and Jimenez-Barbero, J. (2011), *Chembiochem* **12**, 1-17.

[108] Claridge, T. D. W. (1999), *Tetrahedron Organic Chemistry Series* **19**.
[109] Tekely, P. (2002), *Magnetic Resonance in Chemistry* **40**, 800-800.
[110] Claridge, T. D. W. (2008), Elsevier Science.
[111] Calle, L. P., Canada, F. J., and Jimenez-Barbero, J. (2011), *Nat Prod Rep* **28**, 1118-1125.
[112] Karplus, M. (1959), *The Journal of Chemical Physics* **30**, 11-15.
[113] Hu, X., Carmichael, I., and Serianni, A. S. (2010), *J Org Chem* **75**, 4899-4910.
[114] Williamson, M. P., Havel, T. F., and Wuthrich, K. (1985), *J Mol Biol* **182**, 295-315.
[115] Riek, R., Hornemann, S., Wider, G., Glockshuber, R., and Wuthrich, K. (1997), *FEBS Lett* **413**, 282-288.
[116] Combettes, L. E., Clausen-Thue, P., King, M. A., Odell, B., Thompson, A. L., Gouverneur, V., and Claridge, T. D. (2012), *Chemistry* **18**, 13133-13141.
[117] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008), *Nucleic Acids Res* **36**, D402-408.
[118] Jansson, P. E., Stenutz, R., and Widmalm, G. (2006), *Carbohydr Res* **341**, 1003-1010.
[119] Mayer, M., and Meyer, B. (1999), *Angew Chem Int Ed Engl* **38**, 1784-1788.
[120] Mayer, M., and Meyer, B. (2001), *Journal of the American Chemical Society* **123**, 6108-6117.
[121] Angulo, J., Enriquez-Navas, P. M., and Nieto, P. M. (2010), *Chemistry* **16**, 7803-7812.
[122] Claasen, B., Axmann, M., Meinecke, R., and Meyer, B. (2005), *J Am Chem Soc* **127**, 916-919.
[123] Dalvit, C., Pevarello, P., Tatò, M., Veronesi, M., Vulpetti, A., and Sundström, M. (2000), *Journal of Biomolecular NMR* **18**, 65-68.
[124] Trevino, S. G., Zhang, N., Elenko, M. P., Luptak, A., and Szostak, J. W. (2011), *Proc Natl Acad Sci U S A* **108**, 13492-13497.
[125] Ludwig, C., Michiels, P. J., Wu, X., Kavanagh, K. L., Pilka, E., Jansson, A., Oppermann, U., and Gunther, U. L. (2008), *J Med Chem* **51**, 1-3.
[126] Lepre, C. A., Peng, J., Fejzo, J., Abdul-Manan, N., Pocas, J., Jacobs, M., Xie, X., and Moore, J. M. (2002), *Comb Chem High Throughput Screen* **5**, 583-590.
[127] Balaram, P., Bothner-By, A. A., and Breslow, E. (1972), *Journal of the American Chemical Society* **94**, 4017-4018.
[128] Balaram, P., Bothner-By, A. A., and Dadok, J. (1972), *Journal of the American Chemical Society* **94**, 4015-4017.
[129] Fernandez-Alonso, M. d. C., Alvaro Berbis, M., Canales, A., Arda, A., Javier Canada, F., and Jimenez-Barbero, J. (2013), In *New Applications of NMR in Drug Discovery and Development*, pp 7-42, The Royal Society of Chemistry.
[130] Manzenrieder, F., Frank, A. O., and Kessler, H. (2008), *Angew Chem Int Ed Engl* **47**, 2608-2611.
[131] Dalvit, C., and Vulpetti, A. (2011), *ChemMedChem* **6**, 104-114.
[132] Dolbier, W. R. (2009), In *Guide to Fluorine NMR for Organic Chemists*, pp 9-34, John Wiley & Sons, Inc.
[133] Diercks, T., Ribeiro, J. P., Canada, F. J., Andre, S., Jimenez-Barbero, J., and Gabius, H. J. (2009), *Chemistry* **15**, 5666-5668.
[134] Battiste, J., and Newmark, R. A. (2006), *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 1-23.
[135] Dalvit, C. (2007), *Progress in Nuclear Magnetic Resonance Spectroscopy* **51**, 243-271.
[136] Bax, A., and Lerner, L. (1986), *Science* **232**, 960-967.
[137] Nesmelova, I. V., Berbis, M. A., Miller, M. C., Canada, F. J., Andre, S., Jimenez-Barbero, J., Gabius, H. J., and Mayo, K. H. (2012), *Biomol NMR Assign* **6**, 127-129.
[138] Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997), *Proc Natl Acad Sci U S A* **94**, 12366-12371.
[139] Vogtherr, M., and Fiebig, K. (2003), *EXS*, 183-202.

[140] Sibille, N., Hanoulle, X., Bonachera, F., Verdegem, D., Landrieu, I., Wieruszeski, J. M., and Lippens, G. (2009), *J Biomol NMR* **43**, 219-227.
[141] Dalvit, C., Mongelli, N., Papeo, G., Giordano, P., Veronesi, M., Moskau, D., and Kummerle, R. (2005), *J Am Chem Soc* **127**, 13380-13385.
[142] Jackson, J. C., Hammill, J. T., and Mehl, R. A. (2007), *J Am Chem Soc* **129**, 1160-1166.
[143] Lin, D., Sze, K. H., Cui, Y., and Zhu, G. (2002), *J Biomol NMR* **23**, 317-322.
[144] Pellecchia, M., Meininger, D., Shen, A. L., Jack, R., Kasper, C. B., and Sem, D. S. (2001), *J Am Chem Soc* **123**, 4633-4634.
[145] Gardner, K. H., and Kay, L. E. (1998), *Annu Rev Biophys Biomol Struct* **27**, 357-406.
[146] LeMaster, D. M. (1990), *Annu Rev Biophys Biophys Chem* **19**, 243-266.
[147] Bax, A., and Ikura, M. (1991), *J Biomol NMR* **1**, 99-104.
[148] Wittekind, M., and Mueller, L. (1993), *Journal of Magnetic Resonance, Series B* **101**, 201-205.
[149] Grzesiek, S., and Bax, A. (1992), *Journal of Magnetic Resonance (1969)* **96**, 432-440.
[150] Bax, A., Clore, G. M., and Gronenborn, A. M. (1990), *Journal of Magnetic Resonance (1969)* **88**, 425-431.
[151] Jain, N. U., Venot, A., Umemoto, K., Leffler, H., and Prestegard, J. H. (2001), *Protein Sci* **10**, 2393-2400.
[152] Azurmendi, H. F., Martin-Pastor, M., and Bush, C. A. (2002), *Biopolymers* **63**, 89-98.
[153] Martin-Pastor, M., and Bush, C. A. (2000), *Carbohydr Res* **323**, 147-155.
[154] Tian, F., Al-Hashimi, H. M., Craighead, J. L., and Prestegard, J. H. (2000), *Journal of the American Chemical Society* **123**, 485-492.
[155] Canales, A., Jimenez-Barbero, J., and Martin-Pastor, M. (2012), *Magn Reson Chem* **50 Suppl 1**, S80-85.
[156] Martin-Pastor, M., Canales, A., Corzana, F., Asensio, J. L., and Jimenez-Barbero, J. (2005), *J Am Chem Soc* **127**, 3589-3595.
[157] Kummerlowe, G., Kiran, M. U., and Luy, B. (2009), *Chemistry* **15**, 12192-12195.
[158] Annila, A., and Permi, P. (2004), *Concepts in Magnetic Resonance Part A* **23A**, 22-37.
[159] Tian, F., Bolon, P. J., and Prestegard, J. H. (1999), *Journal of the American Chemical Society* **121**, 7712-7713.
[160] Martin-Pastor, M., Canales-Mayordomo, A., and Jimenez-Barbero, J. (2003), *J Biomol NMR* **26**, 345-353.
[161] Erdelyi, M., d'Auvergne, E., Navarro-Vazquez, A., Leonov, A., and Griesinger, C. (2011), *Chemistry* **17**, 9368-9376.
[162] Canales, A., Mallagaray, A., Perez-Castells, J., Boos, I., Unverzagt, C., Andre, S., Gabius, H. J., Canada, F. J., and Jimenez-Barbero, J. (2013), *Angew Chem Int Ed Engl* **52**, 13789-13793.
[163] Zhang, Y., Yamamoto, S., Yamaguchi, T., and Kato, K. (2012), *Molecules* **17**, 6658-6671.
[164] Zhuang, T., Leffler, H., and Prestegard, J. H. (2006), *Protein Sci* **15**, 1780-1790.
[165] Zhuang, T., Lee, H. S., Imperiali, B., and Prestegard, J. H. (2008), *Protein Sci* **17**, 1220-1231.
[166] Erickson, H. P. (2009), *Biol Proced Online* **11**, 32-51.
[167] Groves, P., Palczewska, M., Molero, M. D., Batta, G., Canada, F. J., and Jimenez-Barbero, J. (2004), *Anal Biochem* **331**, 395-397.
[168] French, A. D., and Johnson, G. P. (2011), *Methods Mol Biol* **715**, 21-42.
[169] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008), *J Comput Chem* **29**, 622-655.
[170] Imberty, A., and Perez, S. (2000), *Chem Rev* **100**, 4567-4588.
[171] Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996), *J Mol Recognit* **9**, 1-5.
[172] Moseley, H. N., Curto, E. V., and Krishna, N. R. (1995), *J Magn Reson B* **108**, 243-261.
[173] Jayalakshmi, V., and Krishna, N. R. (2002), *J Magn Reson* **155**, 106-118.

[174] Gomez, H., Polyak, I., Thiel, W., Lluch, J. M., and Masgrau, L. (2012), *J Am Chem Soc* **134**, 4743-4752.
[175] Lammerts van Bueren, A., Ardevol, A., Fayers-Kerr, J., Luo, B., Zhang, Y., Sollogoub, M., Bleriot, Y., Rovira, C., and Davies, G. J. (2010), *J Am Chem Soc* **132**, 1804-1806.
[176] Gomez, H., Lluch, J. M., and Masgrau, L. (2013), *J Am Chem Soc* **135**, 7053-7063.
[177] Viegas, A., Sardinha, J., Freire, F., Duarte, D. F., Carvalho, A. L., Fontes, C. M., Romao, M. J., Macedo, A. L., and Cabrita, E. J. (2013), *Biochem J* **451**, 289-300.
[178] Tsai, J., Lee, J. T., Wang, W., Zhang, J., Cho, H., Mamo, S., Bremer, R., Gillette, S., Kong, J., Haass, N. K., Sproesser, K., Li, L., Smalley, K. S., Fong, D., Zhu, Y. L., Marimuthu, A., Nguyen, H., Lam, B., Liu, J., Cheung, I., Rice, J., Suzuki, Y., Luu, C., Settachatgul, C., Shellooe, R., Cantwell, J., Kim, S. H., Schlessinger, J., Zhang, K. Y., West, B. L., Powell, B., Habets, G., Zhang, C., Ibrahim, P. N., Hirth, P., Artis, D. R., Herlyn, M., and Bollag, G. (2008), *Proc Natl Acad Sci U S A* **105**, 3041-3046.
[179] Schultz, J. (2008), In *Fragment-Based Drug Discovery*, pp 63-98, John Wiley & Sons, Ltd.
[180] Reitz, A. B., Smith, G. R., Tounge, B. A., and Reynolds, C. H. (2009), *Curr Top Med Chem* **9**, 1718-1724.
[181] Jahnke, W., Rudisser, S., and Zurini, M. (2001), *J Am Chem Soc* **123**, 3149-3150.
[182] Jahnke, W., Florsheimer, A., Blommers, M. J., Paris, C. G., Heim, J., Nalin, C. M., and Perez, L. B. (2003), *Curr Top Med Chem* **3**, 69-80.
[183] Li, D., DeRose, E. F., and London, R. E. (1999), *J Biomol NMR* **15**, 71-76.

CHAPTER 2

THE RECOGNITION OF SUGAR DETERMINANTS BY AN ASSORTMENT OF VIRAL FIBER PROTEINS

2.1 Introduction

Viruses are intracellular parasites of all forms of life, which take over the metabolic machinery of the organism they infect to make multiple copies of themselves through self-assembly. Lying at the boundary between life and the inanimate, viruses are generally considered as non-living entities, due to the lack of a metabolism and the requirement of a host cell to produce new progeny. Yet, viruses are the most abundant organisms in the biosphere [1], and are responsible for the greatest selective pressure on the rest of the organisms [2]. In addition, viruses are now widely used as tools in molecular and cellular biology [3], and seen as promising vehicles for the development of novel applications of industrial or therapeutic relevance, such as insecticides [4], antibiotics [5], or vectors for gene therapy [6].

Morphologically, viral particles are assemblies of different components. They consist of a capsid, or head, which is a protein coat that contains the viral genome and may adopt a rich variety of symmetrical shapes. The viral genome is a nucleic acid, which can be single-strained or double-strained DNA or RNA. The size of the viral genome is highly variable, ranging from only 2 genes in porcine circoviruses [7], to more than 900 in mimiviruses [8]. Optionally, the capsid may be surrounded by a lipid envelope derived from the host cell membrane. Finally, viruses possess a series of fibrous appendices, which project receptor proteins involved in the specific recognition of the host precluding infection.

Viruses are highly polymorphic organisms whose ultimate origin is a matter of controversy [9]. Several types of classifications of viruses exist, the simplest of which takes into consideration the type of cell they infect. Hence, there are animal viruses, plant viruses and bacteriophages, also called phages. In addition, viruses can be classified according to the chemical nature of their genetic material, their shape and their life cycle.

Regardless of virus type, one of the most important steps in the biology of viruses is infection, which is triggered by the binding of the virus to the surface of the host cell. This process involves the specific recognition, by viral receptor proteins, of protein or carbohydrate determinants of the host. To this end, some viruses have developed protein structures that project viral receptors away from the viral capsid surface, usually in the shape of fibrous proteins. The general organization of fiber proteins appears constant: they are trimers, with a N-terminal capsid-binding domain, a core shaft domain of a certain length, and a globular C-terminal receptor domain.

Understanding the host-virus interactions is key to get insight into the biology of infection and to explain the exquisite selectivity of viral strains toward their host species. At the same time, understanding these phenomena opens the door for their modulation. In this context, bacteriophages are regarded as potential antibacterial tools against resistant organisms [10-11], while adenoviruses have gathered significant interest as vectors for gene therapy or antitumor vectors in humans [12-13].

The work presented in this chapter has been performed in collaboration with the group of Dr. Mark J. van Raaij (CNB-CSIC). Dr. van Raaij provided a selection of viral fiber proteins of different sources, whose recognition properties toward an array of biologically relevant sugar determinants were investigated. In many cases, microarray data were available supporting the interaction of these receptors with several types of sugars. NMR studies on these systems were designed to obtain a confirmation of binding in solution, also attempting to provide the fine structural details on the interaction from the ligand point of view.

From the methodological perspective, the work presented in this chapter makes extensive use of STD as a straightforward ligand-detected NMR technique for the screening and epitope-fine dissection of sugar-protein interactions.

2.2 Phage T4 fiber protein gp37

2.2.1 Introduction

The phage T4 is a virus belonging to the family *Myoviridae*, which infects *Escherichia coli* cells. It is a relatively large virus, of approximate dimensions of 200 nm × 90 nm. Its icosahedral head hosts the double-stranded DNA genome, which encodes 289 proteins, and is connected through the tail to the base plate, from which a series of tail fibers emerge: long fibers, which are involved in reversible binding to the host, and short fibers, which non-reversibly bind to the bacterium surface (figs. 1, 2).

The first step in the infection of T4 involves non-covalent binding of gene product 37 (gp37), a trimeric fibrous protein of the long tail fibers (fig. 3), to glucosyl-α(1-3)glucose (Glc α(1-3)Glc) residues of the bacterial lipopolysaccharide (LPS) [14], and to the outer membrane protein C (OmpC) [15]. When at least six of the long fibers have interacted with their receptors, the signal is transduced to the basal plate, triggering an extension of the short fibers, which irreversibly bind to the host cell surface. Then, an internal tube of the tail penetrates into the bacterium and the viral DNA is ejected into the cell (fig. 2) [16].

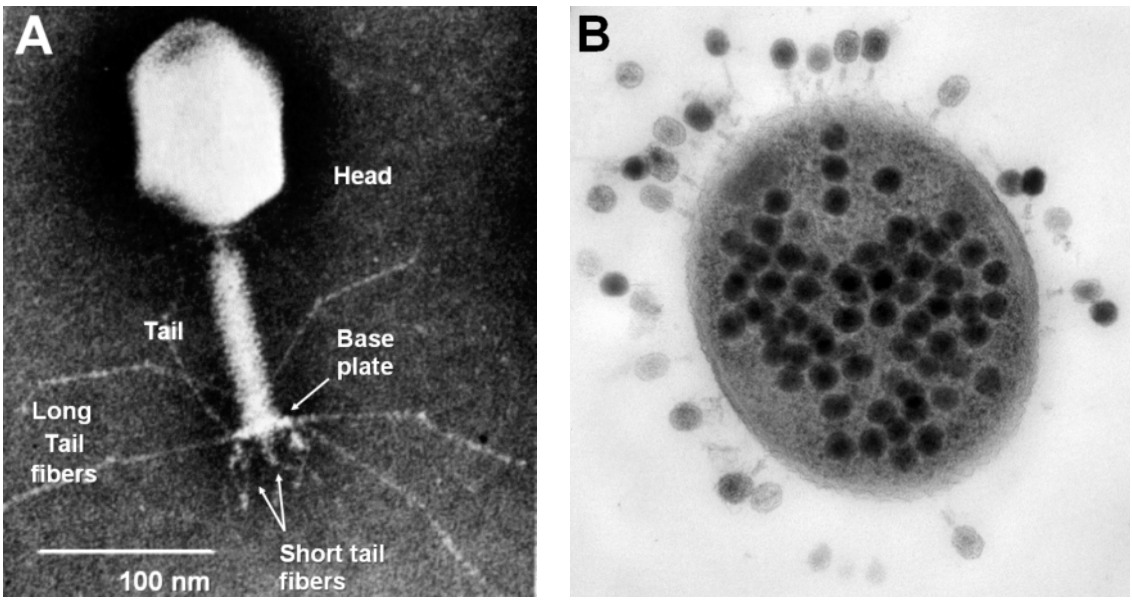


Figure 1: Electron micrographs of Phage T4. A: with labels indicating the different phage components. B: Thin section of T4 phages infecting a microcolony of *E. coli* K12. Courtesy of John Wertz (Yale University)

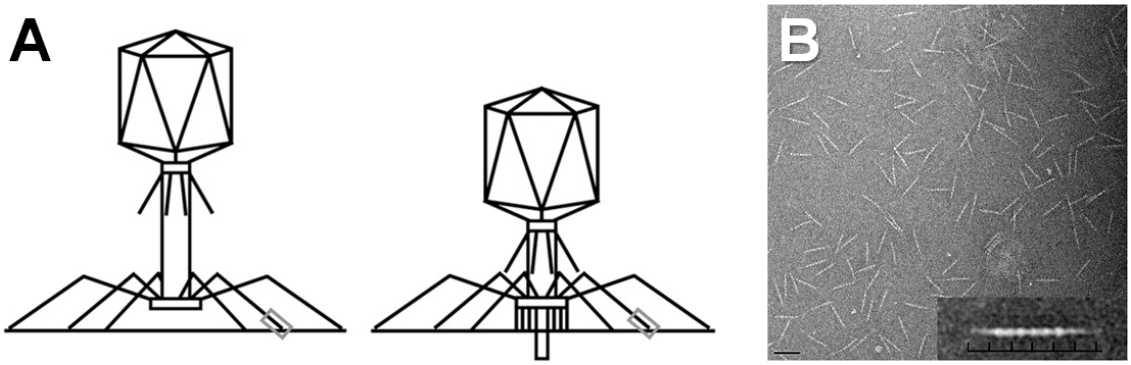


Figure 2: The gp37 tip domain. A: Schematic representations of bacteriophage T4 attached to a bacterial membrane before (left) and after (right) expansion of the short tail fibers. The gp37 tip domain is boxed. Adapted from [17]. B: Electron micrograph of recombinantly expressed gp37 tip domains. Bar size: 50 nm. Courtesy of M. van Raaij (CNB-CSIC).

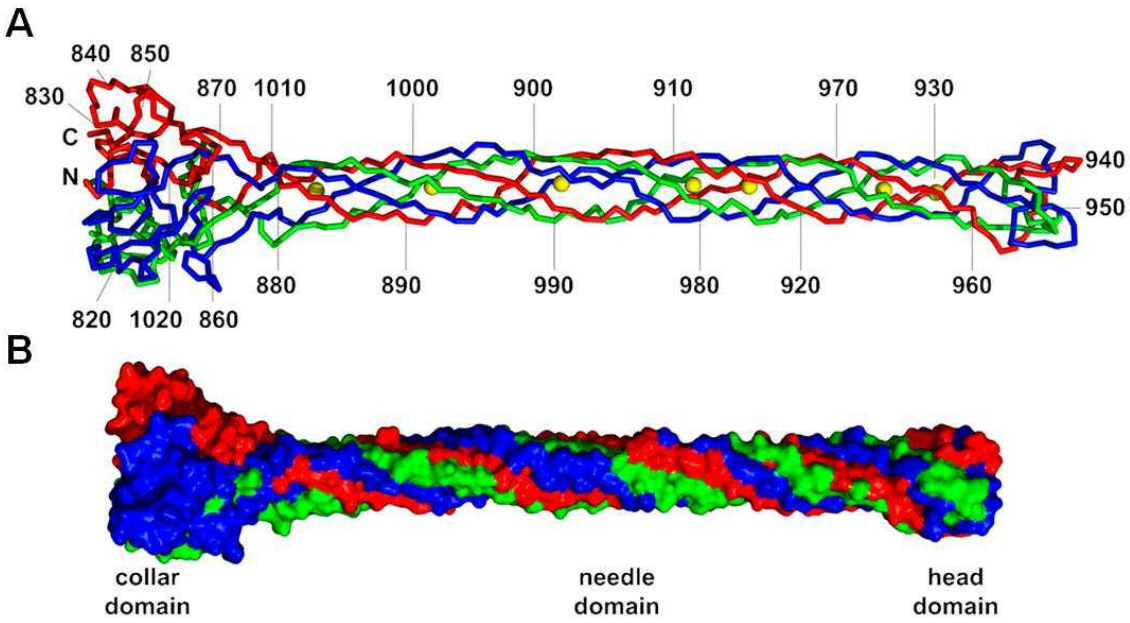


Figure 3: Structure of the gp37 trimer, with each subunit shown in a different color. A: Ribbon representation. Each 10th residue is highlighted. Iron ions are shown as yellow balls. B: Surface representation, with labels showing the collar domain, the core needle domain and the head domain, which includes the receptor motif for LPS binding. Adapted from [17].

2.2.2 Results

In order to demonstrate and characterize the recognition of Glcα(1-3)Glc units by gp37, STD experiments were performed on mixtures of this protein and the simple disaccharide model nigerose (3-*O*-α-D-glucopyranosyl-D-glucopyranose).

STD peaks were clearly observed in the corresponding spectrum, indicating that the disaccharide binds to the protein (fig. 4A). Inspection of the STD spectrum revealed differences in the degree of saturation between the two glucose residues. Specifically, most of the protons at the reducing glucose unit showed considerably smaller STD responses than their counterparts at the non-reducing moiety, suggesting that binding of nigerose to gp37 occurs mainly at the level of the non-reducing glucose. The highest degree of saturation was registered at H4 of the non-reducing end, suggesting that the recognition of nigerose by gp37 occurs mainly around this area. Finally, no significant differences were found between the two anomeric species of nigerose.

These observations are in agreement with previous data suggesting that the initial recognition of *E. coli* cells by phage T4 involve Glcα(1-3)Glc motifs in the *E. coli* lipopolysaccharide [14].

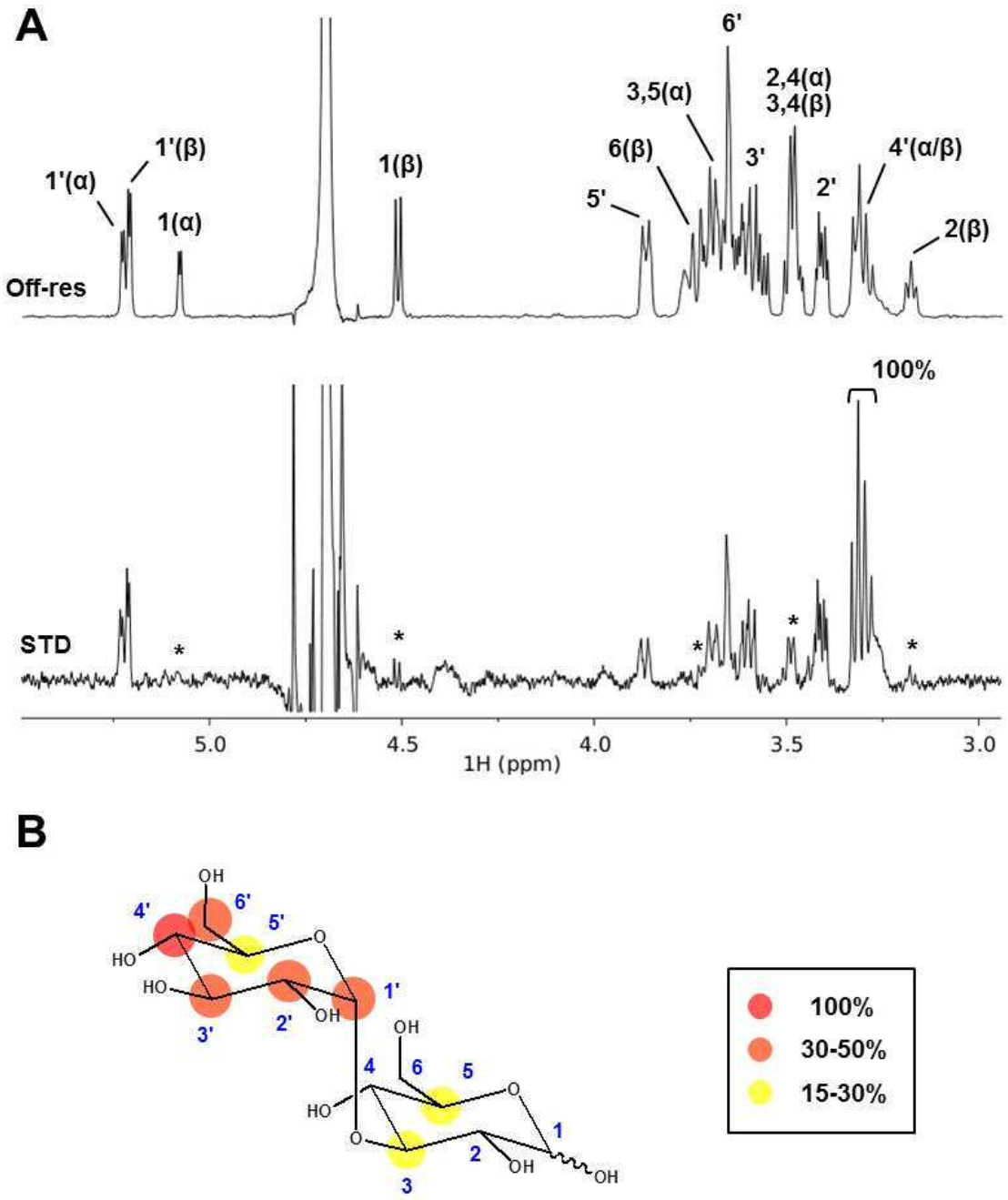


Figure 4: A: STD experiment performed on nigerose in the presence of gp37. Top spectrum: Off-resonance (reference) spectrum, with labels indicating the assignment for a number of representative signals [18-19] . Bottom spectrum: STD spectrum (up-scaled), with a label ("100%") indicating the signals that display the highest degree of saturation (4-protons of the non-reducing unit). The asterisks show the signals with the lowest STD responses (those belonging to the reducing unit). B: Epitope mapping obtained with the STD data for nigerose. Chemical structure of nigerose, with labels indicating the notation for each hydrogen atom, and colors representing the STD intensities of each signal, relative to the STD intensities of the most saturated signal. Red circle: 100%, orange circles: 50% > *I* > 30%, yellow circles: 30% > *I* > 15%.

2.3 Phage T7 fiber protein gp17

2.3.1 Introduction

The phage T7 is a virus belonging to the family *Podoviridae*. T7 infects rough strains of *Escherichia coli*, i.e. those lacking the O-antigen chain in their LPS. Its icosahedral capsid, made out of gp10, contains the DNA genome, which codes for 55 proteins. The capsid is linked at one of its vertices, through a connector, to a tail encompassing six tail fibers, each composed of three copies of the fibrous protein gp17 (fig. 5).

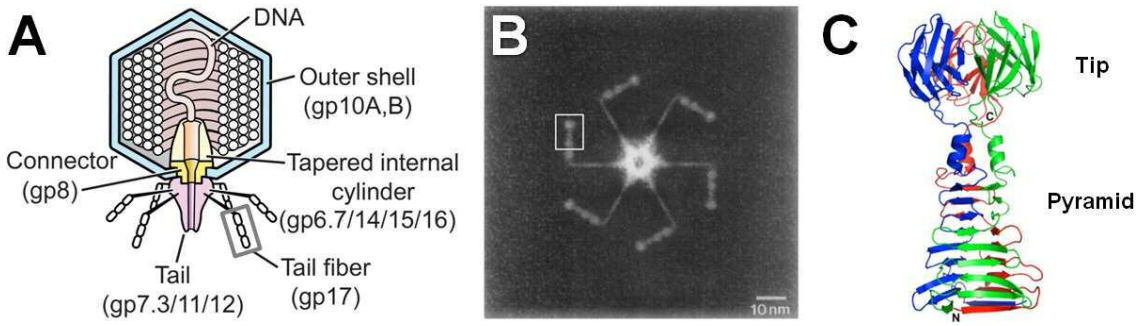


Figure 5: Phage T7 fiber gp17 protein. A: Scheme of T7, with gp17 fibers boxed. Adapted from [20]. B: Negatively stained electron micrograph of the tail complex of phage T7, showing six tail fibers. The boxed region corresponds to the gp17(371-553) fragment used in this chapter. The tip and pyramid domains are distinguishable. C: Structure of the gp17(371-553) fragment, showing its tip and pyramid domains (PDB: 4A0T). Adapted from [21].

Gp17 forms elongated homotrimers (fig. 5C), which are involved in the first specific, albeit reversible, attachment to the *E. coli* LPS [22]. A second, irreversible interaction with the bacterial membrane, followed by inoculation of the viral genetic material is likely mediated by binding of one or more of the tail proteins to an unknown receptor [23].

Each gp17 subunit consists of a N-terminal tail-attachment domain, a core shaft region and a C-terminal domain [21]. The structure of the most C-terminal portion of gp17, i.e. gp17(371-553) was elucidated by the group of M. van Raaij using X-ray crystallography [21]. This distal part of the fiber, which is likely involved in virus-host recognition, encompasses two distinct structural domains, which are also apparent under the electron microscope (fig. 5B): an N-

terminal “pyramid” domain, formed by a mixed β-sheet made up by β-strands of the three different subunits, and a C-terminal globular “tip” domain (fig. 5C).

The most important part of the LPS for binding to T7 fibers has been suggested to be the first glucose and the terminal glucose/heptose of the LPS outer core [24]. Fittingly, microarray data obtained by the group of M. van Raaij suggested binding between the gp17(371-553) fragment to kojibiose (2-*O*-α-D-glucopyranosyl-D-glucopyranose, (Glcα(1-2)Glc), a motif which is present in the outer core of the *E. coli* LPS. This fact prompted the NMR study of the interaction, in solution, between gp17 and the aforementioned disaccharide.

2.3.2. Results

STD experiments performed on kojibiose in the presence of the gp17(371-355) fragment gave rise to peaks in the corresponding spectrum, indicating that the disaccharide binds to the protein (fig. 6). Inspection of the STD spectrum revealed subtle differences in the degree of saturation between the two glucose residues, as well as between the two different species of kojibiose that coexist in solution, due to the α/β mutarrotation process at the reducing glucose. Specifically, several protons at the reducing glucose showed smaller STD responses than their counterparts at the non-reducing moiety. These differences became especially apparent for the β-anomer (fig. 6, fig. 7). These results suggest that binding of kojibiose to the gp17.371 fragment occurs mainly at the level of the non-reducing glucose, and that the β-anomer of the reducing unit is in looser contact to the protein surface than the α-anomer. However, the differences in the degree of saturation between the two glucose units were only minor (for comparison, significantly larger differences between the glucose monomers were obtained with the gp37/nigerose system, see fig. 4 above). It is thus likely that both sugar units of kojibiose are accommodated in the gp17 binding site.

To complement these observations, we also explored the ability of gp17(371-533) to recognize the simple monosaccharide, glucose. STD signals were obtained, although they were significantly less intense than those obtained for kojibiose (fig. 6). These results point towards considerable differences in affinity between the monosaccharide and kojibiose and support the hypothesis that both glucose units are accommodated in the binding site.

In rough *E. coli* strains, the LPS outer core bears heptose-α(1-2)glucose-α(1-2)glucose-α(1-3)glucose termini. [25]. In the past, the terminal glucose residues have been shown to be important for the T7/LPS interaction [24]. Our results indicate that gp17 also binds to the Glcα(1-2)Glc disaccharide. Therefore, more structural details on the actual epitope that is

recognized by the virus fibers have been presented herein. Further experiments with elongated ligands incorporating the heptose unit should clarify whether the terminal heptose is also involved in the recognition by gp17.

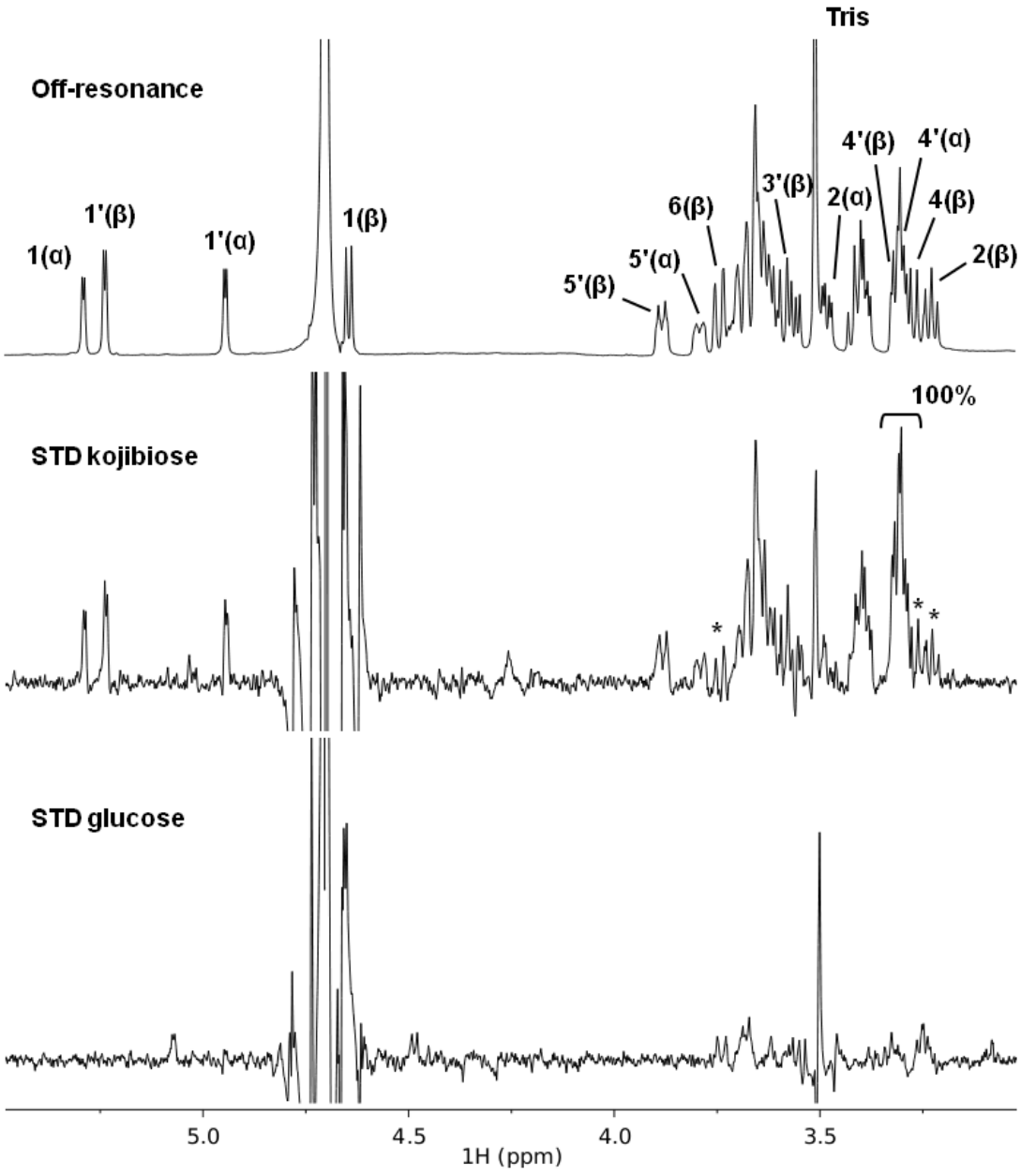


Figure 6: STD experiments performed on kojibiose and glucose in the presence of the gp17.371 fragment. Top spectrum: Off-resonance (reference) spectrum of kojibiose, with labels indicating the assignment for a number of representative signals [18-19]. Middle spectrum: STD spectrum of kojibiose (up-scaled 100X), with a label ("100%") indicating the signals showing the highest degree of saturation (H4-protons of the non-reducing unit), and asterisks indicating the signals with the lowest STD responses (those belonging to the β-anomer of the reducing unit). Bottom spectrum: STD spectrum of glucose (up-scaled 100X).

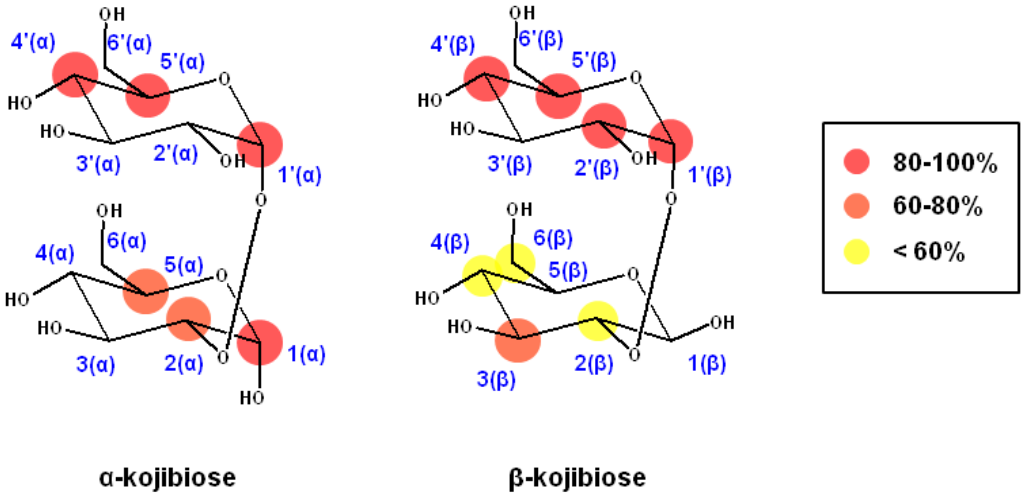


Figure 7: Epitope mapping obtained with the STD data for kojibiose. Chemical structures of kojibiose with the reducing end in α- (left) and β-configuration (right), with labels indicating the notation for each hydrogen atom, and colors showing the STD intensities of each signal, relative to the STD intensities of the most saturated signals. Red circles: 100% > I > 80%, orange circles: 80% > I > 60%, yellow circles: I < 60% (for uncircled protons the corresponding relative STD values could not be calculated due to signal overlapping).

2.4 Turkey adenovirus 3 fiber head protein

2.4.1 Introduction

Adenoviruses are a family of animal viruses with a large, non-enveloped icosahedral capsid and a double-stranded DNA genome. Thanks to its efficient infection and high loading capacity, adenoviruses are regarded as promising tools for gene therapy [13].

The adenovirus capsid consists of three major structural proteins: the hexon, the penton base and the fiber. The hexon is the major coat protein forming the icosahedral structure. The penton is bound to the hexon protein at each vertex of the icosahedron, serving as the base for the insertion of the fiber (fig. 8) [26]. The specific attachment of adenoviruses to their host cells occurs with high affinity and at the level of the fiber knob [27], whose primary receptor in humans is the widely expressed coxsackievirus and adenovirus receptor (CAR) [28]. Other subtype-specific receptors for adenovirus reported in humans include CD46 [29] as well as the sialic acids presented by gangliosides and glycoproteins [30].

In addition, most human adenoviruses bear the tripeptide sequence Arg-Gly-Asp (RGD) in the fiber knob, which has been reported as the binding site of the virus to cellular integrins, causing endocytosis of the virus by the cell [31].

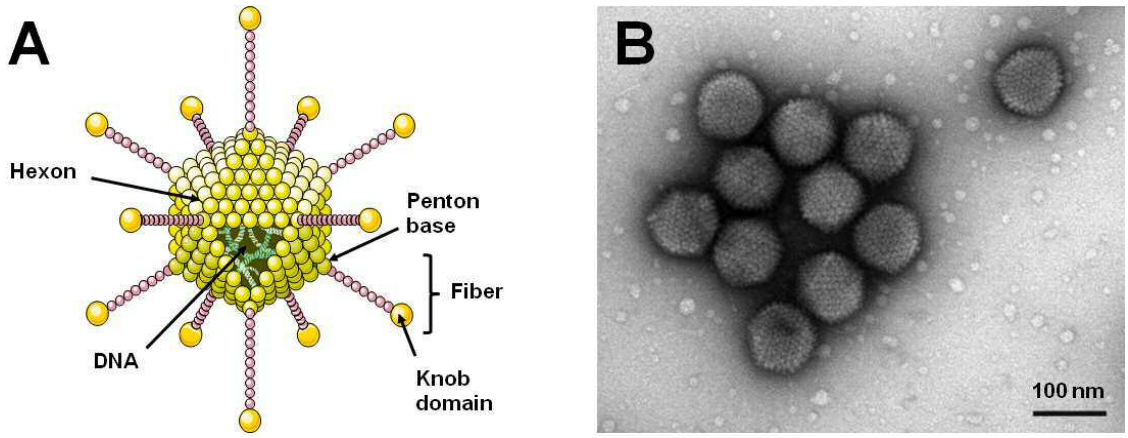


Figure 8: Adenovirus structure. A: Scheme showing the three main protein components of the adenoviral capsule: the hexon, the penton base and the fiber. B: Negatively stained electron micrograph of adenovirus particles isolated from *Pipistrellus pipistrellus* bat. Scale bar: 100 nm. From [32].

Turkey adenovirus 3 (TAdV-3) belongs to the genus *Siadenovirus*. With the smallest genomes in the family, the taxon is named after a genus-specific open-reading frame, whose putative gene product bears a high sequence similarity to bacterial sialidases [33].

TadV-3, which carries a small genome coding for 16 genes [34], causes hemorrhagic enteritis in turkeys and marble spleen disease in pheasants [35]. The general organization of the virus is highly similar to that of other members of the adenovirus family. Sequencing of the TadV-3 genome revealed a high degree of sequence similarity between most of its putative gene products and their homologues throughout the family [33]. The predicted fiber protein consists of an N-terminal virus-attachment domain, a central shaft domain and a C-terminal knob (head) domain. Interestingly, the head domain has no sequence similarity to any previously known analogue [36], nor it carries the RGD sequence, as in other mammalian and avian adenoviruses [33].

The lack of sequence homologues raises the intriguing possibility that the TadV-3 fiber head domain might bear novel recognition capabilities. In any case, the identification of the receptor for this protein is interesting either from the perspective of a novel sequence leading to a novel activity, or to convergent recognition properties. In this regard, microarray data obtained by Dr. van Raaij suggested binding to both 3'- and 6'-sialyllactose, as well as to xylose, for the fiber head domain of TAdV-3. NMR-based test of binding to these three sugars is presented in this section.

The structure of the head domain has recently been solved by the group of M. van Raaij (PDB: 3ZPE). The structure shows a trimeric arrangement, in which each monomer consists of a β-sandwich, plus a protruding β-hairpin akin to a "flap" embracing the neighboring monomer (fig. 9). Of note, structure-based searches could not correlate the structure of TadV-3 fiber head protein with any previously known structure.

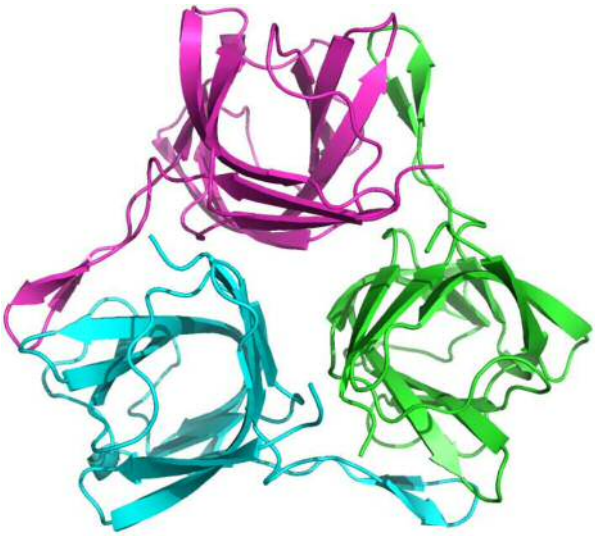


Figure 9: Structure of the turkey adenovirus 3 fiber C-terminal head domain (PDB: 3ZPE), as solved by the group of M. van Raaij. The structure adopts a trimeric arrangement (each monomer is shown in a different color), in which each monomer consist of a β-sandwich, plus a protruding β-hairpin which embraces the neighboring monomer.

In addition to the virulent strains of TadV-3, naturally occurring avirulent strains are known, producing splenomegaly and immunosupression in turkeys, although not resulting in mortality [37]. Intriguingly, the sequence identity of both virulent and avirulent strains is very high, with only a few non-silent point mutations found in genes *ORF1*, *E3* and *fib*, i.e. the one coding for the fiber. Different mutations appearing in avirulent strains have been suggested to result in alteration of cell tropism or affinity to receptors [37]. To test this hypothesis, we have investigated herein the effect of two point mutations in the fiber head domain (^{354}M→I and ^{376}M→T) found in a naturally occurring avirulent strain [37], by studying the recognition abilities of both the virulent and the avirulent versions of the fiber head protein toward sialic acid-containing oligosaccharides.

2.4.2 Results

STD peaks were observed for 3'-sialyllactose and 6'-sialyllactose in the presence of either the virulent or the avirulent versions of TAdV, indicating that both ligands are recognized by these

proteins in solution (figs. 9-10). In contrast, the more intriguing interaction with xylose, suggested by the microarray data, could not be detected by NMR in solution (not shown).

In all instances, the highest degree of saturation was observed at the sialic acid signals, and to a much lesser degree, at the galactose signals (fig 11). Very similar degrees of saturation were found between both virulent and avirulent proteins, as well as between both ligands.

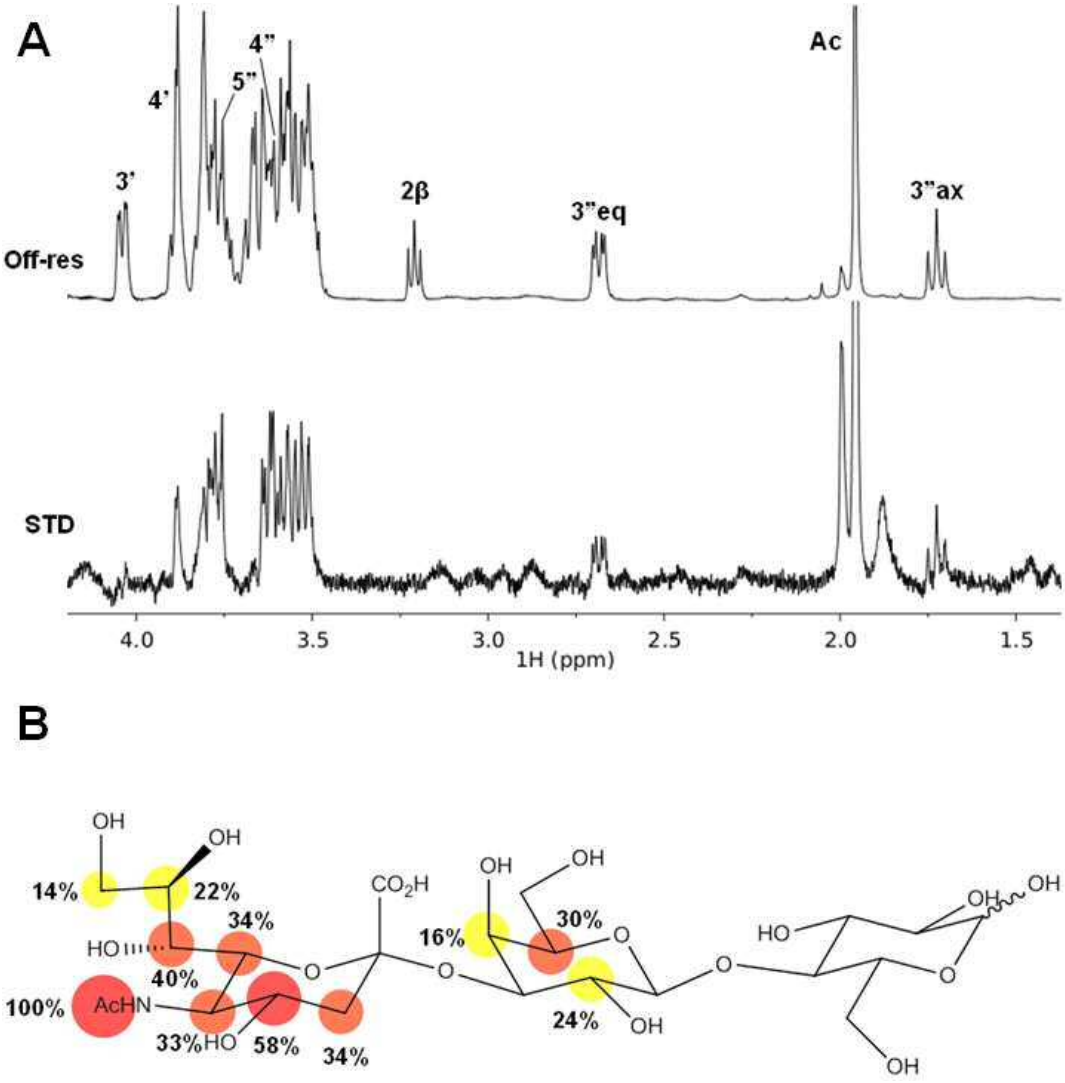


Figure 10: STD experiment performed on 3'-sialyllactose in the presence of the virulent version of TAdV. Top: Off-resonance (reference) spectrum, with labels indicating the assignment for a number of representative ligand signals. Bottom: STD spectrum (up-scaled 100X). B: Epitope mapping deduced from the STD data. Chemical structure of 3'-sialyllactose, with labels indicating the STD intensity for each signal, relative to the STD intensity for the 5''-N-acetyl peak. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%.

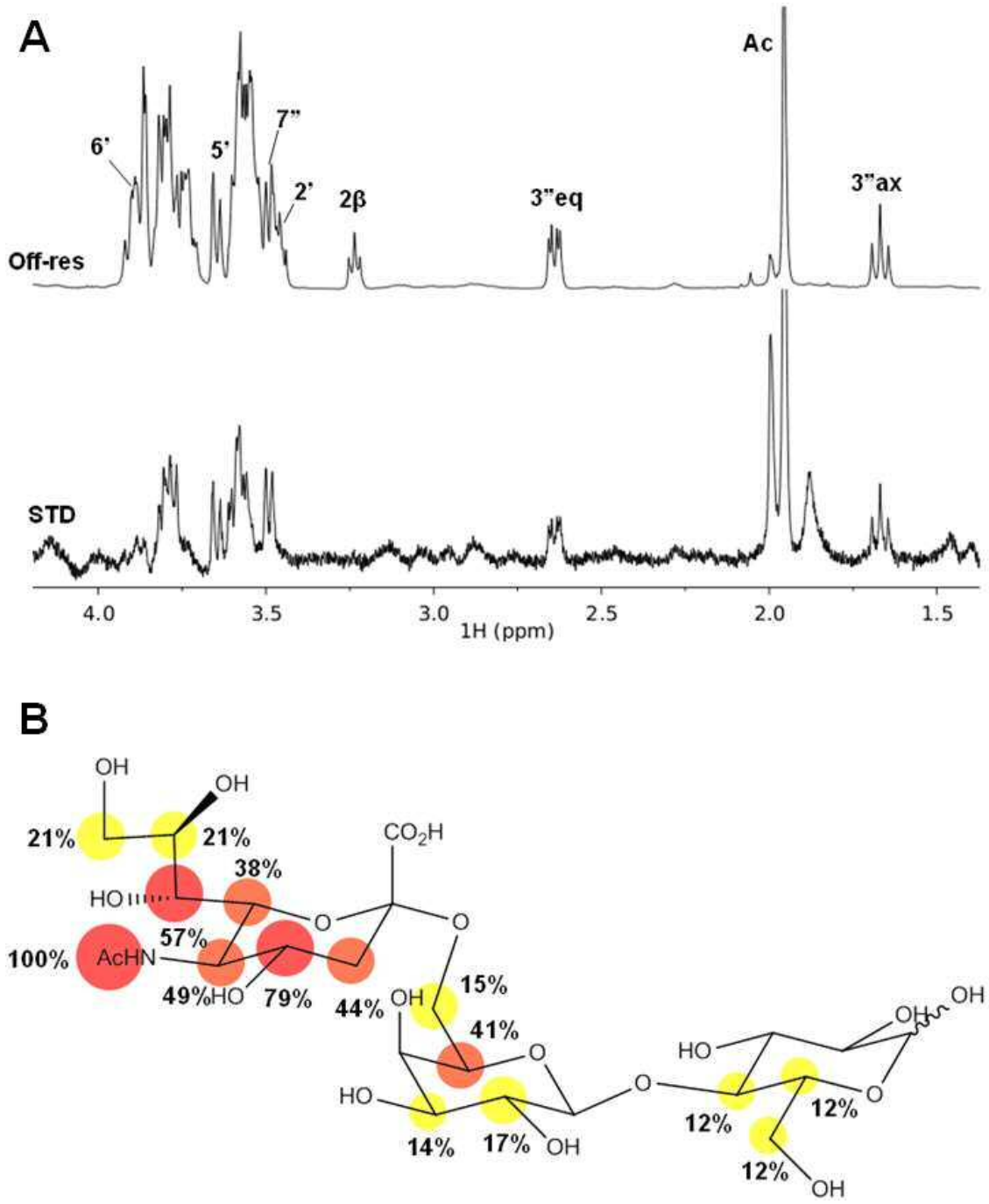


Figure 11: STD experiment performed on 6'-sialyllactose in the presence of the virulent version of TAdV. Top: Off-resonance (reference) spectrum, with labels indicating the assignment for a number of representative ligand signals. Bottom: STD spectrum (up-scaled 100X). B: Epitope mapping deduced from the STD data. Chemical structure of 6'-sialyllactose, with labels indicating the STD intensity for each signal, relative to the STD intensity for the 5''-N-acetyl peak. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%.

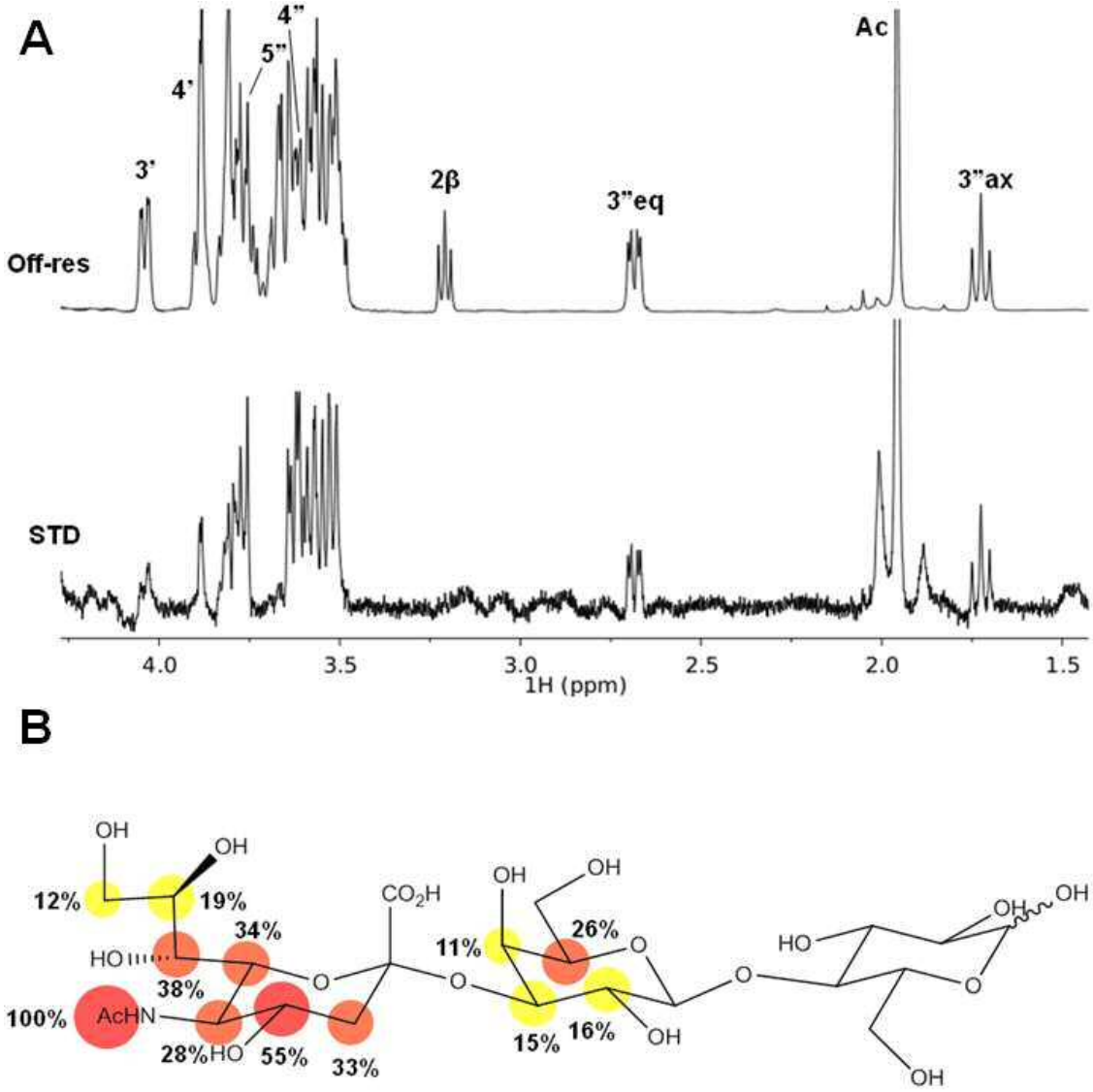


Figure 12: STD experiment performed on 3'-sialyllactose in the presence of the avirulent version of TAdV. Top: Off-resonance (reference) spectrum, with labels indicating the assignment for a number of representative ligand signals. Bottom: STD spectrum (up-scaled 100X). B: Epitope mapping deduced from the STD data. Chemical structure of 3'-sialyllactose, with labels indicating the STD intensity for each signal, relative to the STD intensity for the 5''-N-acetyl peak. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%.

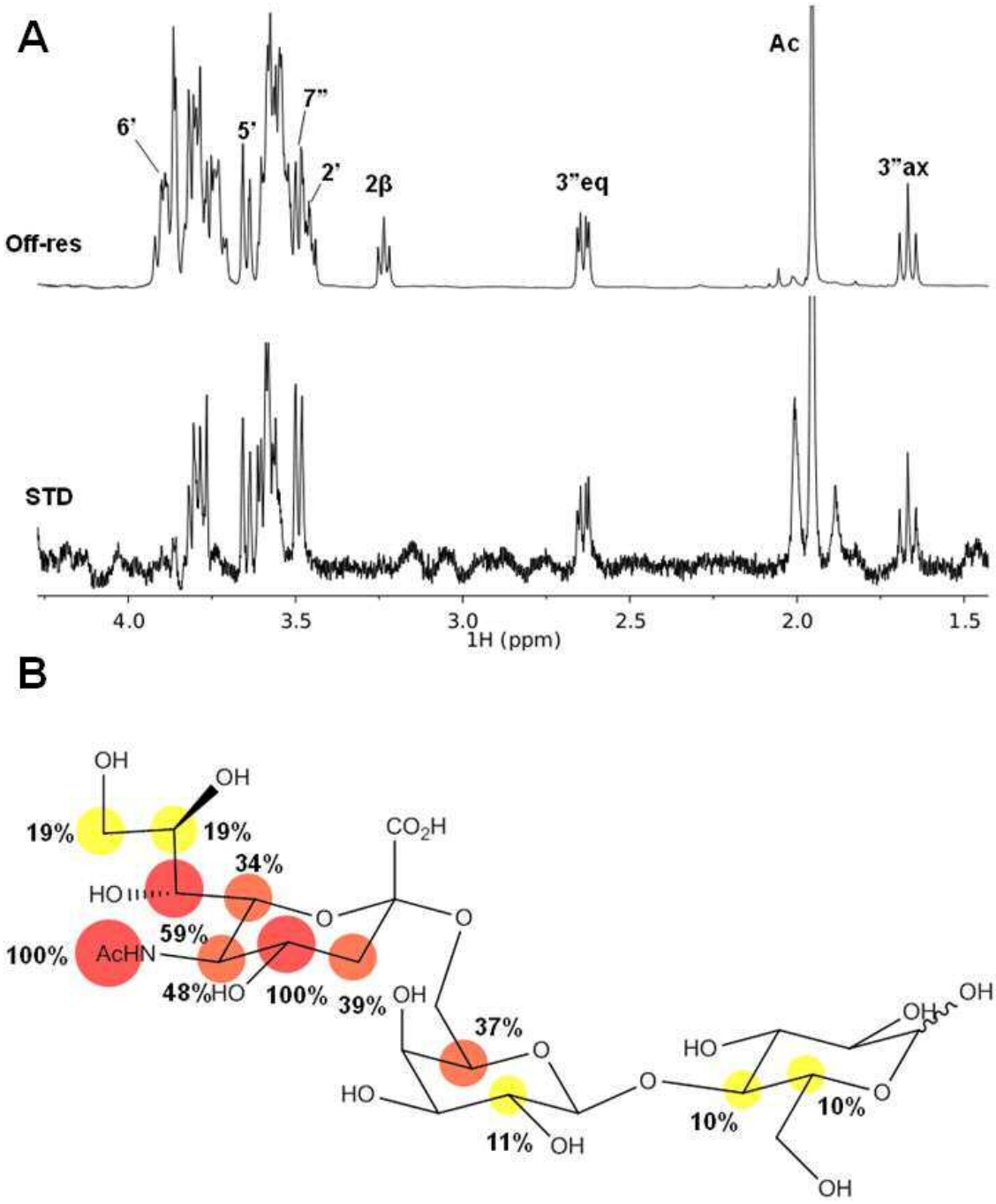


Figure 13: STD experiment performed on 6'-sialyllactose in the presence of the avirulent version of TAdV. Top: Off-resonance (reference) spectrum, with labels indicating the assignment for a number of representative ligand signals. Bottom: STD spectrum (up-scaled 100X). B: Epitope mapping deduced from the STD data. Chemical structure of 6'-sialyllactose, with labels indicating the STD intensity for each signal, relative to the STD intensity for the 5''-N-acetyl peak. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%.

2.4.2.1: Avirulence may be caused by an impaired trimer formation

The virulent and the avirulent versions of the fiber head protein did not show any significant difference in the recognition processes of different sialyllactose derivatives. Provided that sialic acid is, in fact, the receptor for this protein, this observation rules out the possibility that avirulence is caused by a defect in the recognition of the receptor. Thus, we further investigated what the causes for the difference in avirulence phenotype for both protein variants might be. As depicted in Fig. 14, one of the mutated residues (^{354}M→I) lies at the interface between each pair of trimer subunits, within the beta hairpin moiety that embraces the neighboring monomer. We thus raised the question whether the ^{354}M→I substitution might alter the stability of the trimer. DOSY (diffusion-ordered NMR spectroscopy) experiments performed on both protein versions revealed the existence of a smaller apparent size for the avirulent protein, in a concentration-dependent manner. Extrapolation of the obtained *D* values to a calibration line (fig. S1 in Appendix) permitted to estimate apparent molecular weights of 20-45 kDa for samples of the avirulent protein version at concentrations ranging from 3 µM to 40 µM , in contrast with the inferred size of 64 kDa for the virulent version. Thus, these data support that the loss of virulence in the avirulent protein might, at least in part, be explained by an impeded trimer formation.

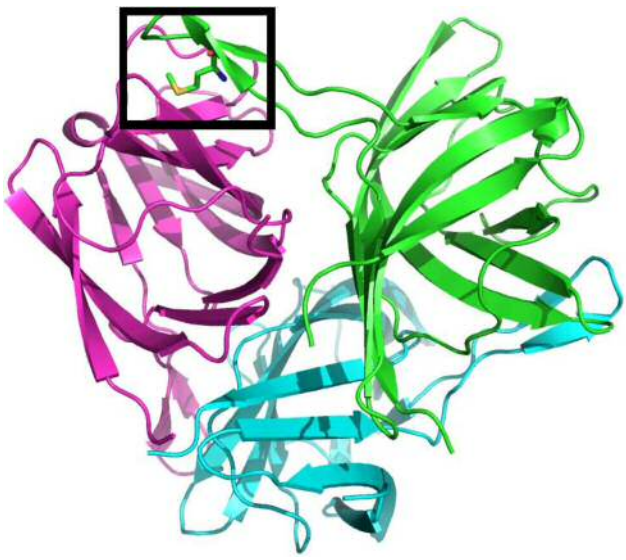


Figure 14: Structure of the turkey adenovirus 3 fiber C-terminal head domain (PDB: 3ZPE), as solved by the group of M. van Raaij, highlighting the position of M354 in the virulent protein (in sticks, in the green monomer). This amino acid, involved in monomer-monomer interactions, is substituted by an isoleucine residue in the avirulent protein.

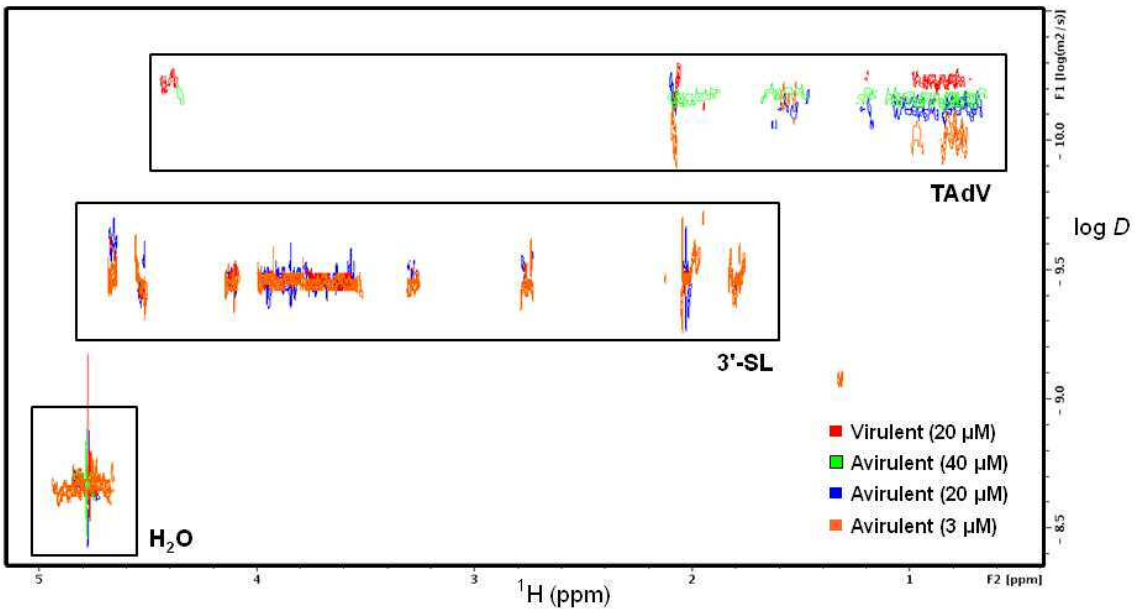


Figure 15: DOSY titration with different concentrations of the avirulent versión of the TadV fiber head protein. Stacked DOSY spectra of either the virulent protein (red), or different concentrations of the avirulent protein (green: 40 µM, blue: 20 µM, and orange: 3 µM). Signals of water and 3'-sialyllactose are indicated for reference.

2.5 Methods

STD spectra were acquired at 298 K in a Bruker AVANCE 600 MHz spectrometer equipped with a cryoprobe. STD experiments were performed in fully deuterated phosphate 20 mM buffer, pD 7.8. Samples contained 2 mM of the screened sugar compounds, and 15 to 30 µM of the correspondent viral fiber proteins. Proteins were provided by M. van Raaij (CNB-CSIC): TadV-3 fiber head protein [36], phage T7 gp17 (371-553) fragment [21] and phage T4 g37 [17] were recombinantly produced as previously reported.

The protein signals were selectively saturated on-resonance at 0.5 ppm with a train of Gaussian-shaped pulses of 50 ms each, totaling an irradiation time of 2 s. The off-resonance frequency was set at 100 ppm. A T_2 relaxation filter consisting of a 15 ms 5 kHz spin-lock was used to reduce the protein background signals. STD spectra were obtained by subtracting the on-resonance from the off-resonance spectrum. STD intensities were measured comparing each STD spectrum with the correspondent off-resonance one and normalizing to the ligand peak receiving the highest degree of saturation. The resonances of 3'-sialyllactose or 6'-

sialyllactose were assigned by standard procedures (^{1}H-^{1}H TOCSY 70 ms, ^{1}H-^{1}H ROESY 600 ms and phase-sensitive ^{1}H-^{13}C HSQC), whereas those of nigerose and kojibiose were obtained using the reported assignments [19]. Data were analyzed with Bruker TopSpin 3.0 and figures were built with MestReNova v.8.0.2.

DOSY spectra of TAdV-3 fiber head proteins were recorded in a Bruker AVANCE 500 MHz spectrometer equipped with a 5 mm inverse probe head, with a diffusion time of 0.5 s and a gradient pulse of 1.5 ms, 32 gradient intensity increments and 32 scans per increment. Spectra were processed with the standard 2D DOSY procedure included in TOPSPIN 2.1 software (Bruker Biospin).

2.6 References

[1] Clokie, M. R., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011), *Bacteriophage* **1**, 31-45.

[2] Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2005), *Curr Opin Struct Biol* **15**, 655-663.

[3] Henry, M., and Debarbieux, L. (2012), *Virology* **434**, 151-161.

[4] Miller, L. K., Lingg, A. J., and Bulla, L. A., Jr. (1983), *Science* **219**, 715-721.

[5] Matsuzaki, S., Rashel, M., Uchiyama, J., Sakurai, S., Ujihara, T., Kuroda, M., Ikeuchi, M., Tani, T., Fujieda, M., Wakiguchi, H., and Imai, S. (2005), *J Infect Chemother* **11**, 211-219.

[6] Mulligan, R. C. (1993), *Science* **260**, 926-932.

[7] Allan, G., Krakowka, S., Ellis, J., and Charreyre, C. (2012), *Virus Res* **164**, 4-9.

[8] Claverie, J. M., Abergel, C., and Ogata, H. (2009), *Curr Top Microbiol Immunol* **328**, 89-121.

[9] Abrescia, N. G., Bamford, D. H., Grimes, J. M., and Stuart, D. I. (2012), *Annu Rev Biochem* **81**, 795-822.

[10] Summers, W. C. (2001), *Annu Rev Microbiol* **55**, 437-451.

[11] Meng, X., Shi, Y., Ji, W., Zhang, J., Wang, H., Lu, C., Sun, J., and Yan, Y. (2011), *Appl Environ Microbiol* **77**, 8272-8279.

[12] Stercz, B., Perlstadt, H., Nagy, K., and Ongradi, J. (2013), *Acta Microbiol Immunol Hung* **60**, 447-459.

[13] Kasala, D., Choi, J. W., Kim, S. W., and Yun, C. O. (2014), *Expert Opin Drug Deliv* **11**, 379-392.

[14] Dawes, J. (1975), *Nature* **256**, 127-128.

[15] Yu, F., and Mizushima, S. (1982), *J Bacteriol* **151**, 718-722.

[16] Leiman, P. G., Arisaka, F., van Raaij, M. J., Kostyuchenko, V. A., Aksyuk, A. A., Kanamaru, S., and Rossmann, M. G. (2010), *Virol J* **7**, 355.

[17] Bartual, S. G., Otero, J. M., Garcia-Doval, C., Llamas-Saiz, A. L., Kahn, R., Fox, G. C., and van Raaij, M. J. (2010), *Proc Natl Acad Sci U S A* **107**, 20287-20292.

[18] Roslund, M. U., Tahtinen, P., Niemitz, M., and Sjoholm, R. (2008), *Carbohydr Res* **343**, 101-112.

[19] Goffin, D., Bystricky, P., Shashkov, A. S., Lynch, M., Hanon, E., Paquot, M., and Savage, A. V. (2009), *Bulletin of the Korean Chemical Society*, 2535-2541.
[20] Serwer, P., Wright, E. T., Hakala, K. W., and Weintraub, S. T. (2008), *BMC Res Notes* **1**, 36.
[21] Garcia-Doval, C., and van Raaij, M. J. (2012), *Acta Crystallogr Sect F Struct Biol Cryst Commun* **68**, 166-171.
[22] Steven, A. C., Trus, B. L., Maizel, J. V., Unser, M., Parry, D. A., Wall, J. S., Hainfeld, J. F., and Studier, F. W. (1988), *J Mol Biol* **200**, 351-365.
[23] Hu, B., Margolin, W., Molineux, I. J., and Liu, J. (2013), *Science* **339**, 576-579.
[24] Qimron, U., Marintcheva, B., Tabor, S., and Richardson, C. C. (2006), *Proc Natl Acad Sci U S A* **103**, 19039-19044.
[25] Nikaido, H. (2003), *Microbiol Mol Biol Rev* **67**, 593-656.
[26] San Martin, C. (2012), *Viruses* **4**, 847-877.
[27] Mei, Y. F., and Wadell, G. (1996), *J Virol* **70**, 3688-3697.
[28] Howitt, J., Anderson, C. W., and Freimuth, P. (2003), *Curr Top Microbiol Immunol* **272**, 331-364.
[29] Russell, I. A., Royds, J. A., and Braithwaite, A. W. (2004), *Prog Mol Subcell Biol* **36**, 207-243.
[30] Arnberg, N. (2009), *Rev Med Virol* **19**, 165-178.
[31] Wickham, T. J., Mathias, P., Cheresh, D. A., and Nemerow, G. R. (1993), *Cell* **73**, 309-319.
[32] Sonntag, M., Muhldorfer, K., Speck, S., Wibbelt, G., and Kurth, A. (2009), *Emerg Infect Dis* **15**, 2052-2055.
[33] Davison, A. J., Wright, K. M., and Harrach, B. (2000), *J Gen Virol* **81**, 2431-2439.
[34] Pitcovski, J., Mualem, M., Rei-Koren, Z., Krispel, S., Shmueli, E., Peretz, Y., Gutter, B., Gallili, G. E., Michael, A., and Goldberg, D. (1998), *Virology* **249**, 307-315.
[35] Sharma, J. M. (1991), *Vet Immunol Immunopathol* **30**, 67-71.
[36] Singh, A. K., Ballmann, M. Z., Benko, M., Harrach, B., and van Raaij, M. J. (2013), *Acta Crystallogr Sect F Struct Biol Cryst Commun* **69**, 1135-1139.
[37] Beach, N. M., Duncan, R. B., Larsen, C. T., Meng, X. J., Sriranganathan, N., and Pierson, F. W. (2009), *J Gen Virol* **90**, 1978-1985.

CHAPTER 3

THE INTERACTION PROPERTIES OF A MONOCLONAL ANTIBODY RAISED AGAINST THE MYCOBACTERIAL CAPSULE

3.1 Introduction

Antibodies, also known as immunoglobulins, are large, Y-shaped molecules synthesized by the immune system to identify and neutralize foreign objects, such as viral or bacterial invaders. Antibodies are able to recognize their ligands, called antigens, with high affinity and specificity.

In mammals, five iso-types exist which differ in origin, functionality and quaternary arrangement. Depending on the isotype, immunoglobulins (Ig) can be secreted in monomeric (IgD, IgE, IgG), dimeric (IgA) and pentameric (IgM) forms, or they can be bound to the surface of immune cells. Structurally, antibodies consist of two identical heavy chains accompanied by two identical light chains, which are covalently attached by a disulfide bond to each other, forming a constant region, which is identical in all antibodies of the same isotype. They play a key role in modulating its immune activity. Additionally, a variable region exist, which contains the antigen binding site (fig. 1A).

Although all antibody molecules share the same basic architecture, they display remarkable variability in the antigen-binding region, which enables them to potentially bind any possible antigen. The size of the human antibody repertoire is potentially as high as 10^{11} different immunoglobulin molecules in every individual [1]. The generation of this diversity is accounted for by a complex mechanism involving the rearrangement of the immunoglobulin genes [2], for whose elucidation S. Tonegawa received the Nobel Prize for Physiology or Medicine in 1987.

The recognition of antigens by antibodies involves non-covalent, reversible interactions between the variable region of immunoglobulins and conformational epitopes of the antigen, particularly of protein or carbohydrate nature.

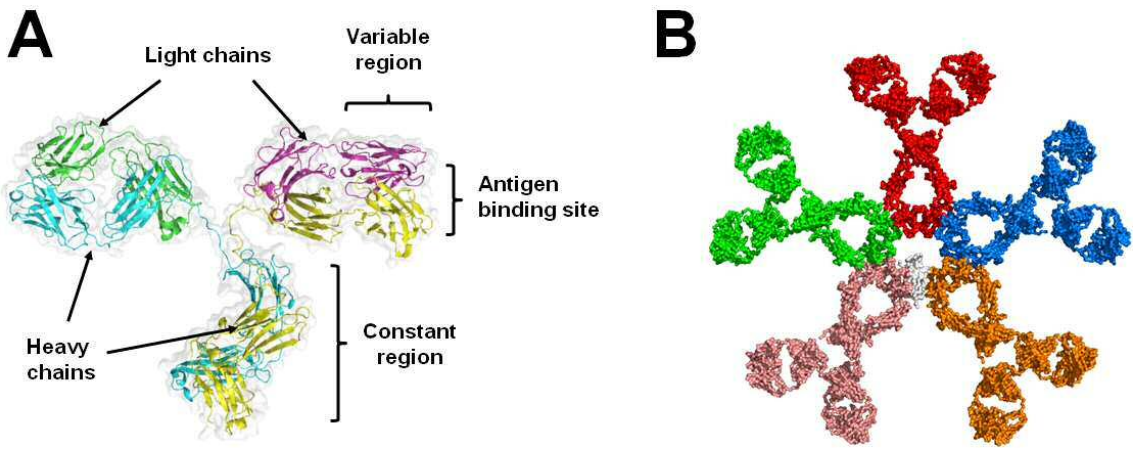


Figure 1: Structure of immunoglobulins. A: Crystallographic structure of an intact IgG antibody (PDB: 1IGT) [3], with labels showing the antibody parts. B: Solution synchrotron X-ray scattering structure of a human pentameric IgM (PDB: 2RCJ) [4].

Carbohydrate antigens are known to act as many disease-causing agents, including cancer and infections caused by viruses, bacteria and parasites [5]. In addition to being naturally recognized by the immune system, they can be exploited for stimulating the immune system in the form of vaccines, [6-7] or being subject to targeting by therapeutic antibodies [8]. Also importantly, antibodies are the base for a variety of immunoassays routinely utilized in biochemistry and molecular biology labs. Hence, understanding antibody-carbohydrate interactions is useful both for the generation of new therapeutic agents and for the development of novel analytical techniques.

The chemistry of carbohydrate-antibody interactions is essentially similar as in other protein-sugar systems, involving hydrogen bonding, van der Waals contacts and hydrophobic effects [9]. Methodologically, antibodies have properties that make them appealing from the NMR viewpoint. They are heavy proteins (> 150 kDa), and can be oligomeric, which makes them very sensitive to ligand-detected methods. In turn, however, this renders them unapproachable using protein-detected techniques. Carbohydrate-antibody interactions have long been studied by NMR methods, such as STD to detect and characterize the recognition of glycans epitopes by immunoglobins [10-11], including studies to deduce the minimum epitope recognized by an antibody [12-13], or trNOESY experiments to study the conformation of saccharides in the antibody-bound state [14].

The work presented in this chapter was done in collaboration with the group of Prof. Ben J. Appelmelk (University of Amsterdam), with the goal of developing useful antibody-based techniques for the investigation of the *Mycobacterium* capsular α-glucan, using a murine IgM antibody raised against glycogen [15].

Pathogenic mycobacteria are associated to important human diseases such as leprosy and tuberculosis (TB). *Mycobacterium tuberculosis*, the causative agent of TB, has infected one-third of the world's population and claims more than 1.4 million lives every year. Thus, insights into the molecular mechanisms that allow the tubercle bacillus to occupy the hostile environment of the human patient and evade its immune responses, is key to the design of novel anti-TB drug therapies that could oppose this pathogen. The tubercle bacillus expresses a broad arsenal of immuno-modulatory factors for successful colonization in its host. Many of these factors are well-known as outermost surface constituents of the mycobacterial cell envelope. The cell envelope of *mycobacteria* is unique for bearing a characteristic second outer membrane of long fatty acids (mycolic acids) and a polysaccharide-rich capsule layer that surrounds these species [16-18]. The main polysaccharide of the mycobacterial capsule is α-glucan, which represents up to 80% of its sugar content [19]. Capsular α-glucan resembles the structure as well as the chemical composition of glycogen, as it is a polymer composed of linear α-D-(1-4) linked glucose (Glc) units with core substitutions at position 6 every 5 or 6 residues by α-D-(1-4) linked oligoglucosides [20-21]. Today, limited data is available on mycobacterial capsular α-glucan and its functional role. Some studies suggest that it interacts with the complement receptor 3 [22], and that it suppresses the production of interleukin 12 by blocking CD1 expression in dendritic cells, stimulating the production of anti-inflammatory cytokine IL-10 via a CD80-dependent manner [23]. Furthermore, Stokes *et al*. showed that *M. tuberculosis* capsular components exhibit anti-phagocytic properties against special types of macrophages [24]. Finally, the group of Ben J. Appelmelk investigated the role of α-glucan in dendritic cells and showed that it abolishes dendritic cell function by a mechanism that is dependent on the C-type lectin DC-SIGN [25]. Taken together, these studies suggest an important role for capsular α-glucan in host-pathogen interactions, which merit further investigation.

However, gathering deep insights into the host-pathogen interaction mediated by this polysaccharide is challenging due to the lack of fast and reliable techniques to qualify and quantify capsular α-glucan levels. Reasons for these limitations include the absence of specific monoclonal antibodies against the mycobacterial capsular α-glucan, and the lack of comprehensive methodologies for capsular α-glucan mutant screening.

In the 1990s, Baba and colleagues developed a hybridoma cell-line for the production of a murine anti-glycogen monoclonal antibody (mAb) (IV58B6) [15]. Recently, Nakamura-Tsuruta analyzed the carbohydrate-binding specificity of this mAb to several natural and enzymatically synthesized glycogens [26]. However, the mAb specificity was not elucidated at atomic detail, and evidence on its interaction with mycobacterial capsular α-glucan was not provided.

In this chapter, we provide structural details on the epitope-fine specificity of the anti-glycogen mAb (IV58B6) toward polyglucosides. In addition, we provide heteronuclear NMR studies on the structure of the mycobacterial capsule of both wild-type and mutant strains.

3.2 Results

3.2.1 Characterization of the mAb IV58B6/α-glucan interaction

We used saturation transfer difference (STD) experiments to study, from the ligand's point of view, the recognition of glycogen and a selection of related oligoglucosides by the mAb.

First, STD experiments were performed on a selection of α-Glc-containing oligosaccharides: maltotetraose and heptaoses. In particular, the study included linear α(1-4) oligosaccharides, α(1-6)-branched maltotriose and a double α(1-6)-branched maltopentose as their respective α(1-6)-branched analogs, keeping the number of Glc units constant. STD experiments performed on these oligosaccharides in the presence of the mAb gave rise to STD peaks (fig. 4), indicating that these ligands were able to bind to the mAb. However, STD responses were rather poor, which is in agreement with this interaction being only very weak. Of note, STD experiments are very sensitive, allowing for the detection of very weak interactions even for K_D ca. 10 mM [27]. The high chemical shift monotony of the oligoglucosides' signals precluded a detailed epitope mapping of the interactions. However, focusing on the anomeric signals, it became apparent that the α(1-4)-linked H1 protons received a higher amount of magnetization than both the reducing and (1-6)-linked H1 protons, in all instances (fig. 2). Similar experiments performed on the monosaccharide α-methylglucose and the α(1-4)-linked Glc disaccharide (maltose) did not provide any STD response (not shown). Taken together, these results suggest that the tri- or tetrasaccharide is the minimum epitope recognized by the mAb, and point towards a recognition of the inner, α(1-4)-linked Glc units by the mAb, rather than the branching points or the reducing ends.

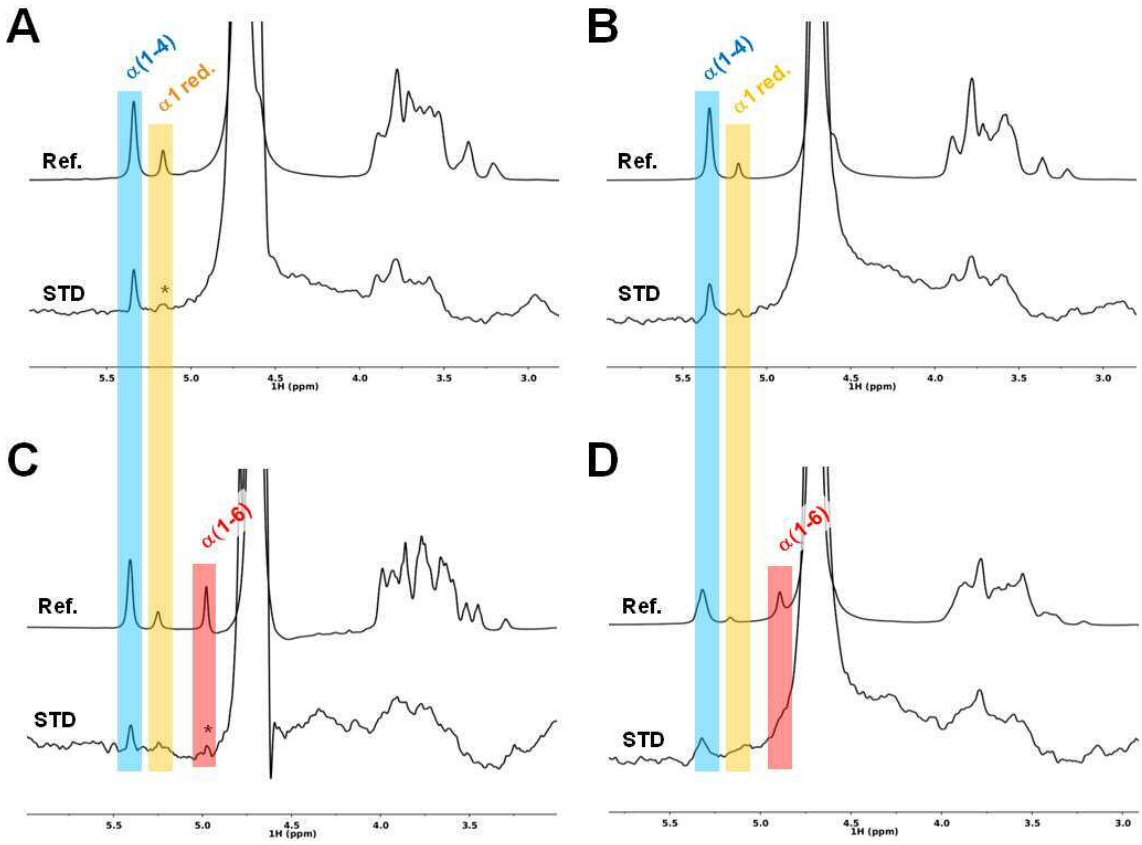


Figure 2: STD experiments performed on a series of α-Glc-containing oligosaccharides. A: maltotetraose; B: maltoheptaose; C: α(1-6)-branched maltotriose; D: a double α(1-6) branched maltopentose. In all instances, the anomeric proton of the linear (1-4) linkages (highlighted in blue) shows a higher STD response than those of the reducing ends (highlighted in orange) and the (1-6)-branches (highlighted in red) (marked with asterisks), suggesting that the recognition takes place at the core (1-4)-linked α-Glc units preferentially.

Next, we performed STD experiments on glycogen. Because glycogen itself, owing to its large size, also becomes inherently saturated to some extent, a blank STD experiment performed in the absence of the mAb was subtracted to the STD spectrum performed on glycogen in the presence of the mAb, in order to obtain a saturation transfer double difference (STDD) spectrum [28], i.e. a dissected contribution of the saturation of glycogen owed to its binding to the mAb. The STDD spectrum (fig. 3A) was analyzed, evidencing that the highest degree of saturation was observed at H3, at the α(1-4)-linked H1, and at the terminal H4 protons (fig. 3B). These results suggest that the recognition of glycogen preferentially takes part at the core α(1-4)-linked H1 protons, as well as at the non-reducing terminal ends.

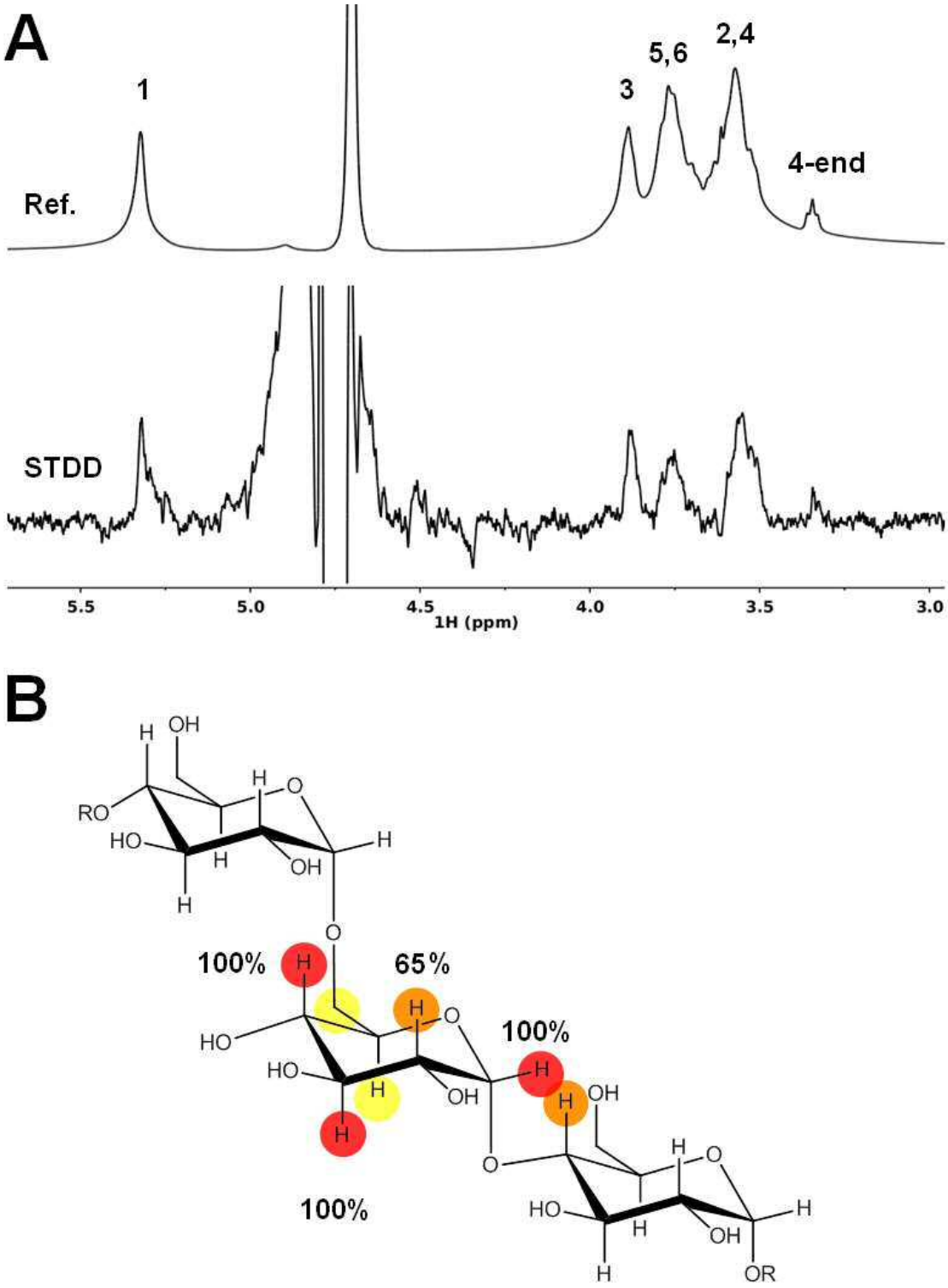


Figure 3: A: STDD experiment on glycogen. Because glycogen itself, as a large molecule, receives inherent magnetization in STD experiments, STD spectra of glycogen in the presence of the mAb were subtracted a control STD spectrum of glycogen in the absence of the mAb to obtained the STDD spectrum. Labels indicate the assignment of the glycogen ^{1}H-NMR signals according to [29]. B: Group epitope mapping as obtained by the relative STDD intensities for each glycogen signal in the spectrum shown in A.

3.2.2 On-cell NMR studies of *M. smegmatis* cell wall glucans

A novel technique for identifying genes involved in the synthesis of the mycobacterial α-glucan using the mAb IV58B6 was developed by the group of Prof. Appelmelk. Briefly, the method is based on the generation of a mutant library through transposon mutagenesis [30], which is screened by immunostaining with the mAb , revealing an array of over-producing and under-producing phenotypes, the genes responsible of which are then identified by ligation-mediated polymerase chain reaction (PCR) [31]. One of the mutants identified as an under-producer was *glgB*::*Tn* (fig. 4). In mycobacteria, the *glgB* gene codes for a glycosyltransferase which introduces α(1-6) branches on the α(1-4) glucan polymer [32]. Thus, *glgB::Tn* mutants are expected to produce a capsular glucan formed by linear α(1-4) chains only.

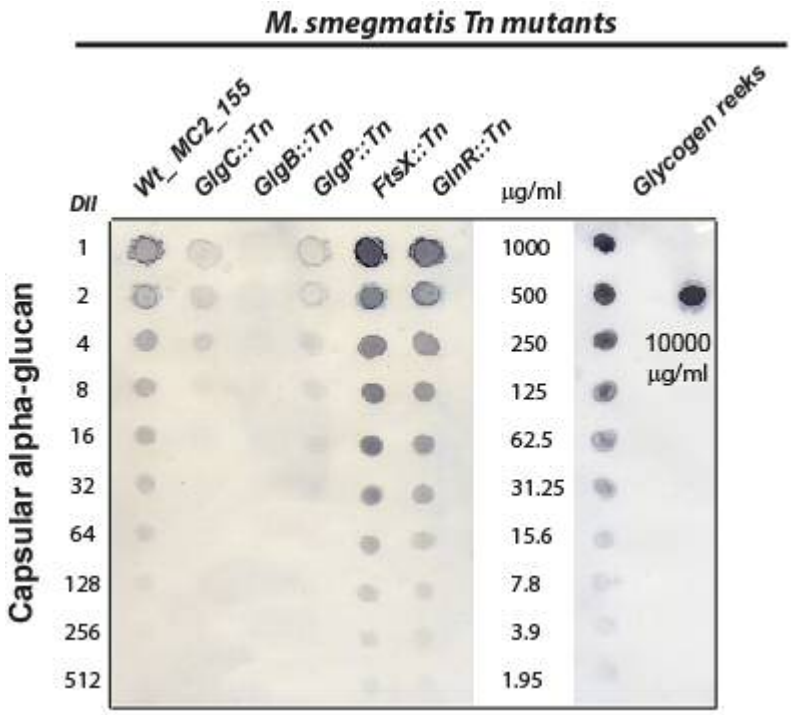


Figure 4: Quantification of capsular α-glucan levels in *M. smegmatis* α-glucan transposon mutants (courtesy of Ben J. Appelmelk).

We used on-cell NMR to get insight into the structure of the cell wall α-glucan envelope of both *M. smegmatis* wild-type and *glgB::Tn* mutants. Suspensions of killed cells of both phenotypes were visible by ^{1}H-NMR in the shape of very broad signals, encompassing chemical shift ranges typical for glycans (3.0 – 5.5 ppm) and lipids (0.8 – 2.5 ppm), and in DOSY spectra

as a single species of log*D* = -11.8, in agreement with it being a single entity of very large size (fig. 5).

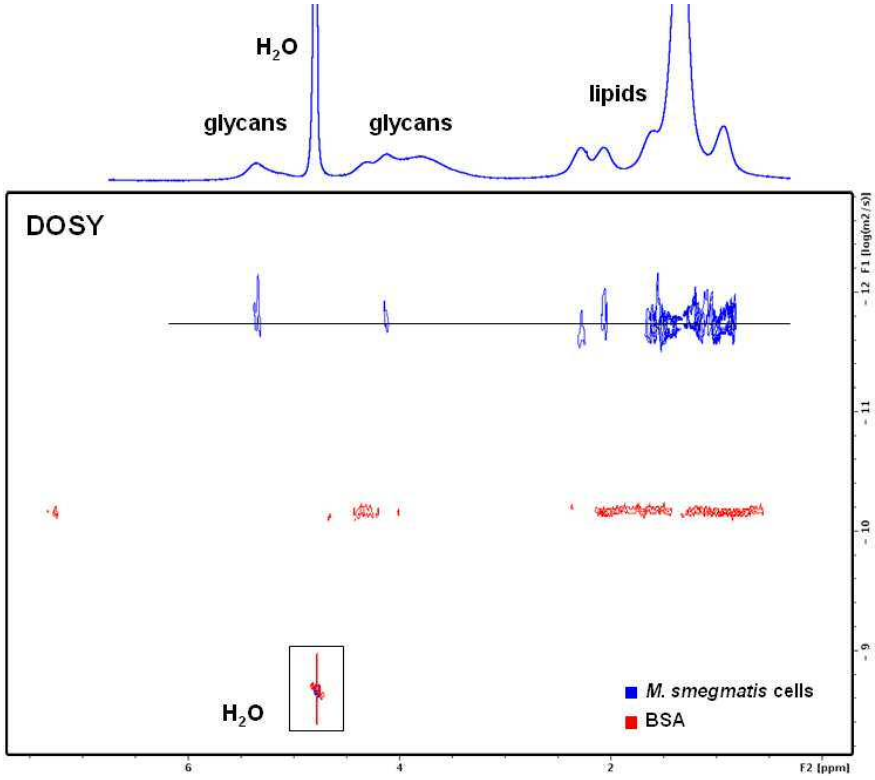


Figure 5: DOSY spectrum of *M. smegmatis* cell preparations (in blue, with 1D ^{1}H-NMR spectrum stacked on the top). The cells are visible as a single entity diffusing with log*D* = -11.8 m²/s, and encompassing signals at chemical shifts characteristic for glycans (3 to 5.5 ppm) and lipids (0.8-2.5 ppm). In red, a DOSY spectrum of the 67 kDa protein bovine serum albumin (BSA) is shown for reference.

^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectra of *M. smegmatis* cell suspensions at ^{13}C natural abundance revealed the presence of a high number of cross-peaks in the chemical shift region typical for glycans (fig. 6). This spectral complexity is due to the rich variety of components of the mycobacterial envelope, including mycolic acids [33], arabinogalactan, peptidoglycan, lipoarabinomannan, phosphatidylinositol mannoside, and others [34]. In order to straightforwardly identify the signals belonging to the α-glucans, these were digested until completion with amyloglucosidase from *Aspergillus niger*, an enzyme that hydrolyzes both Glcα(1-4)Glc and Glcα(1-6)Glc linkages [35]. The resulting spectrum was compared with that of the parent sample, revealing the original signals of the α-glucans.

Expectedly, these signals were very similar to those of glycogen as can be deduced from inspection of the corresponding spectral superimposition (fig. 6A).

Next, spectra from wild type and *glgB::Tn* mutant samples were compared. Two signals characteristic of α(1-6)-linked units were absent in the spectrum of the *glgB::Tn* mutant, i.e. those of 6(6-1) and 4(6-1) (fig. 5B), while those of the linear α(1-4) sequence were maintained. These results indicate that the α-glucan of *glgB::Tn* mutants is primarily formed by a linear, unbranched α(1-4)glucose polysaccharide, lacking α(1-6) branches.

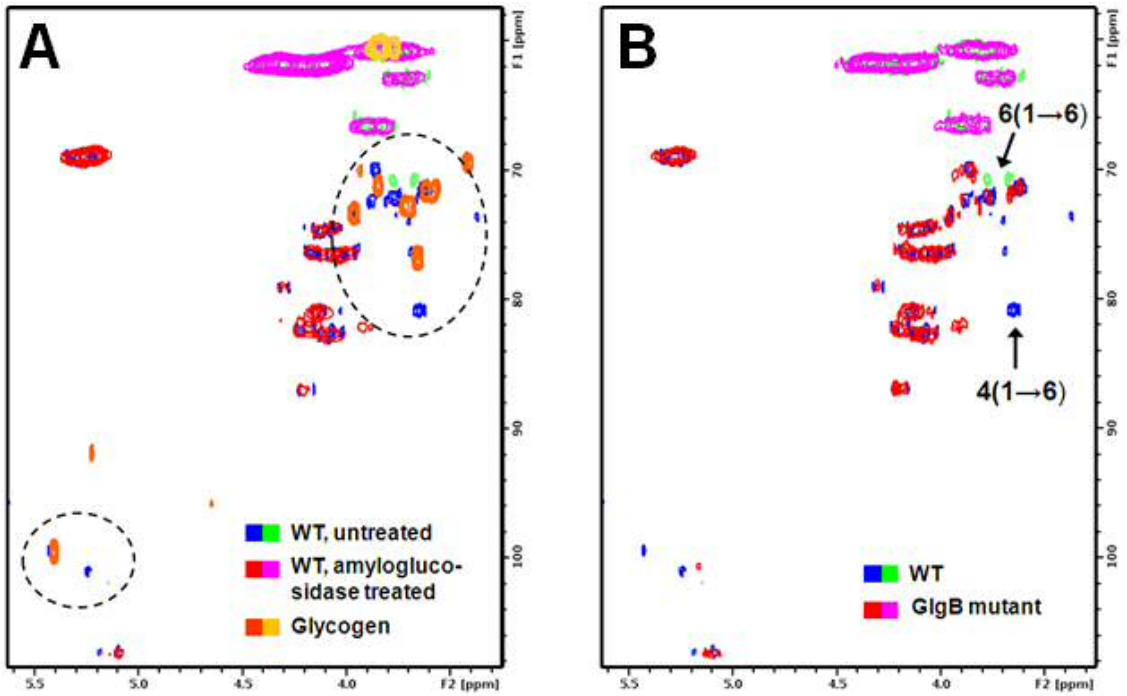


Figure 6: On-cell heteronuclear NMR studies on the mycobacterial capsule. Phase-sensitive ^{1}H-^{13}C HSQC spectra of *M. smegmatis* cells acquired at ^{13}C natural abundance, in which negative cross-peaks belonging to $-CH_2-$ correlations (e.g. position 6 of α-glucans) are shown in lighter colors. A: Identification of the α-glucan signals. Blue/green: untreated wild-type cells. Red/pink: wild-type cells after complete digestion with amyloglucosidase from *Aspergillus niger*. Orange/yellow: glycogen, for reference. Dashed circles highlight the α-glucan signals, which do not appear in the spectrum of amyloglucosidase-treated cells and are highly superimposable with glycogen signals. B: Wild-type vs. *GlgB::Tn* mutant. Blue/green: wild-type cells. Red/pink: *GlgB::Tn* mutant cells. Labels indicate signals identified as those of 6(6-1) and 4(6-1), which are characteristic for α(6-1) branches and do not appear in the spectrum of *GlgB::Tn* mutant cells. ^{1}H-^{13}C correlations were identified using assignments reported by [8] and aided by chemical shift prediction performed with CASPER [36] (figs. S2, S3 in Appendix).

Based on the intensity of the signals in the ^{1}H-^{13}C HSQC spectrum (fig. 6B), the levels of α(1-4) linear glucan present in *glgB::Tn* mutant cells are only about 50% less in comparison to those found in wild-type cells. However, colonies of mutant cells did not show any immunostaining using the mAb, a reason why this mutant seemed to not produce α-glucan at all (fig. 5). These data are in apparent contrast with the ligand epitope mapping deduced from STD data, which seem to indicate that recognition of glycogen and related oligosaccharides by this mAb preferentially involves the α(1-4) linear core (figs. 2-3). This fact raises the question why *glgB::Tn* mutants do not appear to bind the anti-glycogen mAb.

There are two main reasons that may explain this discrepancy. First, because α(1-6) branching provides the α-glucan with additional α(1-4) chains, the capsule of *glgB::Tn* mutants could actually bear lower levels of α-glucan epitopes for the mAb. Since the ^{1}H-^{13}C HSQC spectrum does not allow to discriminate between signals from intracellular and capsular α-glucan, it is possible that the greatest portion of the α-glucan signals in the spectrum of *glgB::Tn* mutants arise from intracellular α-glucan, which is not accessible to the antibody. However, this does not appear to be the case, since previous data showed that *glgB* knockout mutants produced 57% and 36% of normal capsular and intracellular α-glucan levels, respectively [32].

Secondly, amylose-like, linear α(1-4) linked polymers are known to be largely insoluble in hydrophilic conditions and to form aggregates [37]. Therefore, we hypothesize that binding of the mAb to *glgB::Tn* mutants is precluded due to strongly reduced availability of the α(1-4) linked α-glucan epitope caused by insolubility and aggregation of the unbranched α-glucan polymers. Electron microscopy images of *glgB::Tn* mutant cells obtained by the group of B. J. Appelmelk, showing electron-dense insoluble aggregates not visible in preparations of wild-type cells provide support to this hypothesis.

3.3 Methods

3.3.1 STD experiments

STD experiments were acquired at 25 ºC in a Bruker AVANCE 600 MHz spectrometer equipped with a cryoprobe in fully deuterated Tris-d_{11} 25 mM buffer, pD 8.0, 150 mM NaCl. Samples contained 5 mg glycogen from bovine liver (Sigma) or 2 mM of the studied oligosaccharides and 0.2 mg/ml of the monoclonal antibody IV58B6 (mAb), kindly ceded by O. Baba [15]. The

mAb was saturated on-resonance at 0 ppm and off-resonance at 100 ppm with a train of Gaussian-shaped pulses of 50 ms each, totaling an irradiation time of 2 s. STD spectra were obtained by subtracting the on-resonance to the off-resonance spectra. In the case of glycogen, a separate STD spectrum was acquired in the absence of the mAb and subtracted to the STD spectrum in the presence of the mAb to obtain a saturation transfer double difference (STDD) spectrum. STD(D) intensities were measured comparing each STD(D) spectrum with the correspondent off-resonance spectrum and normalizing to the ligand peak receiving the highest degree of saturation. Data were analyzed with Bruker TopSpin 3.0 and figures were built with MestReNova v.8.0.2.

3.3.2 On-cell NMR

M. smegmatis mc^2155 wild-type and *glgB::Tn* strains were cultured in the laboratory of Ben J. Appelmelk (University of Amsterdam). Bacteria were grown for 2 days in 7H9 liquid medium supplemented with 10% albumin-dextrose-catalase and tween-80. Then, they were washed 2 times with PBS, pelleted and heat-killed for 2 hours at 80 ºC. For NMR observarions, they were washed (x2) in D_2O and resuspended in D_2O. ^{1}H-NMR and multiplicity-edited ^{1}H-^{13}C HSQC spectra were acquired at 37 ºC in a Bruker AVANCE 700 MHz spectrometer equipped with a cryoprobe. For identification of the cell wall α-glucan signals, wild-type cell samples were incubated with 60 U amyloglucosidase from *Aspergillus niger* (Sigma, St Louis, MO), in 50 mM sodium phosphate buffer, pH 5.6 at 55 ºC until digestion was complete. Next, cells were pelleted, washed and spectra were recorded again and compared with those recorded before treatment. Diffusion-ordered spectroscopy (DOSY) experiments were run using the standard Bruker sequence, in a Bruker AVANCE 500 MHz spectrometer. 32 ^{1}H-NMR spectra were collected with a diffusion time (big delta) of 1 s for cell samples and 0.4 s for a sample containing BSA, and an gradient pulse (little delta) of 3 ms. The 1D spectra were processed and baseline-corrected, and the diffusion dimensions were exponentially fitted according to pre-set windows.

3.4 References

[1] Glanville, J., Zhai, W., Berka, J., Telman, D., Huerta, G., Mehta, G. R., Ni, I., Mei, L., Sundar, P. D., Day, G. M., Cox, D., Rajpal, A., and Pons, J. (2009), *Proc Natl Acad Sci U S A* **106**, 20216-20221.

[2] Hozumi, N., and Tonegawa, S. (1976), *Proc Natl Acad Sci U S A* **73**, 3628-3632.

[3] Harris, L. J., Larson, S. B., Hasel, K. W., and McPherson, A. (1997), *Biochemistry* **36**, 1581-1597.

[4] Perkins, S. J., Nealis, A. S., Sutton, B. J., and Feinstein, A. (1991), *J Mol Biol* **221**, 1345-1366.

[5] Huang, Y. L., Hung, J. T., Cheung, S. K., Lee, H. Y., Chu, K. C., Li, S. T., Lin, Y. C., Ren, C. T., Cheng, T. J., Hsu, T. L., Yu, A. L., Wu, C. Y., and Wong, C. H. (2013), *Proc Natl Acad Sci U S A* **110**, 2517-2522.

[6] Astronomo, R. D., and Burton, D. R. (2010), *Nat Rev Drug Discov* **9**, 308-324.

[7] Johnson, M. A., and Bundle, D. R. (2013), *Chem Soc Rev* **42**, 4327-4344.

[8] Slovin, S. F., Keding, S. J., and Ragupathi, G. (2005), *Immunol Cell Biol* **83**, 418-428.

[9] Holgersson, J., Gustafsson, A., and Breimer, M. E. (2005), *Immunol Cell Biol* **83**, 694-708.

[10] Houliston, R. S., Jacobs, B. C., Tio-Gillen, A. P., Verschuuren, J. J., Khieu, N. H., Gilbert, M., and Jarrell, H. C. (2009), *Biochemistry* **48**, 220-222.

[11] Johnson, M. A., Jaseja, M., Zou, W., Jennings, H. J., Copie, V., Pinto, B. M., and Pincus, S. H. (2003), *J Biol Chem* **278**, 24740-24752.

[12] Rademacher, C., Shoemaker, G. K., Kim, H. S., Zheng, R. B., Taha, H., Liu, C., Nacario, R. C., Schriemer, D. C., Klassen, J. S., Peters, T., and Lowary, T. L. (2007), *J Am Chem Soc* **129**, 10489-10502.

[13] Johnson, M. A., Cartmell, J., Weisser, N. E., Woods, R. J., and Bundle, D. R. (2012), *J Biol Chem* **287**, 18078-18090.

[14] Bundle, D. R., Baumann, H., Brisson, J. R., Gagne, S. M., Zdanov, A., and Cygler, M. (1994), *Biochemistry* **33**, 5183-5192.

[15] Baba, O. (1993), *Kokubyo Gakkai Zasshi* **60**, 264-287.

[16] Brennan, P. J., and Nikaido, H. (1995), *Annu Rev Biochem* **64**, 29-63.

[17] Daffe, M., and Etienne, G. (1999), *Tuber Lung Dis* **79**, 153-169.

[18] Hoffmann, C., Leis, A., Niederweis, M., Plitzko, J. M., and Engelhardt, H. (2008), *Proc Natl Acad Sci U S A* **105**, 3963-3967.

[19] Ortalo-Magne, A., Dupont, M. A., Lemassu, A., Andersen, A. B., Gounon, P., and Daffe, M. (1995), *Microbiology* **141 (Pt 7)**, 1609-1620.

[20] Dinadayala, P., Sambou, T., Daffe, M., and Lemassu, A. (2008), *Glycobiology* **18**, 502-508.

[21] Lemassu, A., and Daffe, M. (1994), *Biochem J* **297 (Pt 2)**, 351-357.

[22] Cywes, C., Hoppe, H. C., Daffe, M., and Ehlers, M. R. (1997), *Infect Immun* **65**, 4258-4266.

[23] Gagliardi, M. C., Lemassu, A., Teloni, R., Mariotti, S., Sargentini, V., Pardini, M., Daffe, M., and Nisini, R. (2007), *Cell Microbiol* **9**, 2081-2092.

[24] Stokes, R. W., Norris-Jones, R., Brooks, D. E., Beveridge, T. J., Doxsee, D., and Thorson, L. M. (2004), *Infect Immun* **72**, 5676-5686.

[25] Geurtsen, J., Chedammi, S., Mesters, J., Cot, M., Driessen, N. N., Sambou, T., Kakutani, R., Ummels, R., Maaskant, J., Takata, H., Baba, O., Terashima, T., Bovin, N., Vandenbroucke-Grauls, C. M., Nigou, J., Puzo, G., Lemassu, A., Daffe, M., and Appelmelk, B. J. (2009), *J Immunol* **183**, 5221-5231.

[26] Nakamura-Tsuruta, S., Yasuda, M., Nakamura, T., Shinoda, E., Furuyashiki, T., Kakutani, R., Takata, H., Kato, Y., and Ashida, H. (2012), *Carbohydr Res* **350**, 49-54.
[27] Meyer, B., and Peters, T. (2003), *Angew Chem Int Ed Engl* **42**, 864-890.
[28] Claasen, B., Axmann, M., Meinecke, R., and Meyer, B. (2005), *J Am Chem Soc* **127**, 916-919.
[29] Zang, L. H., Howseman, A. M., and Shulman, R. G. (1991), *Carbohydr Res* **220**, 1-9.
[30] Blake, K. L., and O'Neill, A. J. (2013), *J Antimicrob Chemother* **68**, 12-16.
[31] Abdallah, A. M., Verboom, T., Hannes, F., Safi, M., Strong, M., Eisenberg, D., Musters, R. J., Vandenbroucke-Grauls, C. M., Appelmelk, B. J., Luirink, J., and Bitter, W. (2006), *Mol Microbiol* **62**, 667-679.
[32] Sambou, T., Dinadayala, P., Stadthagen, G., Barilone, N., Bordat, Y., Constant, P., Levillain, F., Neyrolles, O., Gicquel, B., Lemassu, A., Daffe, M., and Jackson, M. (2008), *Mol Microbiol* **70**, 762-774.
[33] Verschoor, J. A., Baird, M. S., and Grooten, J. (2012), *Prog Lipid Res* **51**, 325-339.
[34] Brennan, P. J. (2003), *Tuberculosis (Edinb)* **83**, 91-97.
[35] Pazur, J. H., and Ando, T. (1959), *J Biol Chem* **234**, 1966-1970.
[36] Jansson, P. E., Stenutz, R., and Widmalm, G. (2006), *Carbohydr Res* **341**, 1003-1010.
[37] Buleon, A., Colonna, P., Planchot, V., and Ball, S. (1998), *Int J Biol Macromol* **23**, 85-112.

CHAPTER 4

STUDIES ON UDP-GLUCOSE PYROPHOSPHORYLASE OF STREPTOCOCCUS PNEUMONIAE, A DRUG TARGET CANDIDATE

4.1 Introduction

Carbohydrate-processing enzymes catalyze the synthesis of activated sugar donors [1], as well as the transfer of sugar residues from these donors to other carbohydrates or aglycons [2], and thus, they are responsible for the genesis of the whole cellular glycan repertoire.

Glycosyl phosphoesters of nucleoside pyrophosphates, more often referred to as sugar-nucleotides, occupy a relevant place in carbohydrate metabolism as the activated forms serving as glycosyl donors for the synthesis of complex carbohydrates and the sugar moieties of glycoconjugates. Among these molecules, UDP-glucose is a highly multifunctional metabolite, being essential for a rich variety of processes and a crossroad point in carbohydrate metabolism. In animal and fungal metabolism, UDP-glucose is the sugar donor for glycogen synthesis and, as in all eukaryotes, in the formation of the carbohydrate moieties of glycolipids, glycoproteins and proteoglycans [3]. Besides, it is also required for galactose utilization [4]. In plants, UDP-glucose is used for the synthesis of sucrose and cellulose and is involved in starch metabolism [5]. In prokaryotes, perhaps the most distinguishing role of UDP-glucose is its participation in the synthesis of different components of the bacterial envelope, particularly the lipopolysaccharide (LPS) and the capsule, structures that represent necessary virulence factors of many microorganisms.

UTP:α-D-glucose-1-phosphate uridylyltransferase (EC 2.7.7.9), commonly referred to as UDP-glucose pyrophosphorylase, and abbreviated as UGP or by the capitalization of its coding gene in *E. coli* (GalU), catalyzes the reversible formation of UDP-glucose and pyrophosphate (PPi) from glucose-1-phosphate and UTP (fig. 1). UGP activity is ubiquitous to all domains of life, given the functional importance of its product. However, and interestingly, prokaryotic UGPs are evolutionary unrelated to their eukaryotic counterparts [6].

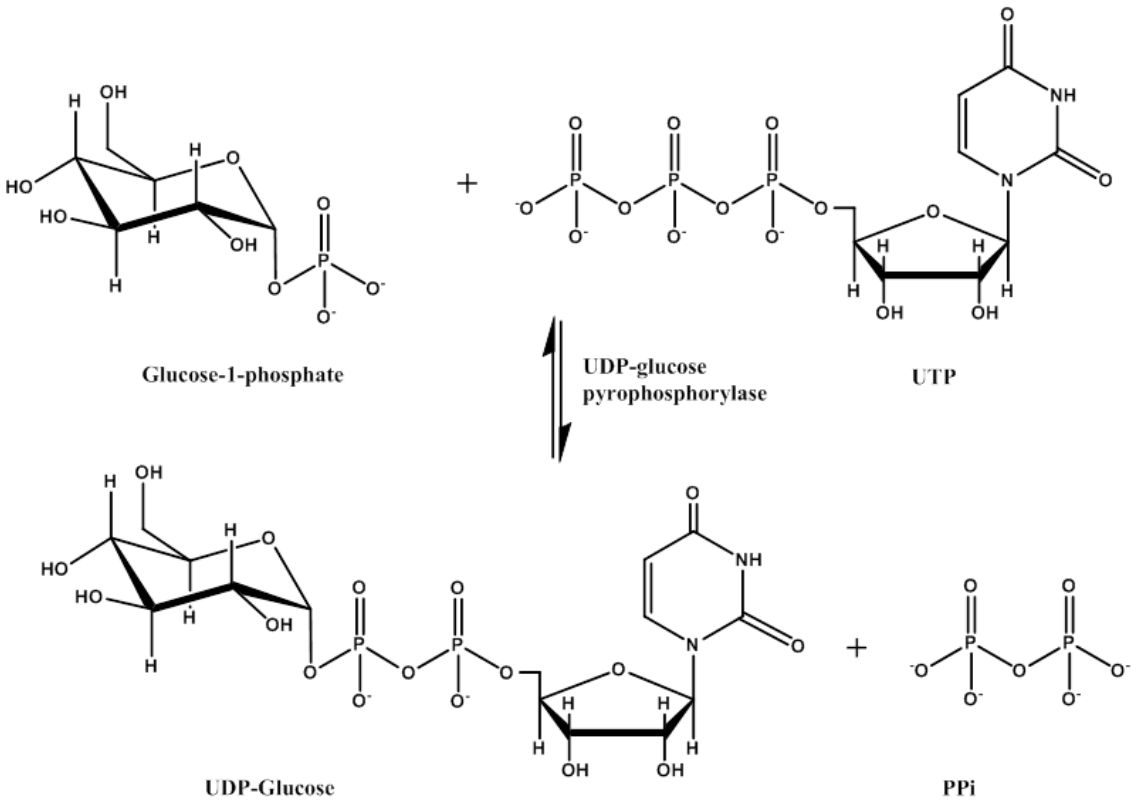


Figure 1: Reaction scheme for UDP-glucose pyrophosphorylase

Given its essential activity for the cell metabolism, UDP-glucose pyrophosphorylases are ubiquitously distributed throughout all domains of life. UGPs are, in addition, very well conserved proteins: for instance, the sequence of the *E. coli* UGP bears a 29% identity with that from the archean *Sulfolobus*, and the human and barley enzymes share a 55% identity. However, and interestingly, sequence similarities between UGPs of prokaryotic origin and their eukaryotic counterparts lie around 8%, which is considered non-significant. For comparison, human UDP-galactose-4-epimerase, another enzyme rendering UDP-glucose as product, shares an identity of 51% with its *E. coli* homologue [7]. Thus, the UGPs of prokaryotic and of eukaryotic origin are believed to be evolutionarily unrelated [6, 8]. Fittingly, Hartman and Fedorov classified the eukaryotic UGP as a “eukaryotic signature protein”, being itself one of only a handful enzymes in a list of 347 proteins that were considered to be genuine eukaryotic inventions [9].

In light of these data, a number of authors have underlined the attractiveness of bacterial UGPs as potential antimicrobial targets, as a means of selectively inhibit the pathogen enzyme without disrupting the host metabolism [6, 8, 10].

4.1.1 Structure

Up to date, the crystallographic structures of four bacterial UGPs have been deposited in the Protein Data Bank: those from *Escherichia coli* (acc.: 2E3D) [11], *Corynebacterium glutamicum* (2PA4) [12], *Sphingomonas elodea* (2UX8) [13] and *Helicobacter pylori* (3JUJ and 3JUK) [8], all of which share a very similar structure. Here, the structure of the *H. pylori* enzyme is discussed as representative of the group.

All four UGP structures were crystallized as homotetramers (fig. 4), exhibiting mean dimensions of 85 Å × 75 Å × 50 Å. Each monomer, of approximately 30-35 kDa, presents a characteristic folding, dominated by a central mixed β-sheet, reminiscent of the nucleotide-binding Rossmann fold, but including nine β-strands, surrounded by eleven α-helices and two additional β-strands (fig 5). The overall quaternary structure can be understood as a dimer of dimers, with 'tight' dimers resulting of extensive interactions established between two adjacent subunits, and 'loose' dimers arisen from weaker packing interactions between the alternate pairs of subunits (fig. 4). In the *H. pylori* UGP, the 'tight' dimers result from interactions made by residues located mainly in the α1, α4 and α12 helices and the α2-α3 loop, between subunits A and B, and C and D, respectively. In other UGPs (e.g. that of *S. elodea*), an additional helix exists at the C-terminus, which gives rise, together with the preceding helix, to a V-shaped domain that further stabilizes the dimer by interacting with its homologous in the other monomer. The 'loose' dimers arise from contacts between the □3 strands from each alternate subunit A and C, and B and D, respectively.

4.1.1.1 Active center

As in other pyrophosphorylases, such as thymidylyltransferases, each bacterial UGP subunit carries on an active site made up by residues located in the same polypeptide chain [14]. This is in contrast with cytidylyltransferases, in which active sites are formed by residues located in different subunits [12]. In UGPs, the active center is located in a deep cleft of approximately 20 Å × 20 Å × 15 Å, delimited by β1, β4, β8, α9, α7, α11 and nearby loops (fig. 5). Residues

involved in binding and catalysis are extremely conserved among UGPs of different species. These residues are located mainly in loops facing the active site cleft, rather than in the neighboring α-helices or β-strands.

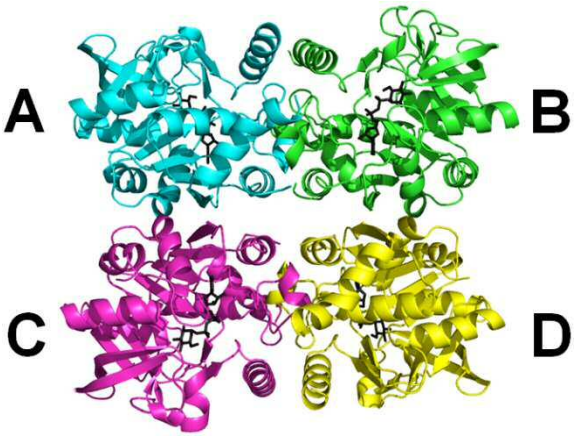


Figure 2: Tetrameric structure of *H. pylori* UGP (PDB: 3JUK) [8]. Subunits A and B (in blue and green, respectively) and C and D (in pink and yellow, respectively) interact within a region including three α-helices and one loop, forming the respective 'tight dimers'. In addition, subunits A and D, and B and C are packed, in a weaker manner, with participation of residues located in two β-strands.

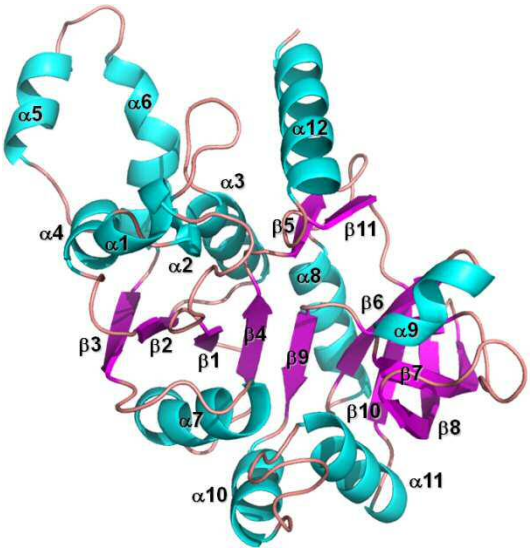


Figure 3: *H. pylori* UGPase in its unliganded form (PDB code: 3JUJ), with labels indicating the identity of the individual alpha helices and beta strands.

Fig. 6 shows UDP-glucose anchored to the *H. pylori* UGPase active site, showing key contacts between the enzyme atoms and those of its product. In the bound state, both the nucleoside and sugar moieties are buried inside the active site pocket, unlike the phosphate groups, which remain fully accessible to the solvent. The nucleoside moiety of UDP-glucose interacts with the Ala10 and Gly11 backbone NHs from the conserved N-terminal motif, at the level of the O2 and O2' atoms, respectively. In addition, the Gln102 ε-amide and the Gly107 NH interact with O4 of the uracyl moiety. Negatively charged phosphate groups are docked to the enzyme by electrostatic interactions with a pair of conserved lysine residues. The first phosphate interacts with the Lys191 side-chain from the VEKP signature motif, whereas the second phosphate contacts the Lys25 side-chain. Two phosphoryl oxygens from the phosphate pair are coordinated with a Mg^{2+} ion, whose coordination sphere is completed by the Glu130 carboxylate and three ordered water molecules. Gly171 can hydrogen-bond to any of both the

3’- and 4’-hydroxyl groups of the glucose moiety, Asp131 contacts the 6’-hydroxyl group, and Glu190 carboxyl oxygens from the VEKP motif interact with the sugar 2’ and 3’-hydroxyl groups, a bidentate interaction commonly found in other glucose-protein interactions, such as those found in lectins [15].

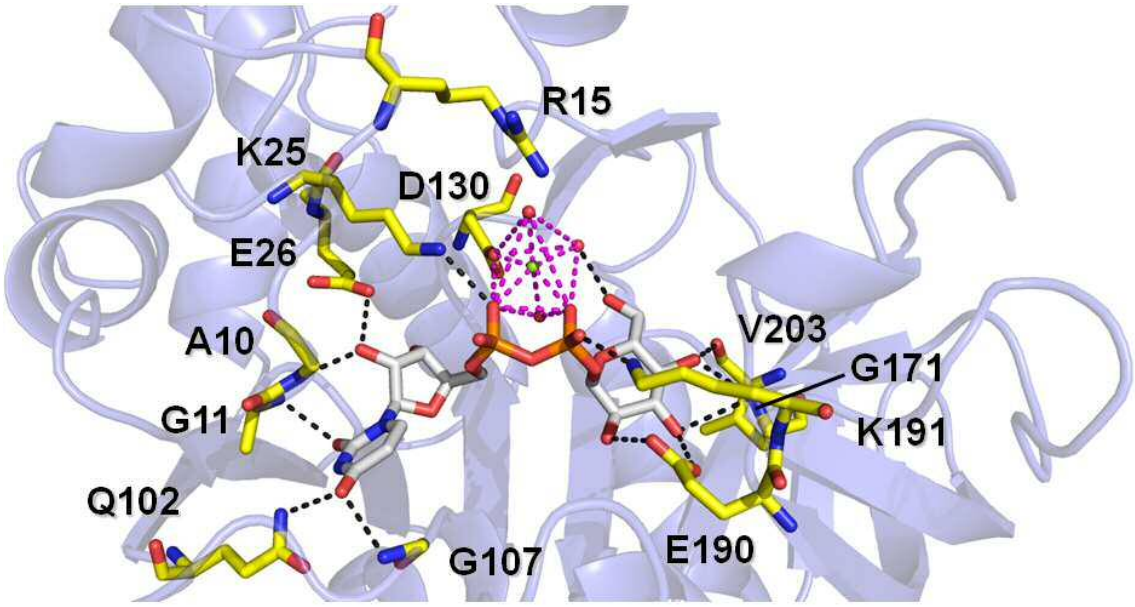


Figure 4: Diagram of the *H. pylori* UGP active center, in complex with UDP-glucose [8]. The protein residues participating in substrate binding are depicted in yellow, and a magnesium ion, coordinated with Glu130, two phosphoryl groups and three water molecules, is shown in green. Black dotted lines represent the inferred interactions implicated in substrate binding. Arg15 is located apart from the ligand and is not involved in binding, since mutation of this residue has no significant effects on the affinity of the enzyme towards its substrates [8]. Instead, it plays an important role in the catalytic activity.

4.1.2 Mechanism of action

In UDP-glucose pyrophosphorylases, the reaction is initiated by a nucleophilic attack of the glucose-1-P phosphoryl oxygen on the phosphorous atom of the UTP α-phosphate. It is thought that a Mg^{2+} ion participates in the catalysis by both stabilizing the excess of negative charge and properly positioning the phosphoryl oxygen of glucose-1-phosphate for its nucleophilic attack [16]. Indeed, the presence of magnesium has been shown to be absolutely required for the catalysis [17], as well as enhancing the binding affinity of UGP towards UTP and UDP-glucose [8]. After the reaction takes place, the β- and γ-phosphates are displaced and

leave the enzyme/UDP-glucose complex as inorganic pyrophosphate, followed by dissociation of UDP-glucose. In the reverse reaction, pyrophosphate nucleophilically attacks UDP-glucose and glucose-1-P is displaced, followed by release of UTP.

4.1.3 Function and value as a therapeutic target

The relevance of UGPs for bacterial cell physiology is determined by the pivotal role of its product UDP-glucose in the bacterial metabolic pathways (Fig. 7). Occupying this central position in the bacterial metabolism, UDP-glucose serves both as a signaling molecule and as a building block for the synthesis of complex polysaccharide structures. As noted above, one of the main functions of UDP-glucose, in prokaryotes is to serve as glucosyl donor for carbohydrate biosynthesis. UDP-glucose is required for the synthesis of LPS, cell wall sugar moieties, capsular polysaccharides and membrane-derived oligosaccharides, as well as exopolysaccharides, either directly or via UDP-galactose. In addition, production of UDP-glucose has been linked with osmotolerance and the complex coordination of cell size and the control of the bacterial cell cycle.

The central position of UDP-glucose in the bacterial biochemical pathways, very prominently those involved in the synthesis of envelope structures, determines that strains defective for this enzyme show markedly impaired pathogenic phenotypes. Not unexpectedly, the gene *galU* can be found as an essential gene for one third of all bacteria taxa included in the database of essential genes [18].

In *S. pneumoniae*, ninety different capsular types have been described [19], and the organization of the gene cluster responsible for the biosynthesis of several capsular types has been analyzed [19-21]. UDP-glucose is needed for the addition of glucose, galactose (via the epimerization of the activated sugar), glucuronic and galacturonic acid residues of the pneumococcal capsule. Since at least one of these four sugars is found in all capsular polysaccharides, pharmacological inhibition of UGP would affect the biosynthesis of the pneumococcal capsule, considered a *sine qua non* component of pathogenicity of this microorganism [10, 17]. Figure 8C shows the unencapsulated phenotype of *galU*-defective pneumococci, which is correlated with its apathogenicity.

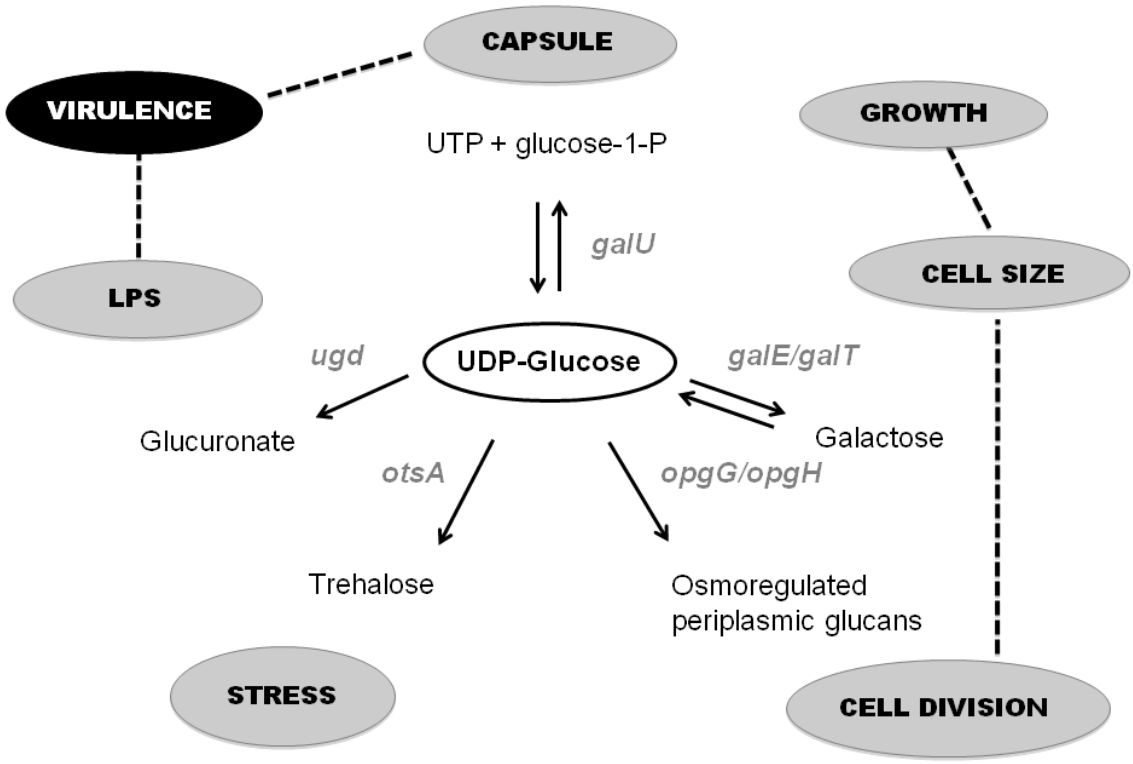


Figure 5: Diagram showing the central position of UDP-glucose in the bacterial biochemical pathways. Arrows indicate the enzymatic transformations of UDP-glucose into several metabolites in *E. coli*, catalyzed by enzymes denoted by codes in gray. Gray ovals represent bacterial functions associated with those enzymatic activities. These functions, especially the presence of capsule and/or LPS in the bacterial envelope, are related with bacterial virulence (represented by a black oval).

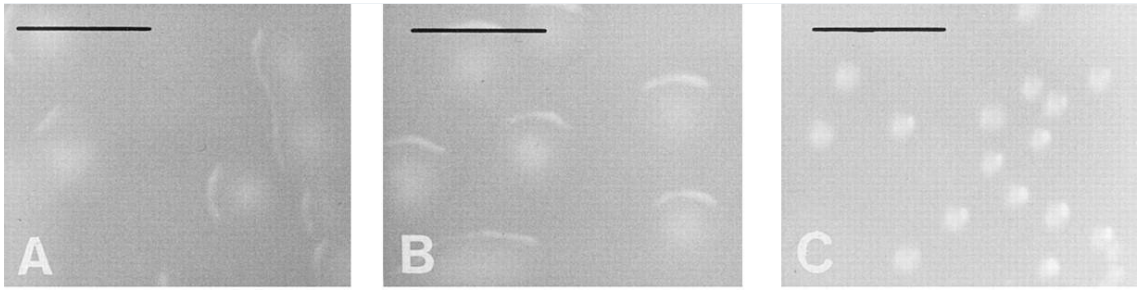


Figure 6: Capsule phenotypes for three pneumococcal strains: A: M23, B: M23c, and C: M23g (GalU), showing the non-encapsulated phenotype of the *galU* mutant (C). Bar, 1 mm. Adapted from [22].

In addition to *S. pneumoniae*, it has been shown that galU mutation produces aberrant envelope structures in a large number of organisms. Consequently, the galU UDPG:PP enzymatic activity may represent an important target in fighting bacterial infectious diseases

[17]. In *E. coli*, such mutants were reported to produce incomplete LPS, containing only heptose and 3-deoxy-D-manooctulosonic acid as sugar moieties. Moreover, these mutants showed motility impairment due to lack of functional flagella formation [23]. Loss of motility has been also shown to occur in other pathogenic organisms, such as *Pseudomonas*, upon galU mutation [24]. In addition, mutations in genes coding for UGPs have been shown to lead to decreased virulence of a number of diverse bacterial pathogens. In *Klebsiella pulmoniae*, galU mutation led to loss of mucoid colony phenotype and virulence in mice and high sensitiveness in human serum [25]. galU$^-$ *Vibrio colerae* was defective in colonization and lost its capability to synthesize an exopolysaccharide biofilm involved in the formation of a resistant rugose variant [26]. Similarly, cornea infecting and systemic spreading capacities of *Pseudomonas aeruginosa* were impeded upon galU mutation [27], while the plant pathogen *Xanthomonas citri* was shown to require UGP activity for *in planta* growth and pathogenicity [28].

Considering all above, the advent of new molecules targeted against the activity of bacterial UGPs will obviously open new ways for broad applications in biotechnology and biomedicine. Specifically, inhibitors of this enzyme would be of great value for the treatment of diseases caused by pathogens that are resistant to the current therapeutic arsenal. The promising profile of bacterial UGPs as potential target candidates has been referred by many authors, given its participation in key metabolic pathways and the synthesis of some of the most important bacterial virulence factors, together with the fact that there is no relation, aside from the catalytic activity, between bacterial UGPs and their eukaryotic counterparts. This last feature very well suits bacterial UGPs to provide the required specificity to avoid undesired toxicities, therefore possibly accelerating its development and speeding up its process to market.

The objectives of this chapter are to develop NMR-based procedures useful for mechanistical studies on this family of enzymes, using the biomedically relevant UGP from *S. pneumoniae* as a model, which can be used further in the characterization of inhibitory drugs or in the exploration of metabolic pathways in which sugar-nucleotides are involved. In addition, we present a fragment-based screening against spUGP, which represents the first step toward drug targeting of bacterial UGPs.

4.2 Results

4.2.1 Insights into the spUGP reaction mechanism

To get insight into the mechanism of spUGP by NMR, STD experiments were performed on its three ligands that are visible by proton NMR, i.e. UDP-Glc, UTP and Glc-1-P, in the presence of the enzyme. The outcome of these experiments can be seen in figs. 7-9. Both UDP-Glc and UTP gave rise to STD signals in the presence of spUGP, indicating that its binding can be monitored by NMR. Group epitope mapping studies showed that a similar pattern of selective saturation was found at the nucleoside part of both substrates, with H1 of uracyl receiving the highest degree of saturation, followed by the ribose signals, especially H1' and H4' protons.

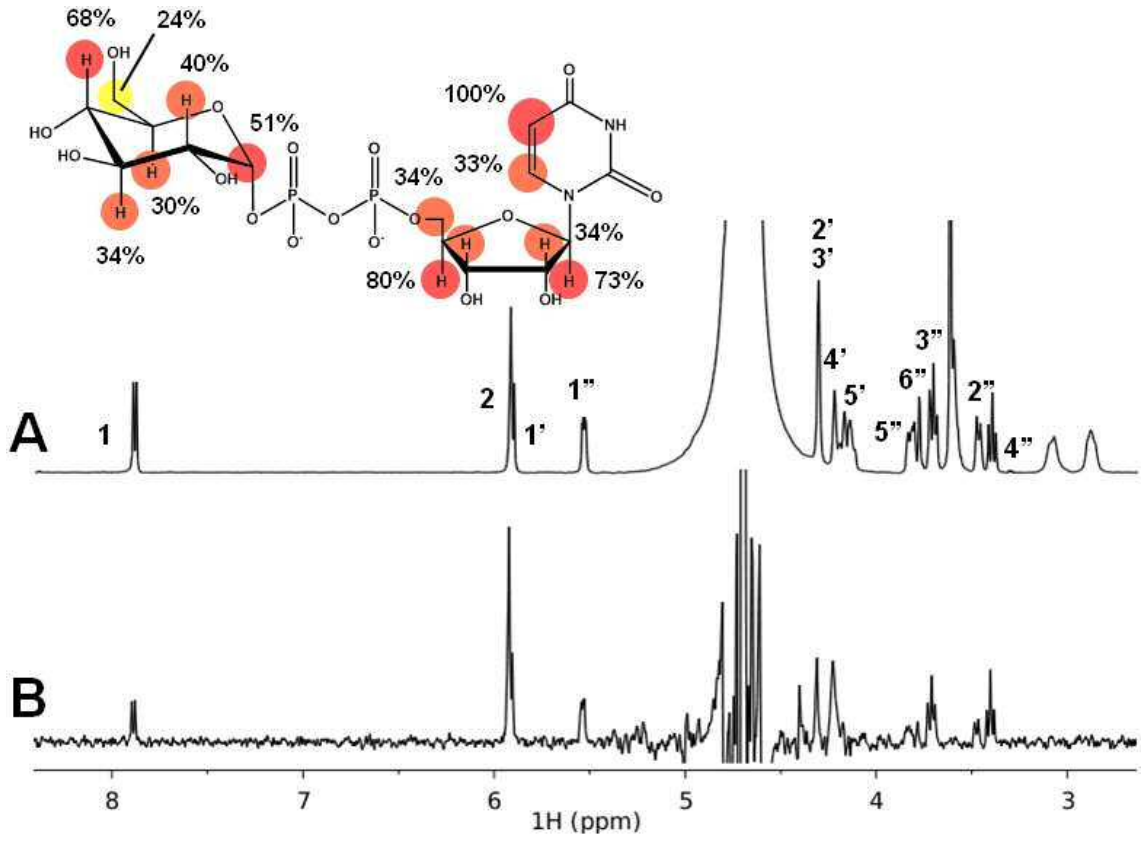


Figure 7: Recognition of UDP-glucose. Off-resonance (A) and STD (B) spectrum of UDP-Glc in the presence of spUGP, with labels indicating ^{1}H-NMR assignments. A molecule of UDP-Glc is shown, with labels indicating the STD intensity of each signal, relative to the STD intensity of the most saturated signal. Red circles: I > 50%, orange circles: 50% I > 30%, yellow circles: 30% > I > 10%.

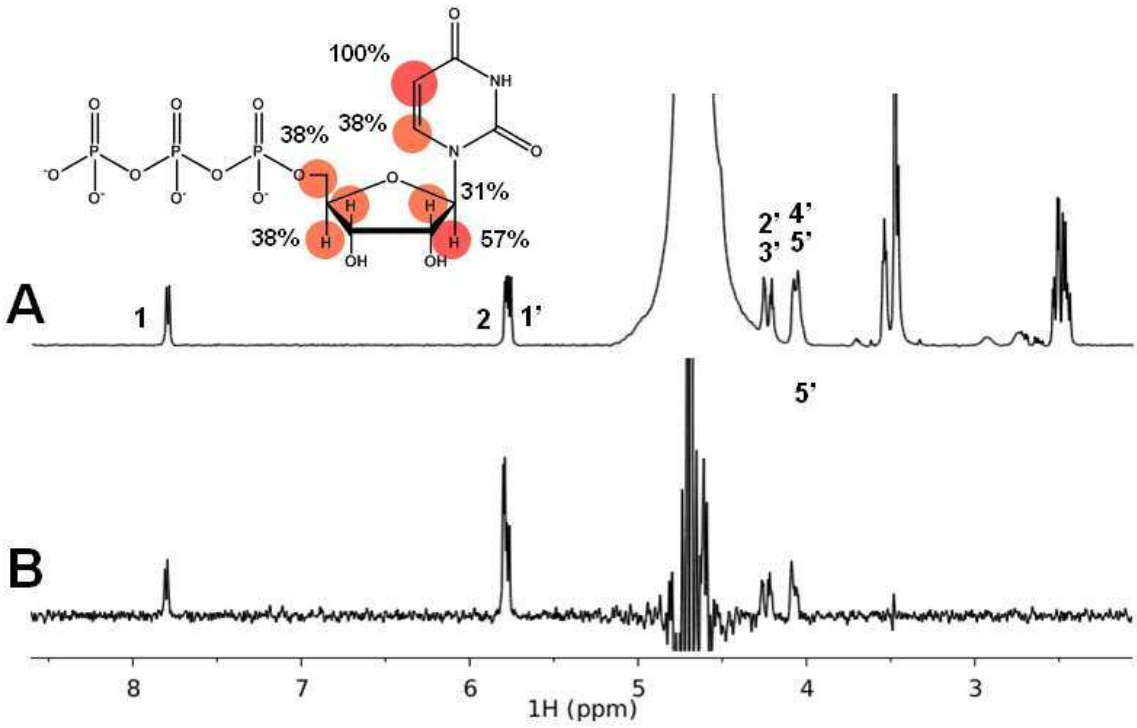


Figure 8: Recognition of UTP. Off-resonance (A) and STD (B) spectrum of UTP in the presence of spUGP, with labels indicating ^{1}H-NMR assignments. A molecule of UTP is shown, with labels indicating the STD intensity of each signal, relative to the STD intensity of the most saturated signal. Red circles: I > 50%, orange circles: 50% I > 30%.

In contrast, STD experiments performed on Glc-1-P in the presence of spUGP were negative (fig. 9B). However, addition of UTP to the same samples triggered the rising of signals in the STD spectrum not only for the nucleotide, but also for the sugar phosphate (fig. 9D). In those instances, signals of UDP-Glc were also observed in the spectra, because co-incubation of UTP and Glc-1-P with the enzyme in the NMR tube resulted in their enzymatic transformation into UDP-Glc and PPi.

Lastly, STD signals of the different substrates were abolished upon addition of EDTA to the samples (fig. 9E), thus indicating that magnesium is needed not only for catalysis, but also for ligand binding. Together, these results can give insight into the reaction mechanism of this enzyme.

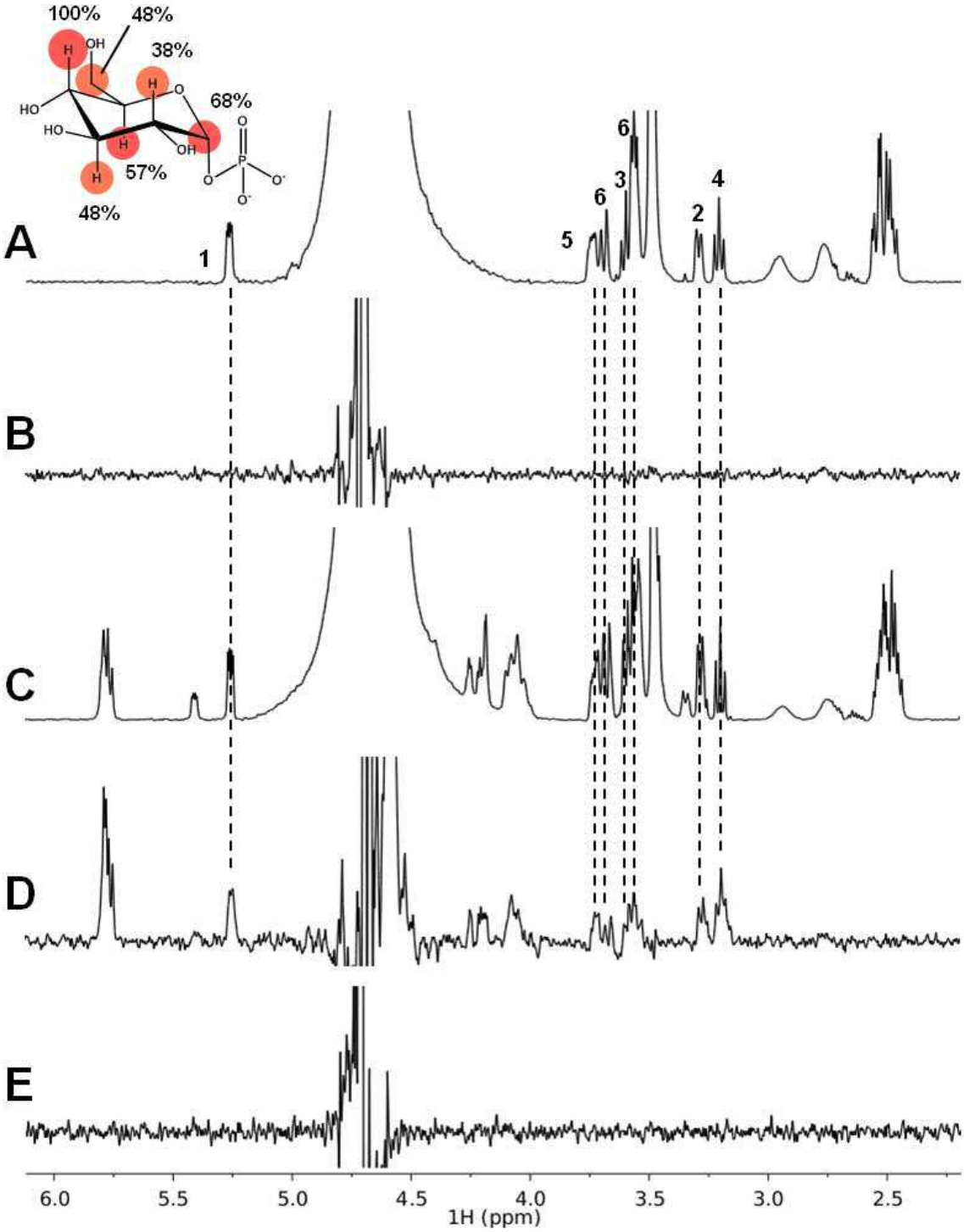


Figure 9: Recognition of glucose-1-P. Off-resonance (A) and STD (B) spectrum of Glc-1-P in the presence of spUGP, with labels indicating the ^{1}H-NMR assignments. Off-resonance (C) and STD (D) spectrum of the same sample, after addition of 1 equivalent UTP, and STD spectrum (E) of this sample after addition of 5 mM EDTA-d_{12}. A molecule of Gcl-1-P is shown, with labels indicating the STD intensity of each signal, relative to the STD intensity of the most saturated signal. Red circles: I > 50%, orange circles: 50% I > 30%. These results indicate that Glc-1-P only binds to spUGP after UTP, and that Mg^{2+} is absolutely required for ligand binding.

The catalytic mechanism followed by NDP-sugar pyrophosphorylases has historically been a subject of controversy. Early insights into the mode of action of a member of the family suggested a sequential ordered Bi-Bi mechanism, in which the substrates enter the active site in an orderly fashion, react, and depart from the enzyme in a precise order. Nonetheless, it was speculated that the phosphorolysis reaction could operate also through a ping-pong mechanism, via a covalent intermediate between the nucleotide and the enzyme [29], a possibility that was discarded by the absence of ^{14}C incorporation to the enzyme from the labeled sugar-phosphate. Speculation persisted, nonetheless, if the second substrate binds and simultaneously reacts (the so-called 'hit and run' mechanism) or binds, and then reacts [14].

Our results point toward a sequential ordered Bi-Bi catalytic mechanism, in which UTP binds to the enzyme first, followed by Glc-1-P binding to the enzyme/nucleotide complex. These results are supported by recent findings using isothermal titration calorimetry (ITC) measurements, together with crystallographic data for a number of pyrophosphorylases [8, 14].

4.2.2 NMR-based enzyme assay

One of the advantages of NMR spectroscopy is that it allows observing individual nuclei of particular molecules, even in complex mixtures of compounds. This feature makes NMR-based assays very convenient to monitor chemical reactions in real-time.

Although most UTP and Glc-1-P signals overlap with those of UDP-Glc, the reaction catalyzed by spUGP could be monitored by looking at the interconversion between the glucose-1-P proton H1 to the UDP-glucose proton H1'' (fig. 10). These signals, which appear as characteristic doublets of doublets, due to ^{3}J-coupling with both the glucosyl H2 and the ^{31}P of the vicinal phosphate, are around 0.15 ppm apart from each other, and well-isolated in the ^{1}H-NMR spectrum.

Focusing on these signals, the reversible interconversion between Glc-1-P and UDP-Glc was monitored inside the NMR tube, both in the forward reaction, i.e. formation of UDP-Glc (fig. 10), and in the reverse reaction, i.e. formation of Glc-1-P and UTP (fig. 11). The NMR-based assay consisted in a cascade of ^{1}H-NMR experiments of 30 s each, until the reaction reached a steady state (ca. 20 min). To help mix the reactants and minimize precipitation of the insoluble magnesium pyrophosphate, the NMR spectra were recorded while spinning the sample at 20 Hz.

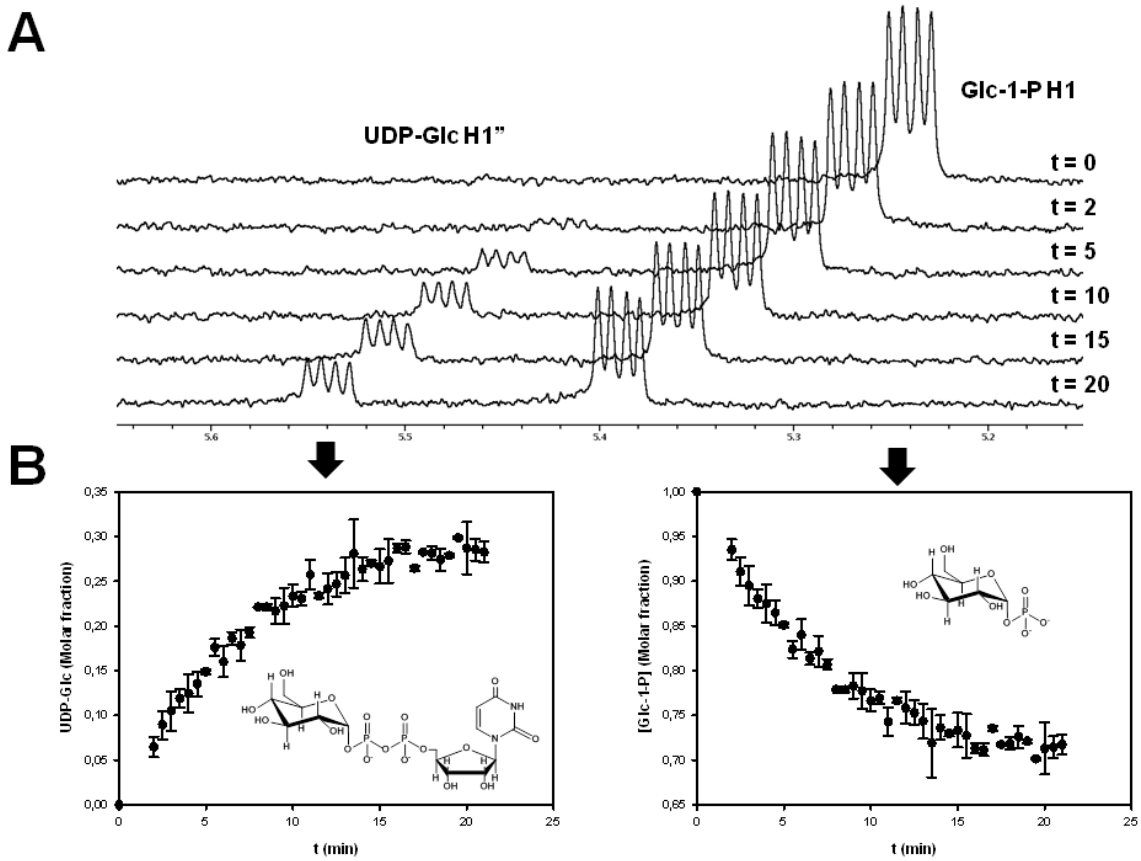


Figure 10: NMR-based spUGP activity assay. Monitoring of the forward reaction (UTP + Glc-1-P → UDP-Glc + PPi). A: Stacked ^{1}H-NMR spectra of a reaction mixture at different reaction times (in min), focusing on the UDP-Glc H1'' and Glc-1-P H1 signals. Reaction mixtures contained 1 mM UTP, 1 mM Glc-1-P, 5 mM $MgCl_2$ and 350 nM spUGP. Spectra at t=0 were obtained before enzyme addition. B: Plots showing substrate-product conversion as a function of time.

The maximum reaction rates for the forward and reverse reactions measured in these assays were 16 nmol/min and 48 nmol, respectively. Considering one unit of activity (U) as the amount of enzyme required to catalyze 1 µmol of substrate per minute, and the amount of enzyme used (6 µg), these values account for specific activities of 2.7 U/mg and 8 U/mg, respectively. The activity measured in these assays is thus comparable to the reported value of 13.6 U/mg for the reverse reaction, using freshly purified spUGP [17].

This NMR-based assay, used here for the first time for the characterization of enzymes from this family, represents a direct approach to monitor their activity, because substrate interconversion is directly observed. Thus, this method is advantageous over the currently most used one for measuring nucleotidyltransferase activity, which relies on its coupling to NADH/NAHD^{+} conversion by UDP-Glc 6-dehydrogenase [6, 17].

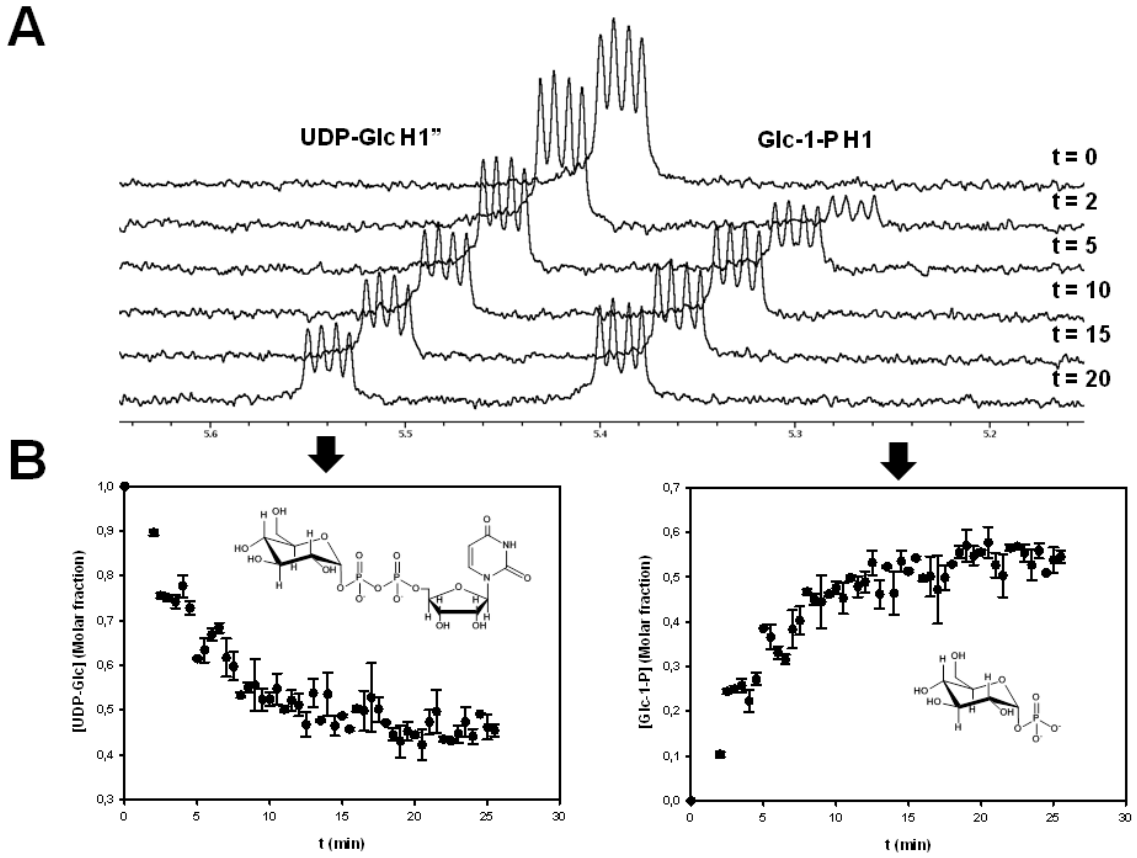


Figure 11: NMR-based spUGP activity assay. Monitoring of the reverse reaction (UDP-Glc + PPi ⟶ UTP + Glc-1-P). A: Stacked ^{1}H-NMR spectra of a reaction mixture at different reaction times (in min), focusing on the UDP-Glc H1'' and Glc-1-P H1 signals. Reaction mixtures contained 1 mM UDP-Glc, 1 mM PPi, 5 mM $MgCl_2$ and 350 nM spUGP. Spectra at t=0 were obtained before enzyme addition. B: Plots showing substrate-product conversion as a function of time.

4.2.3 Fragment-based targeting of spUGP

As already mentioned, the involvement of UGPs in bacterial pathogenicity makes them an excellent target for the discovery of new pharmaceuticals, especially against highly resistant bacterial species, such as *S. pneumoniae*.

A useful approach to target carbohydrate-processing enzymes is by using non-hydrolyzable analogs, such as aza- or carbasugars [30] and exo-glycals [31]. However, since UTP:α-D-glucose-1-phosphate uridylyltransferase activity is key for biological processes also in humans, care must be taken when using this strategy, because the enzymatic activity of the host would likely be inhibited as well. Instead, the lack of evolutionary relation and sequence homology between prokaryotic and eukaryotic UGPs represents an opportunity for lead discovery as result of screening programs to selectively inhibit the bacterial enzyme without problems derived from promiscuity toward the human protein.

We have targeted spUGP through fragment-based drug discovery (FBDD), which represents the first reported efforts to target bacterial UGPs.

In total, 200 molecules were screened against spUGP. Nearly 30% of the tested fragments displayed STD signals. This was correlated with the existence of a big binding pocket (that of the active center), making it prone for fragments to bind therein. Fig. 12 exemplifies the finding of a fragment hit within a mixture. In order to pick only the best performing fragments, a cut-off was applied for hit selection, which left fragments showing an STD response below 12% out of subsequent study. Positive hits satisfying this requirement are shown in Table I.

All hits had in common an elevated hydrophobicity and a similar scaffold, consisting of two non-fused aromatic rings, either five- or six-membered. All of them were shown to interact with protein epitopes in or very close to the active site, as inferred from their ability to compete with UDP-glucose (fig. 13). Besides, in all instances, UTP was shown to compete with these fragments in STD experiments (as exemplified for fragment 275 in fig. 12), indicating that fragment binding occurs at the nucleotide subsite of the active center.

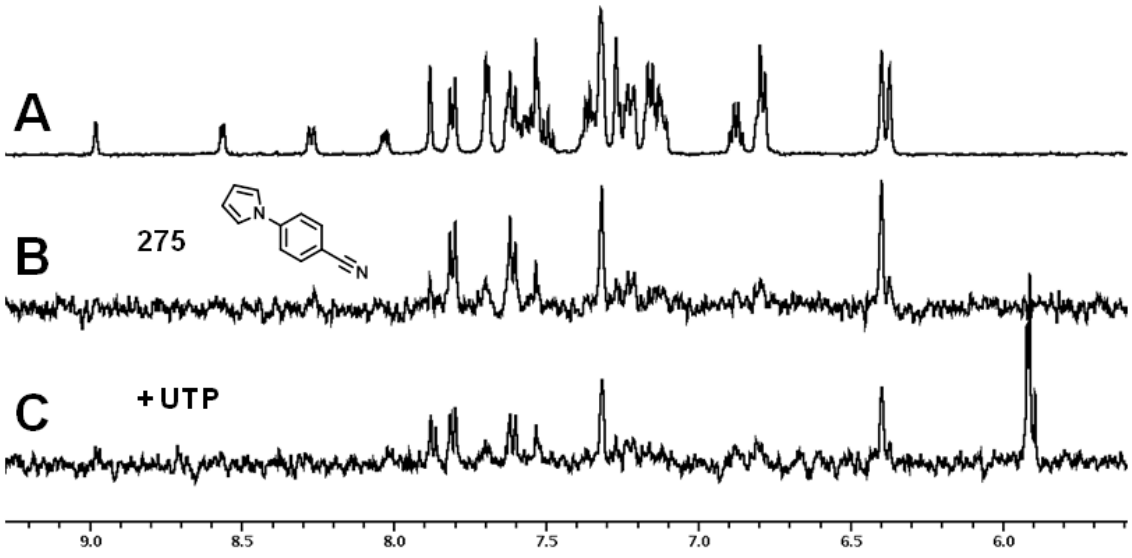


Figure 12: Example of fragment hit occurrence in a mixture targeting spUGP. A: off-resonance spectrum of a fragment mixture (scaled 10%); B: STD spectrum of that fragment mixture, with the structure of fragment 275, identified as the hit. C: STD spectrum after addition of UTP to the sample.

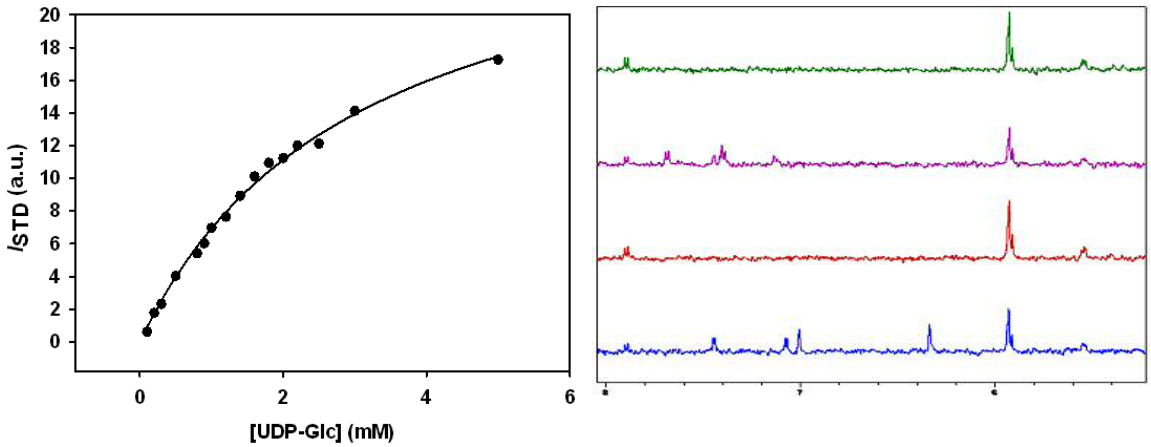


Figure 13: Fragment ranking. Left: STD titration with UDP-Glc. Right: Competition STD experiments with UDP-Glc. Green: 3.2 mM UDP-Glc; purple: + 2 mM 209; red: 3.2 mM UDP-Glc; blue: + 2 mM 358.

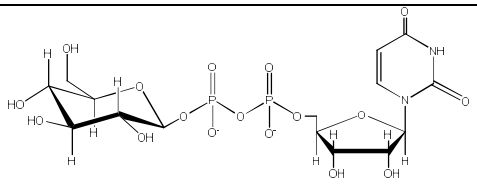
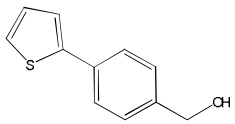
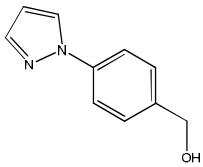
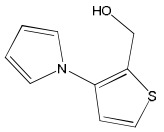
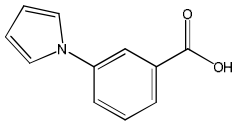
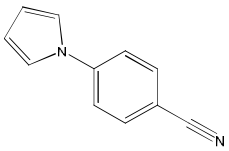
ID	Structure	K_D (mM)	LE (kcal/mol)
UDP-glc		3.2 ± 0.2	0.09
209		0.9	0.32
338		1.2	0.31
358		1.2	0.33
355		2.1	0.26
275		1.1	0.31

Table I: List of fragment hits for spUGP and their estimated K_D and LE values

Dissociation constants (K_D) for these fragment hits were estimated through competition experiments, using UDP-Glc as a spy molecule (Table I). This procedure was advantageous over direct K_D calculation for every molecule, not only because it obviated the need to perform multiple titrations, but also because they in fact permitted to estimate a K_D for such highly hydrophobic compounds. Indeed, in some cases, the obtained K_Ds were higher than their solubility limit.

Because competition experiments are more powerful when using weak-binding spy molecules, they were performed with UDP-Glc in the absence of Mg^{2+} (in fact, with only the amount of Mg^{2+} present in the protein samples, likely coming from the bacterial expression host), which lowered the affinity of spUGP toward UDP-Glc down to the mM level, but still permitted to observe significant STD signals (fig. 13).

In agreement with their small size, fragment hits were estimated to bind to spUGP with rather low affinities (with K_Ds in the low mM range), but with high ligand efficiencies. As expected from these low affinities, enzymatic assays performed in the presence of these fragments led to virtually no difference in the activity profile of the enzyme (fig. 14). However, it is expected that larger, more potent inhibitors will lead to a decrease in spUGP activity observable in the enzyme assay.

These studies are the first reported small-molecule screening against a bacterial UGP and represent the first step toward the targeting of pneumococcal UGP, a therapeutic target candidate.

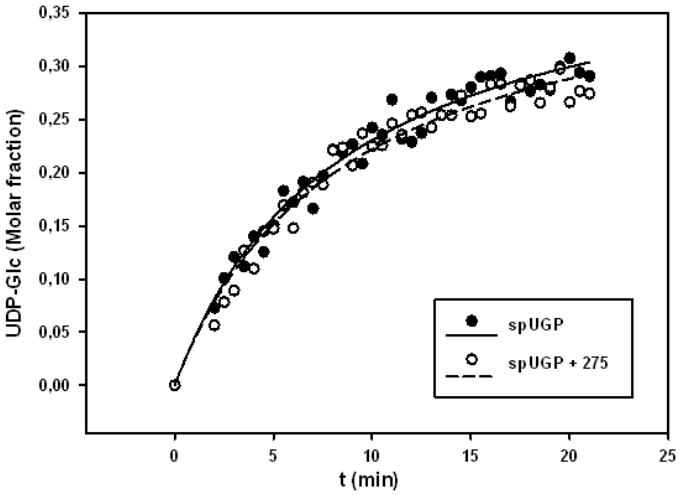


Figure 14: NMR-based spUGP activity assay. Plot showing Glc-1-P to UDP-Glc conversion, as a function of time, by spUGP, both in the absence (filled circles and solid regression curve) and in the presence of 1 mM fragment 275 (open circles and dashed regression curve).

4.3 Methods

STD studies on spUGP substrates

STD experiments were performed on UGP recombinantly produced as previously reported [17]. Experiments were run at 298 K in a Bruker Avance 500 MHz spectrometer, using 1 mM of each substrate and 5 mM $MgCl_2$ in 25 mM Tris-d_6, pH 8.0, 50 mM NaCl in D_2O. The protein was saturated on resonance at -0.5 ppm and off resonance at 100 ppm with a train of 40 selective Gaussian-shaped pulses of 50 ms duration with a 100 µs delay between each pulse. Substrates were purchased from Sigma (St. Louis, MO, USA). UGP samples were generously provided by the laboratory of Prof. Antonio Romero (CIB-CSIC).

NMR-based enzyme assay

The progress of the reaction catalyzed by spUGP was monitored by NMR. Reaction samples within the NMR tubes contained 1 mM of each pair of reactants (UTP and Glc-1-P for the forward reaction; UDP-Glc and PPi for the reverse reaction) and 5 mM $MgCl_2$ in 25 mM Tris-d_6, pH 8.0 in D_2O. Reactions were initiated by the addition of 6 µg (360 nM) recombinant spUGP and subsequently monitored by a cascade of ^{1}H-NMR experiments of 30 s each until reaction reached the equilibrium, in a Bruker Avance 500 MHz spectrometer. Substrate to product conversion was measured by monitoring the interconversion between the glucose-1-P proton H1 to the UDP-glucose proton H1''. One unit of activity was defined as the amount of enzyme required to catalyze 1 µmol of substrate per minute at 25 ^{o}C.

Fragment screening

Mixes of 10 chemically diverse fragments were screened through STD experiments against spUGP in a Bruker Avance 500 MHz spectrometer. Screening mixtures contained 150 µM of each fragment and 5 µM spUGP. Buffer conditions were 25 mM Tris-d_6, pH 8.0, 50 mM NaCl in D_2O. The protein was saturated on resonance at -0.5 ppm and off resonance at 100 ppm with a train of 40 selective Gaussian-shaped pulses of 50 ms duration each.

STD-positive hits were identified from the reference off-resonance spectra and binding was tested independently. Next, they were ranked according to their estimated K_Ds.

K_D estimation

To quantify the affinity of spUGP toward UDP-Glc, a K_D was estimated from titration measurements of the intensity of its STD signal as a function of its concentration.

$$I = \frac{I_{\max} \cdot c}{K_D + c} \quad (1),$$

where I is the observed intensity for the H1 peak in the STD spectrum at a concentration c of UDP-Glc, I_{max} is the maximum intensity and K_D is the dissociation constant. A set of experiments at varying concentrations of UDP-Glc were performed and the K_D was estimated by regression of these data to eq. 1 using SigmaPlot 12.0.

For fragments, affinities were estimated based on competition experiments in the presence of UDP-Glc as a spy ligand. The decrease in the STD signals of UDP-glucose after the addition of each fragment was monitored, and an inhibition constant (K_I) was calculated for each fragment assuming a competitive inhibition model, where:

$$K_D{}^{ap} = K_D\left(1 + \frac{F}{K_I}\right) \quad (2),$$

being $K_D{}^{ap}$ the apparent dissociation constant of UDP-glucose for spUGP observed after the addition of a concentration F of a fragment binding with a dissociation constant K_I. Therefore, the relation between the observed STD signals of UDP-glucose at a concentration L after and before the addition of a fragment could be represented as:

$$\frac{I}{I_0} = \frac{\dfrac{I_{\max} \cdot L}{K_D\left(1 + \dfrac{F}{K_I}\right) + L}}{\dfrac{I_{\max} \cdot L}{K_D + L}} = \frac{K_D + L}{K_D + \dfrac{K_D \cdot F}{K_I} + L} = \frac{K_I \cdot K_D + K_I \cdot L}{K_I \cdot K_D + K_D \cdot F + K_I \cdot L} \quad (3)$$

And, since we used $L = K_D$,

$$\frac{I}{I_0} = \frac{2 \cdot K_I}{2 \cdot K_I + F} \quad (4)$$

A value for I_0 was obtained for UDP-glucose at this concentration in each measure, and a value for I was obtained after the addition of each fragment. Next, K_Is were calculated from eq. 4.

4.4 References

[1] Ginsburg, V. (1978), *Prog Clin Biol Res* **23**, 595-600.

[2] Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008), *Annu Rev Biochem* **77**, 521-555.

[3] Flores-Diaz, M., Alape-Giron, A., Persson, B., Pollesello, P., Moos, M., von Eichel-Streiber, C., Thelestam, M., and Florin, I. (1997), *J Biol Chem* **272**, 23784-23791.

[4] Roeben, A., Plitzko, J. M., Korner, R., Bottcher, U. M., Siegers, K., Hayer-Hartl, M., and Bracher, A. (2006), *J Mol Biol* **364**, 551-560.

[5] Kleczkowski, L. A., Geisler, M., Ciereszko, I., and Johansson, H. (2004), *Plant Physiol* **134**, 912-918.

[6] Lai, X., Wu, J., Chen, S., Zhang, X., and Wang, H. (2008), *Protein Expr Purif* **61**, 50-56.

[7] Daude, N., Gallaher, T. K., Zeschnigk, M., Starzinski-Powitz, A., Petry, K. G., Haworth, I. S., and Reichardt, J. K. (1995), *Biochem Mol Med* **56**, 1-7.

[8] Kim, H., Choi, J., Kim, T., Lokanath, N. K., Ha, S. C., Suh, S. W., Hwang, H. Y., and Kim, K. K. (2010), *Mol Cells* **29**, 397-405.

[9] Hartman, H., and Fedorov, A. (2002), *Proc Natl Acad Sci U S A* **99**, 1420-1425.

[10] Garcia, E., Llull, D., Munoz, R., Mollerach, M., and Lopez, R. (2000), *Res Microbiol* **151**, 429-435.

[11] Thoden, J. B., and Holden, H. M. (2007), *Protein Sci* **16**, 432-440.

[12] Thoden, J. B., and Holden, H. M. (2007), *Protein Sci* **16**, 1379-1388.

[13] Aragao, D., Fialho, A. M., Marques, A. R., Mitchell, E. P., Sa-Correia, I., and Frazao, C. (2007), *J Bacteriol* **189**, 4520-4528.

[14] Blankenfeldt, W., Asuncion, M., Lam, J. S., and Naismith, J. H. (2000), *EMBO J* **19**, 6652-6663.

[15] del Carmen Fernandez-Alonso, M., Diaz, D., Berbis, M. A., Marcelo, F., Canada, J., and Jimenez-Barbero, J. (2012), *Curr Protein Pept Sci* **13**, 816-830.

[16] Sivaraman, J., Sauve, V., Matte, A., and Cygler, M. (2002), *J Biol Chem* **277**, 44214-44219.

[17] Bonofiglio, L., Garcia, E., and Mollerach, M. (2005), *Curr Microbiol* **51**, 217-221.

[18] Luo, H., Lin, Y., Gao, F., Zhang, C. T., and Zhang, R. (2014), *Nucleic Acids Res* **42**, D574-580.

[19] Bentley, S. D., Aanensen, D. M., Mavroidi, A., Saunders, D., Rabbinowitsch, E., Collins, M., Donohoe, K., Harris, D., Murphy, L., Quail, M. A., Samuel, G., Skovsted, I. C., Kaltoft, M. S., Barrell, B., Reeves, P. R., Parkhill, J., and Spratt, B. G. (2006), *PLoS Genet* **2**, e31.

[20] Mollerach, M., and Garcia, E. (2000), *Gene* **260**, 77-86.

[21] Bonofiglio, L., Garcia, E., and Mollerach, M. (2012), *FEMS Microbiol Lett* **332**, 47-53.
[22] Mollerach, M., Lopez, R., and Garcia, E. (1998), *J Exp Med* **188**, 2047-2056.
[23] Komeda, Y., Icho, T., and Iino, T. (1977), *J Bacteriol* **129**, 908-915.
[24] Deng, W. L., Lin, Y. C., Lin, R. H., Wei, C. F., Huang, Y. C., Peng, H. L., and Huang, H. C. (2010), *Mol Plant Microbe Interact* **23**, 1184-1196.
[25] Chang, H. Y., Lee, J. H., Deng, W. L., Fu, T. F., and Peng, H. L. (1996), *Microb Pathog* **20**, 255-261.
[26] Nesper, J., Lauriano, C. M., Klose, K. E., Kapfhammer, D., Kraiss, A., and Reidl, J. (2001), *Infect Immun* **69**, 435-445.
[27] Priebe, G. P., Dean, C. R., Zaidi, T., Meluleni, G. J., Coleman, F. T., Coutinho, Y. S., Noto, M. J., Urban, T. A., Pier, G. B., and Goldberg, J. B. (2004), *Infect Immun* **72**, 4224-4232.
[28] Guo, Y., Sagaram, U. S., Kim, J. S., and Wang, N. (2010), *Appl Environ Microbiol* **76**, 2234-2242.
[29] Melo, A., and Glaser, L. (1965), *J Biol Chem* **240**, 398-405.
[30] Lahiri, R., Ansari, A. A., and Vankar, Y. D. (2013), *Chem Soc Rev* **42**, 5102-5118.
[31] Lin, C. H., Lin, H. C., and Yang, W. B. (2005), *Curr Top Med Chem* **5**, 1431-1457.

CHAPTER 5

TARGETING GALECTINS

5.1 Introduction

Complex carbohydrates on cell surfaces are increasingly viewed as biochemical messages read and translated into effects by endogenous lectins [1]. Among them, the members of the galectin family have recently attracted special interest due to their context-dependent multifunctionality and involvement in disease processes, [2] and the possibility of being used as therapeutic targets.

Galectins are a family of animal lectins defined by a shared consensus of amino acid sequences and a carbohydrate recognition domain (CRD) of around 130 amino acids, with affinity for β-galactose containing oligosaccharides. Galectins are best known for mediating adhesion and regulate cell growth (as mitogens or as anti/pro-apoptotic effectors) through binding to cell surface glycoconjugates. However, members of this family are not only expressed extracellularly, but also inside the cell and in the nucleus, and they have been often found to perform moonlighting functions.

The galectin CRD fold is comprised of two antiparallel β-sheets of five and six β-strands, arranged in a β-sheet sandwich motif with a "jelly-roll" topology. The carbohydrate binding site is located within the six-stranded β-sheet face of the β-sheet sandwich, and is composed of conserved amino acid residues, i.e. two arginines, glutamic acid, histidine, and a tryptophan (see fig. 6 in Chapter 1). The β-galactose unit of the bound carbohydrate ligand is positioned between two loops that essentially fold over it, and is maintained in place by multiple hydrogen bonds formed with its O4, O5 and O6 atoms. From one of these loops, the planar aromatic side chain of the conserved tryptophan is positioned under the galactose ring to provide further stabilization for the bound carbohydrate ligand.

Structurally, galectins are divided into three classes. Prototype galectins are composed of a single CRD, which can self-associates to form homodimers (gal-1, -2, -5, -7, -10, -11, -13, -14 and -15. Tandem repeat galectins have two homologous, yet distinct, CRDs that are connected to each other via a linker polypeptide chain (galectin-4, -6, -8, -9 and -12). Finally, chimera-type galectins (galectin-3) are composed of one CRD attached to a non-lectin part.

In this chapter, a series of studies combining ligand- and protein-detected NMR methods is presented, probing the structure and interacting capabilities of two members of the galectin family: galectin-7, a prototype, dimeric galectin, and galectin-3, the only chimera-type galectin identified until now.

5.2 Galectin-7: a prototype galectin

Galectin-7 (gal-7) is a prototype, dimeric galectin. Also referred to as the p53-induced gene 1 product based on its up-regulation by the tumor suppressor p53 in human DLD-1 colon cancer cells [3], gal-7 is a potent pro-apoptotic effector, e.g. for neuroblastoma cells by binding ganglioside GM1 [4]. Gal-7 also plays a key role in the development and differentiation of the pluristratified epithelium, and has a complex, dual role in tumor progression [5], depending on the tissue type: while gal-7 overexpression in sunburn keratinocytes is co-related with an increased apoptosis, gal-7 induces expression of MMP-9 in lymphoma cells [6].

Several crystallographic structures have been reported for gal-7 dimers, both unliganded and in complex with simple carbohydrates, like lactose and *N*-acetyl lactosamine. In addition to binding glycans, gal-7 can interact with the anti-apoptotic protein bcl-2, a property shared with gal-3 [7].

Studies presented in this section, initiated by full NMR assignment of this galectin, are directed toward the characterization of the structure and dynamics of the gal-7 homodimer. These results are meant to pave the way for the study of interacting capabilities of this protein, both at and beyond the carbohydrate binding site.

5.2.1 Full NMR assignment of galectin-7

Complete NMR assignment of human galectin-7 was achieved in collaboration with the group of Prof. Kevin H. Mayo (University of Minnesota, MN, USA). Our contribution consisted in the collection of backbone and side-chain carbon and proton NMR assignments through a set of homo- and heteronuclear 2D experiments, as well as heteronuclear triple resonance 3D experiments carried out at 800 MHz.

Specifically, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ resonances were sequentially obtained through 3D HNCACB experiments, as exemplified in fig. 1. Their correspondent protons were identified in ^{1}H-^{13}C HSQC spectra. In addition, 3D HNCO experiments afforded the assignment of all ^{13}CO resonances. Besides, several side-chain resonances were identified based on ^{1}H-^{1}H NOESY spectra recorded with different mixing times.

In total, 1,045 ^{13}C, 264 ^{15}N and 1,624 ^{1}H chemical shifts were unambiguously assigned. Backbone resonances (^{1}HN, ^{15}N, ^{13}CO, $^{13}C_{\alpha}$ and $^{1}H_{\alpha}$) were fully assigned with the exception of the first two residues (simply not observed in the ^{1}H-^{15}N HSQC spectrum) and for $^{1}H_{\alpha}$ of Pro15, whose signal was completely overlapped with the water signal. In addition, sequential connectivity of the Pro26-Pro27 pair could not be established. Furthermore, more than 80% of side-chain resonances were assigned, and stereospecific assignments were made for all Val and Leu methyl groups. These chemical shift assignments were deposited in the BioMagResbank [8] under the accession number 17826, and published in the *Biomol NMR Assign* journal [9].

Figure 2 shows a ^{1}H-^{15}N HSQC spectrum of gal-7, in which all cross-peaks are labeled according to these NMR assignments. The spectrum shows doubling of several cross-peaks, which are due to the co-existence of two distinct conformations of gal-7 in solution, in slow exchange on the chemical shift timescale. The nature of this conformational duality will be presented in the next section.

5.2.2 The conformational duality of gal-7

As seen in the previous section, assignment of the gal-7 resonances revealed the doubling of peaks for a number of residues distributed along the whole gal-7 sequence, compatible with the presence of two stable conformations in slow exchange on the chemical shift timescale.

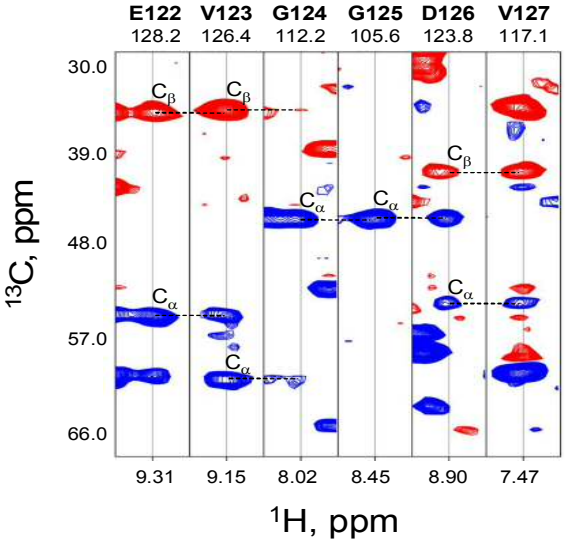


Figure 1: A series of strip plots from an HNCACB experiment taken at the amide ^{15}N and ^{1}H chemical shifts of residues E122-V127. Chemical shifts of C_α (blue cross-peaks) and C_β (red cross-peaks) atoms are indicated on the vertical axis. ^{1}H chemical shifts are indicated on the horizontal axis. Horizontal dotted lines indicate sequential connectivity.

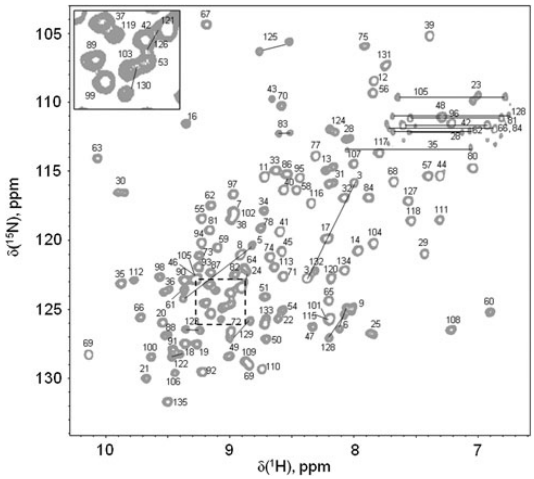


Figure 2: ^{1}H-^{15}N HSQC spectrum of gal-7. The insert shows an expansion of a congested region of the spectrum. Horizontal lines connect resonances arising from side-chain groups of Asn and Gln residues. Other lines shown connect cross-peaks that are doubled in the spectrum due to the presence of two conformations in slow exchange on the chemical shift scale.

The number of affected residues is 22, which accounts for 18% of the gal-7 amino acids observable in the ^{1}H-^{15}N HSQC spectrum (fig. 3). As shown in fig. 4, these residues cover a large portion of the gal-7 surface, and they are distributed all along the gal-7 sequence, although they appear mainly concentrated at the N- and C-termini. There are several modifications that the N-terminus of even recombinantly expressed proteins can suffer in *E. coli*, including deformylation and enzymatic cleavage of the initiating *N*-formylmethionine, which can result in the presence of a heterogeneous population of the recombinant protein. This may lead to spectral changes, including splitting of peaks belonging to the N-terminal amino acids and those of nearby regions [10]. To rule out the possibility that doubling of gal-7 peaks might be due to a heterogeneous modification of the N-terminus in the bacterial expression system, we performed mass spectrometry measurements on our gal-7 samples. The result confirmed the

homogeneous presence of a single major species corresponding to a molar mass of 14946.9 Da (fig. S4 in Appendix), close to the theoretical value of 14943.8 Da for the gal-7 sequence without the initial methionine, thus ruling out the presence of a heterogeneous protein in our samples, and confirming that the spectral duality of gal-7 is due to the coexistence of two stable conformations of the same protein, in slow exchange on the chemical shift timescale.

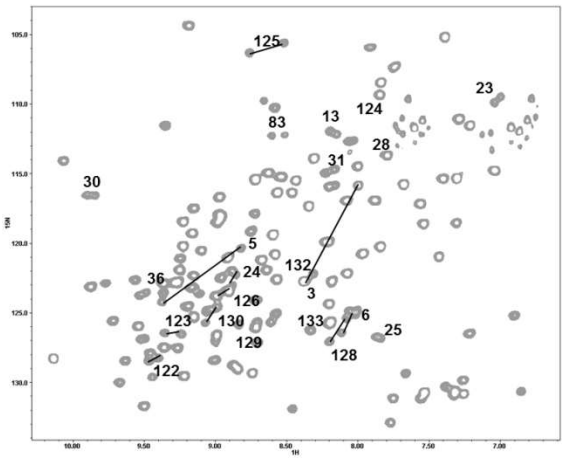


Figure 3: ^{1}H-^{15}N HSQC spectrum of human gal-7, with connectors and labels indicating the doubled peaks.

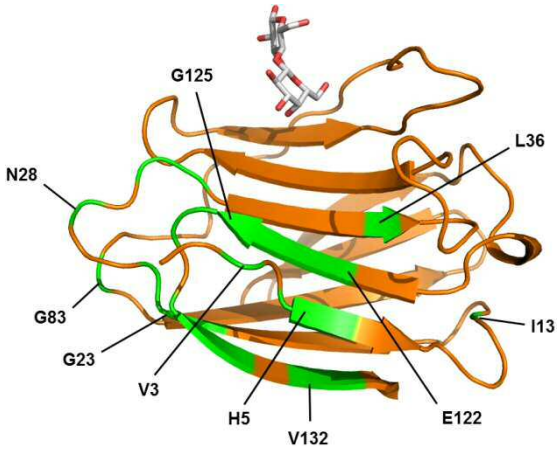


Figure 4: Representation of one of the monomers in the X-ray structure of lactose-bound gal-7 (PDB: 4GAL). Residues with doubled signals are highlighted in green.

5.2.2.1 Pro4 is directly responsible for the conformational duality

The presence of two populations in slow chemical exchange requires that the energy barrier ($\Delta G^{\ddagger}$) between both conformers is high. The Eyring equation [11] permits to calculate the barrier between two species at coalescence:

$$k_c = \frac{k_B T_C}{h} e^{\frac{-\Delta G}{RT_C}} \quad (1)$$

where k_c is the coalescence rate, calculated as:

$$k_c = \frac{\pi \Delta\nu}{\sqrt{2}} \cong 2.22\ \Delta\nu \quad (2)$$

being Δν the frequency separation between the two signals, in Hz. Since, for gal-7, the situation is still far from coalescence, application of the Eyring equation provides that the energetic barrier between both gal-7 conformers is $\Delta G^{\ddagger}$ > 17 kcal/mol. Such high a barrier indicates a slow process, the most typical of which in the realm of proteins is proline *cis*/*trans* isomerization. Indeed, the activation energy for the isomerization of the Xaa-Pro amide is about 20 kcal/mol, due to the highly energetic *syn*-Pro transition state [12].

Galectin-7 has 10 prolines scattered all along its sequence, but isomerization of Pro4 is the one which would best explain the region affected by doubling of peaks. To test this hypothesis, the gal-7 P4L mutant was constructed in the laboratory of Prof. Hans-Joachim Gabius (Ludwig Maximilian University, Munich). The ^{1}H-^{15}N HSQC spectrum of the uniformly ^{15}N-labeled mutant protein is highly superimposable with that of the wild type, indicating that the overall folding of the mutant is maintained, but it completely lacks one set of the doubled peaks (fig. 5), thus confirming that Pro4 is the direct responsible of the gal-7 conformational duality.

5.2.2.1 Fine characterization of the gal-7 *cis*-conformer

The observation that two conformers of gal-7, caused by a switch in the V3-P4 amide bond, coexist in solution, is in contrast with the configuration of that peptide bond in both monomers of all seven gal-7 dimer X-ray structures released to date (1BKZ, 2GAL, 3GAL, 4GAL, 5GAL, 3ZXE, 3ZXF). In all crystallographic structures, the gal-7 N-terminus is found in an “open” conformation, protruding from the β-sandwich. As a result, there is a spatial proximity between the Asn2 sidechain and the β-strand 9 that results in the establishment of a hydrogen bond between the Asn2 side-chain and the Gly125 backbone NH (fig. 6).

In order to gain insight into the hydrogen bond pattern of both conformers, the HN temperature coefficients were obtained from a temperature titration, by running ^{1}H-^{15}N HSQC spectra of gal-7 in a range of temperatures between 292 K and 313 K (fig. 7). Detailed quantification of the correlation between amide proton temperature coefficients ($\Delta\delta_{HN}/\Delta T$) have shown a good prediction value for a hydrogen bond for temperature coefficients smaller than 4.6 ppb/K [13]. The titration showed a differential shifting among the cross-peaks, with a

number of signals showing poor shifting along with the temperature gradient and thus likely being involved in hydrogen bonds.

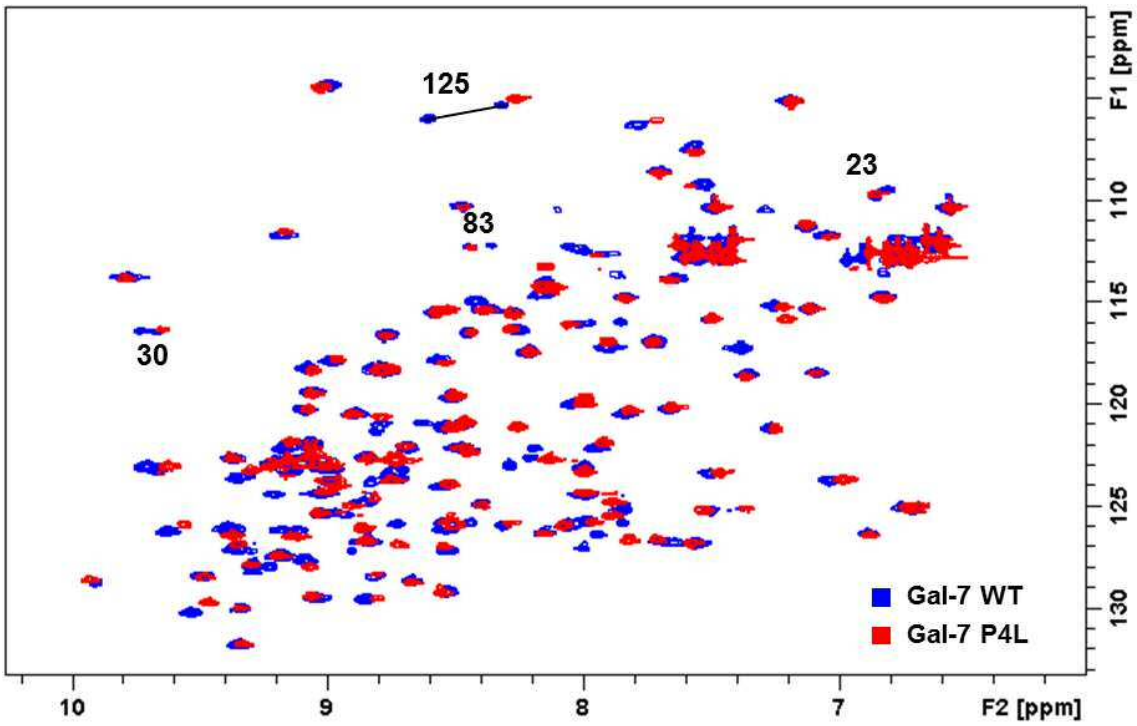


Figure 5: ^{1}H-^{15}N HSQC spectra of human gal-7. Blue: wild-type. Red: P4L mutant. Some of the doubled peaks in the wild-type spectrum are labeled for reference.

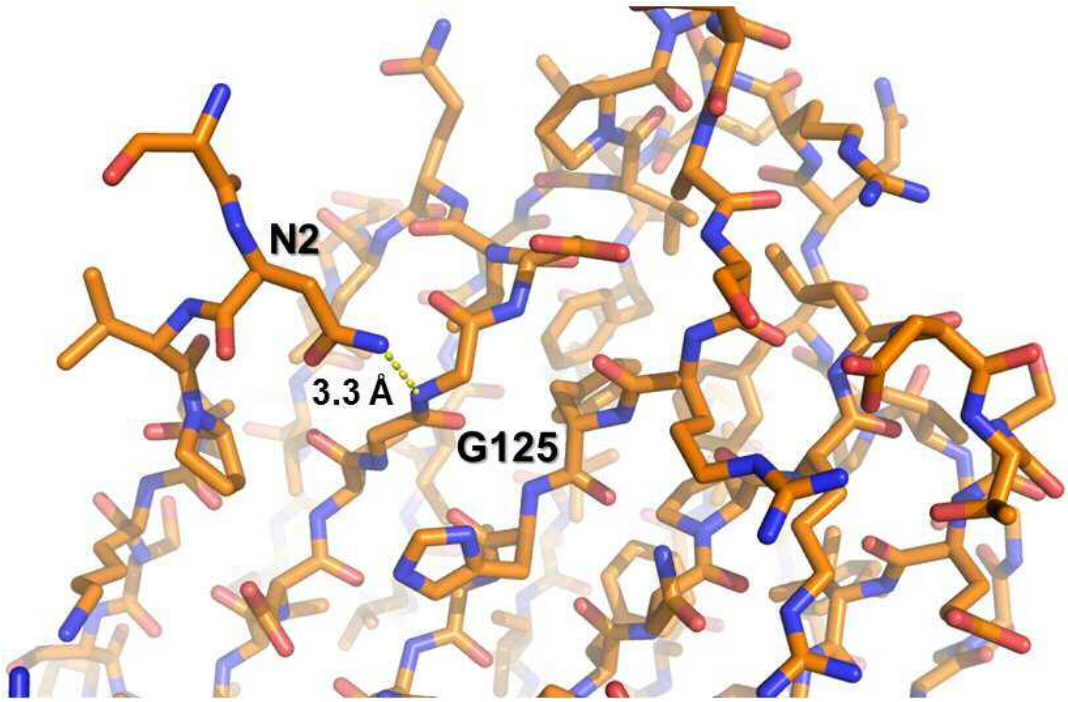


Figure 6: Detail from the X-ray structure of human gal-7 (PDB: 4GAL), in which the *trans*-conformation for P4 is present. This figure shows the establishment of a hydrogen bond between the G125 amide and the N2 side-chain.

Focusing on the G125 signal HN pair, the peak belonging to the *trans*- conformation remained stable over the whole temperature range investigated, with $\Delta\delta_{HN}/\Delta T$ = –2.26 ppb/K, in agreement with the hydrogen bond observed in the crystallographic structure. In contrast, the peak from the *cis*- conformation showed a temperature coefficient of –6.63 ppb/K, which accounts for the loss of the G125/N2 hydrogen bond in the cis- conformer. No other significant differences were found for the other doubled signals.

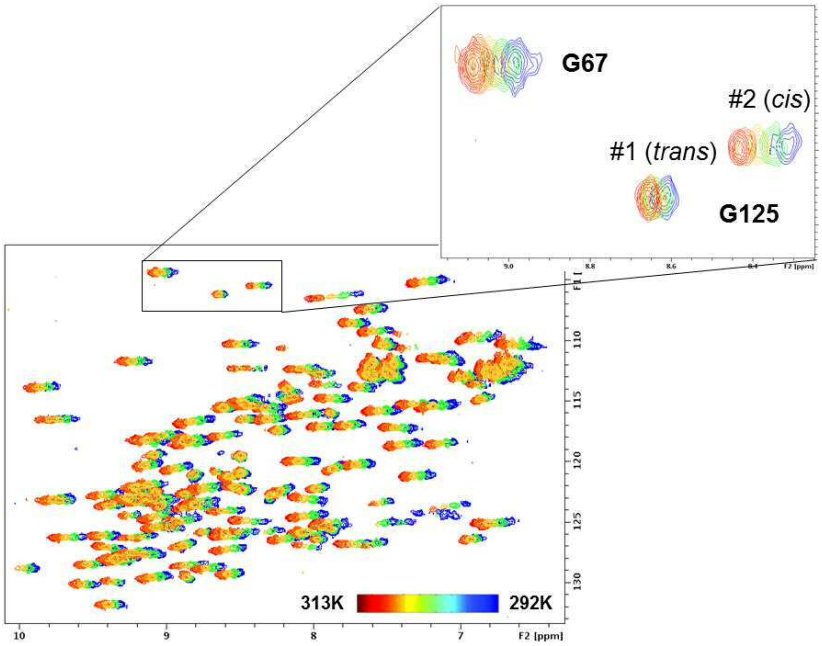


Figure 7: Overlay of ^{1}H-^{15}N HSQC spectra of gal-7 corresponding to a temperature titration. The less shifted are the signals, the higher is the probability that NHs are involved in a hydrogen bond, with a good predictive value for a hydrogen bond for temperature coefficients less negative than –4.6 ppb/K [13]. Temperature coefficients for the *trans*- and *cis*- signals of the G125 pair are –2.26 ppb/K and –6.63 ppb/K, respectively, which is in agreement with the loss of a hydrogen bond between G125 and N2 in the *cis*- conformer.

Further corroboration to this hypothesis was found by probing the gal-7 solvent accessible NH groups with dimethyl sulfoxide (DMSO). Those protein NHs that are solvent accessible can show spectral shifts upon perturbation of the solvent properties, e.g. its dielectric constant via the addition of a small quantity of an organic solvent, such as DMSO [14], while the rest are unaffected. Addition of 2.5% to 7.5% DMSO to gal-7 resulted in the shift of a number of peaks, which permitted to map the gal-7 NHs exposed to the solvent. Of note, the G125 NH peak of the *cis*-conformation showed a significant shift in the presence of DMSO (fig. 8A), while the same peak of the *trans*-conformation remained unaffected. This is in agreement with the burying of G125 by the N2 side-chain, as a result of the hydrogen bond established between the two, in the *trans*-conformation. In contrast, in the *cis*-conformation, the N-terminus is no longer close to the vicinal strand, and the latter becomes solvent-accessible (fig. 8B).

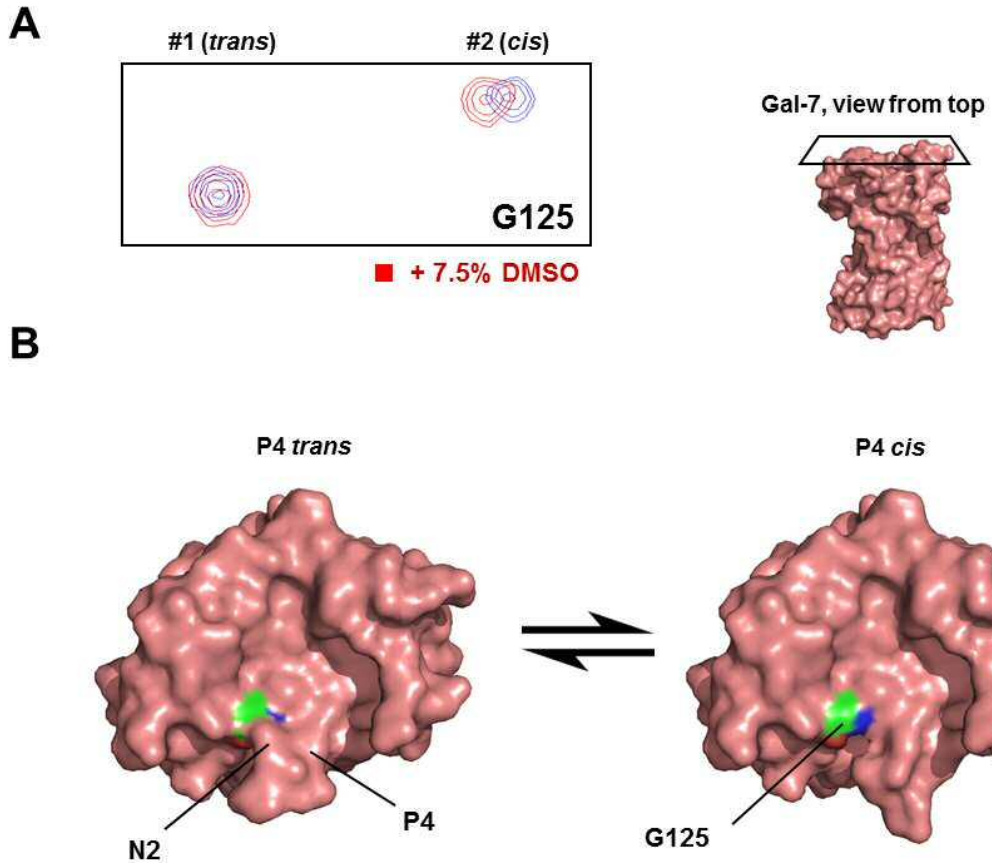


Figure 8: Differential solvent-accessibility of the G125 amide pair in either conformer. A: the addition of 7.5% DMSO to the sample selectively perturbs the G125 amide signal from the *cis*- conformation, remaining that from the *trans*- unaffected. B: Accordingly, the G125 amide nitrogen appears "capped" by the N2 side-chain in the *trans*- conformation present in the crystal (PDB: 4GAL), while in the model for the *cis*- conformation it becomes fully solvent-accessible.

Next, a new mutant of gal-7 was prepared, by mutating the initial Ser into Ala (Gal-7 S1A). Since the orientation of the N-terminus is one of the most distinctive features between both conformers of gal-7, a change in the chemical nature of the first amino acid could be useful to map the region which this terminus points to in each conformer. Figure 9 shows a ^{1}H-^{15}N HSQC spectrum of the S1A mutant. Doubling of peaks persisted in the same manner and extent, with no perceptible changes in the ratio between conformers. In addition, the spectrum was virtually identical to that of wild type gal-7, with minor shifts for a small number of peaks. With the exception of those of N2 and V3, the rest of the shifted peaks belong to amino acids which are distant in the gal-7 sequence to the point mutation. Such is the case for R22, L129 and V132, which lie in a hydrophobic pocket close to the dimer interface in the gal-7 3D structure.

Noteworthy, only the peaks belonging to the *cis*-conformation are shifted, indicating that, in this conformation, the N-terminus is close to those residues.

A model for the gal-7 *cis*- conformer was built by a manual 180° torsion of the V3-P4 peptide bond in the apo-gal-7 crystal structure (PDB: 1BKZ), followed by an energy minimization in an OPLS-2005 force field (fig. 10). This model is in agreement with the experimental data in that, upon V3-P4 isomerization, the N-terminus moves away from the vicinity of G125, towards a spatial area close to L129 and V132.

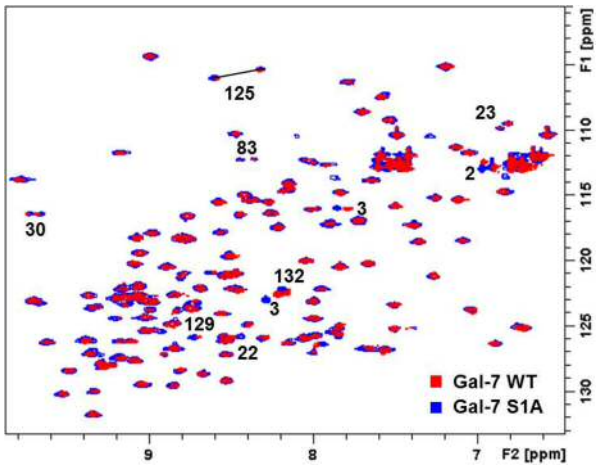


Figure 9: ^{1}H-^{15}N HSQC spectra of human gal-7. Blue: wild-type. Red: A1S mutant. Some of the doubled peaks as well as significantly perturbed signals (located at a hydrophobic patch around L129 and V132) are labeled.

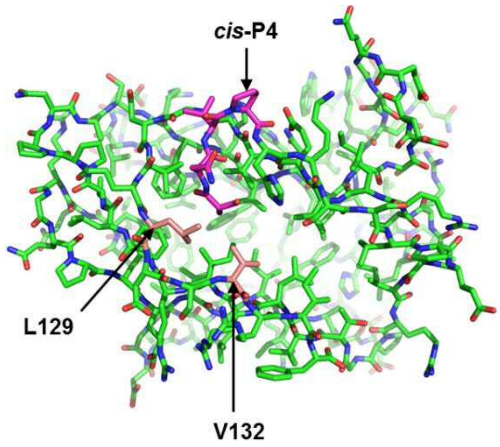


Figure 10: Model for the gal-7 P4 *cis* conformer, generated by a manual 180° torsion of the V3-P4 bond followed by a MM minimization. The N-terminus (pink) points toward a hydrophobic cluster close to residues L129 and V132 (salmon).

5.2.2.2 The gal-7 *cis*-conformation is stabilized by non-polar forces

For non-proline residues, the overwhelming majority of peptide bonds adopt the *trans* configuration, mainly because the steric conflicts between the functional groups attached to the C_α atoms are larger in the *cis* isomer. For proline residues, however, the cyclic nature of its side-chain causes that both conformers are nearly isoenergetic, and it is common for short peptides to populate both isomers under unstrained conditions. Still, the *trans*- isomer is

slightly more favorable, and their relative populations tend to favor it. The exact ratio depends on the substituents, especially on the bulkiness of the side-chain of the preceding amino acid, with aromatic residues slightly favoring the *cis*- isomer. Valine, which is the preceding amino acid to P4, can be considered rather bulky, with a "bulkiness factor" (defined as the ratio of the side-chain volume to its length [15]) of 21.57 $Å^2$, comparable to that of Trp (21.67 $Å^2$) and substantially larger than those of Thr (15.77 $Å^2$) or Pro (17.43 $Å^2$), with similar molecular weights. Accordingly, previous data with the model dipeptide Val-Pro [16] showed that, at physiological pH, the *cis*- and *trans*- isomers were nearly equally populated (40:60, respectively). However, the ratios were found to be strongly pH-dependent, with low pH favoring the *trans*- isomer due to stronger steric repulsions between the protonated amino group and the carboxylic group in the *trans*- isoform of the cationic dipeptide.

In order to evaluate the intrinsic proneness of P4 to populate the *cis* isomer under unstrained conditions, but in a sequence longer than a dipeptide, and in which the rest of the nearby substituents could play a role comparable to that in the full protein, a nonapeptide was synthesized, corresponding to the most N-terminal portion of gal-7 (SNVPHKSSL). The synthesis of this molecule was carried out in the laboratory of Prof. M. Reza Ghadiri, during a short stay at The Scripps Research Institute (La Jolla, CA), financed by the FPI Fellowship. The aromatic and amide region of the ^{1}H-NMR spectrum of the nonapeptide is shown in fig. 11A. A single main set of signals was found, including 7 amide protons (one for each residue with the exception of proline), the side-chain amides of N2, the Ser OH's and two sharp singlets belonging to the H5 side-chain. Although a second, minor set of signals attributable to a second conformer can be seen, its population relative to the major species is < 10%, thus far from the 55:45 proportions in the full-length gal-7. Thus, no doubling of peaks as a result of Pro isomerization was observed. To determine which isomer (*cis* or *trans*) was the one present in solution, the NOE pattern of the V3 α-proton was analyzed in a ROESY (Rotating frame Overhauser Effect Spectroscopy) experiment. As illustrated in fig. 11B, NOE cross-peaks between the α-proton of the preceding amino acid, and the proline α-proton or δ-protons, are diagnostic for the *cis*- or the *trans*- configuration, respectively. The ROESY spectrum of the synthetic nonapeptide revealed NOE cross-peaks between the V3 α-proton and the P4 δ-protons (fig. 11C), which are unambiguously indicative for a *trans* configuration (fig. 11D). Taken together, these results indicate that the gal-7 N-terminal sequence has no natural tendency to isomerize, and thus the *cis*-isomer present in the folded protein in solution is the result of stabilizing forces in the native-state macromolecule.

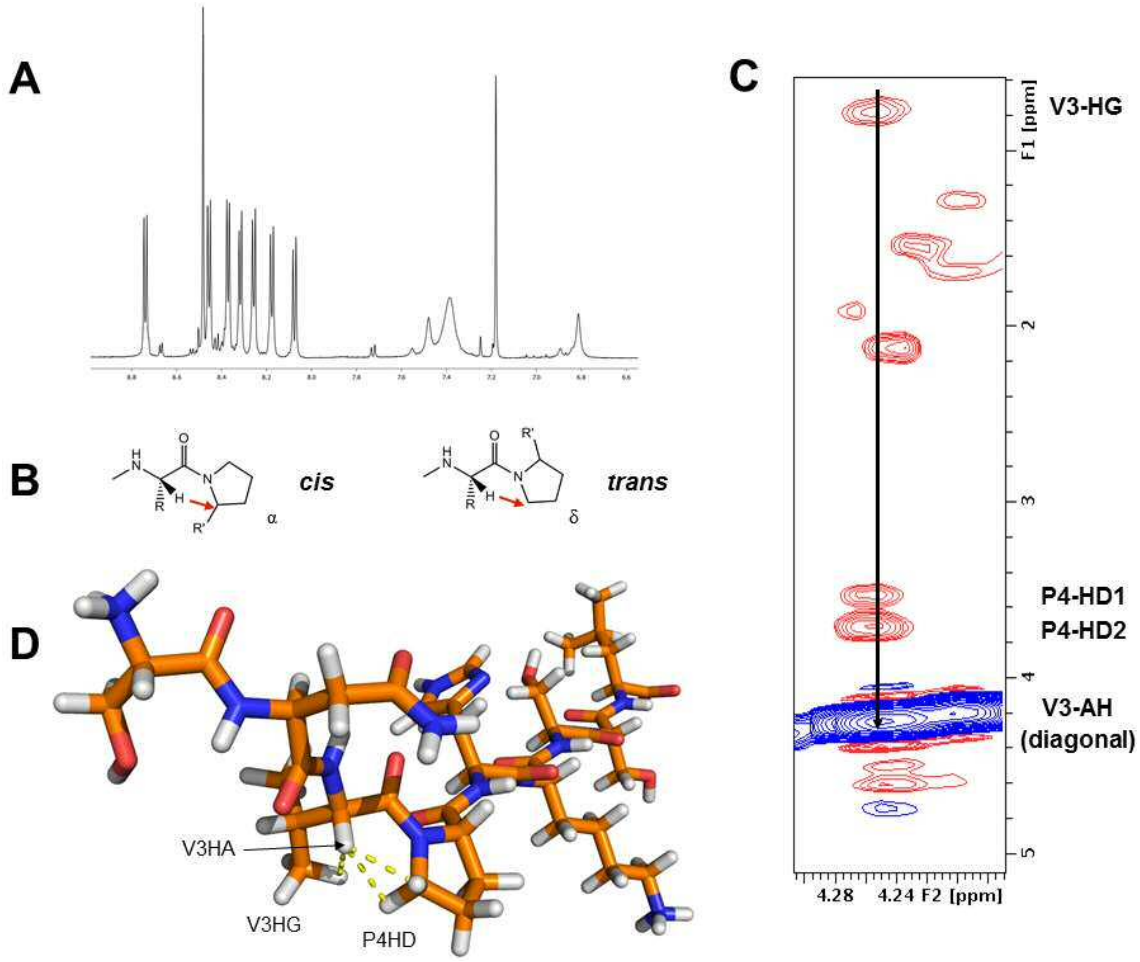


Figure 11: A synthetic nonapeptide corresponding to the N-terminus of human gal-7 (SNVPHKSSL) does not have a natural tendency favorable to yield the *cis*- isomer. A: ^{1}H-NMR spectrum of the nonapeptide in H_2O, revealing the existence of only one major conformation. B: Scheme showing the differential NOE pattern (denoted by red arrows) for the α-proton of the residue preceding proline in each configuration of the peptidyl-prolyl peptide bond. C: ^{1}H-^{1}H ROESY spectrum of the synthetic nonapeptide, showing the NOE pattern for the V3 α-proton, which is unambiguosly indicative for a *trans*-conformation of the V3-P4 peptide bond. D: Model for the nonapeptide built in xleap and subject to an energy minimization, showing the expected NOEs for the V3 α-proton in the *trans*-conformation, also found in the ^{1}H-^{1}H ROESY spectrum.

However, and since there is a penalty for the intrinsically less favorable *cis*-isomer, which adds to the loss of one hydrogen bond between N2 and G125 in the gal-7 *cis*-conformer, the question arises how it is possible that the latter exists in the protein in solution, or, in other words, which are the forces that compensate the added penalties of i) the *cis*- configuration of the V3-P4 peptide bond and ii) the loss of a hydrogen bond between N2 and G125.

In agreement with the experimental data, our model for the gal-7 P4 *cis* conformer missed the hydrogen bond between N2 and G125 (fig. 12, left panel), but in turn it gained close contacts between the N-terminus and residues from a hydrophobic patch close to the C-terminus (fig. 12, right panel). Thus, the electrostatic and torsional energy penalties in the P4 *cis* model are compensated by van der Waals interactions, leading to two nearly isoenergetic states that co-exist in solution.

5.2.2.3 Functional implications

Galectin-7 first appeared in mammals, likely as a result of a gene duplication from a common ancestor to gal-10, with which gal-7 shares the highest sequence similarity [17]. As can be seen in the alignment in fig. 13, the switchable proline is conserved throughout the entire mammalian lineage, with the exception of rodents (*Rattus norvegicus* and *Mus musculus*). Notably, human galectin-10 also possesses the conserved proline residue (P6), which is found in the *cis*- configuration in its X-ray structure (PDB: 1LCL). This configuration shares many similarities with our model for the gal-7 P4 *cis* structure (fig. 14), including the packing of the valine residue preceding the conserved proline towards a hydrophobic patch around residues 126-132. However, and since the N-terminus of gal-10 is a few residues longer than that of gal-7, there are further contacts that contribute to stabilize the structure, including one CH-π interaction between L3 and Y8 and one hydrogen bond between S1 and L132. This could explain that the *cis*- conformer of gal-10 is the energetically most favorable, and thus the one found in the crystals.

The conformational duality of gal-7 could be of functional relevance, since the area affected by the proline switch does not lie far from the sugar recognition site: G125 is only 10 Å from the galactose residue. More complex oligosaccharides, glycoproteins or even other types of protein partners could be selectively recognized by one of the two conformers.

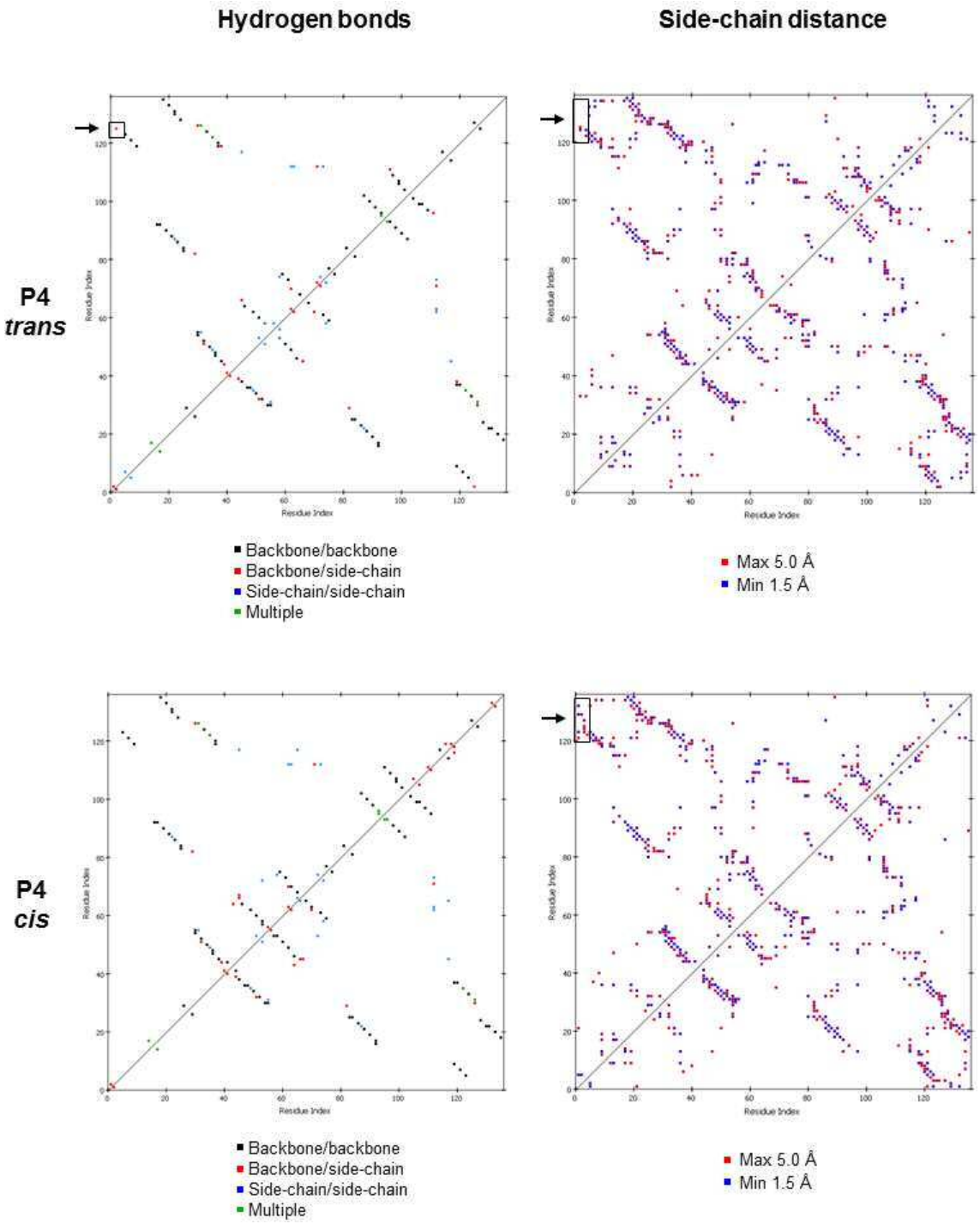


Figure 12: Hydrogen bonds and side-chain contacts in the gal-7 P4 *trans* (upper panel) and P4 *cis* (lower panel) conformers. The loss of a hydrogen bond between N2 and G125 is compensated between close contacts between the N-terminus and a hydrophobic cluster of amino acids on the C-terminus.

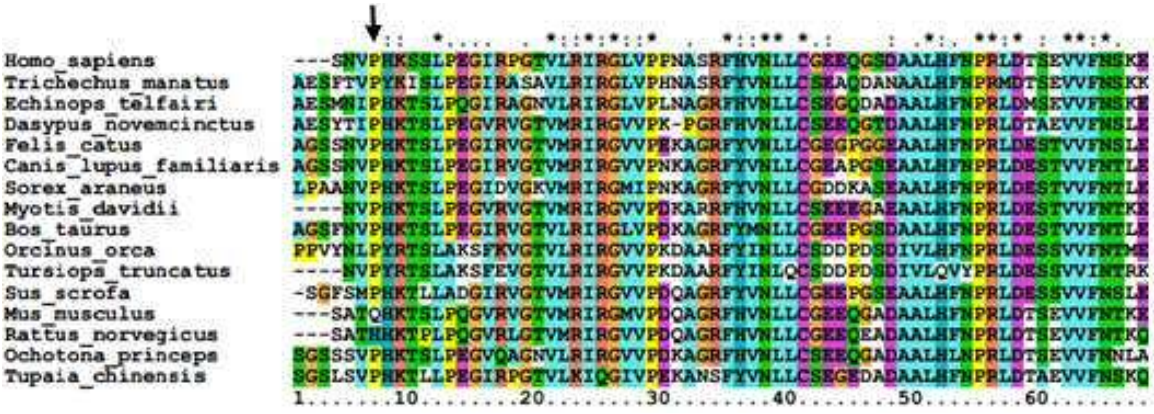


Figure 13: Alignment of gal-7 sequences from different mammal species. P4 (indicated by an arrow) is conserved in all cases, with the exception of rodents *Rattus norvegicus* and *Mus musculus*.

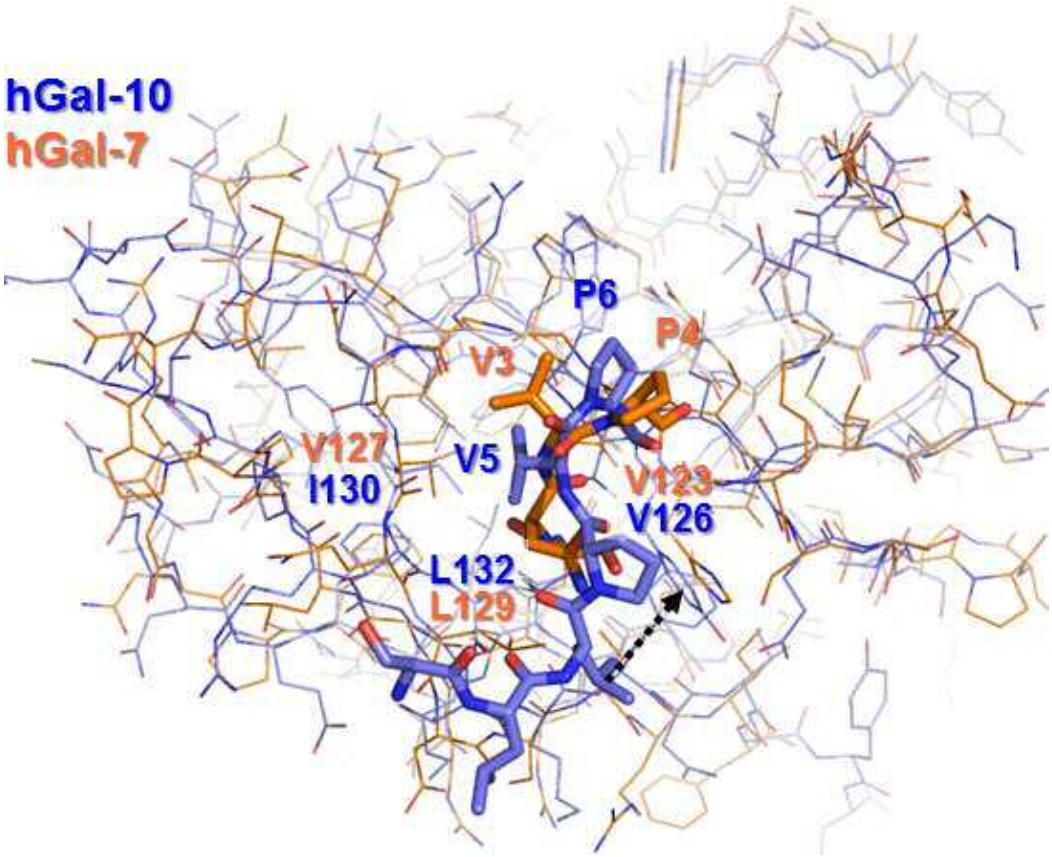


Figure 14: Superimposition of the gal-10 P6 *cis* crystal structure (PDB: 1LCL, in blue) and our model for the gal-7 P4 *cis* conformer (in orange), with arrows indicating those conserved residues participating in similar interactions in both structures.

5.2.3 FBDD targeting of gal-7

The involvement of gal-7 in such heavily-regulated processes as cancer, apoptosis and development, is presumably achieved through the establishment of a prolific interactive network with effector proteins. Being the exact role of gal-7 in most of these processes still poorly understood, physical interactions have been demonstrated between gal-7 and at least two non-glycosylated partners: the apoptotic regulator Bcl-2 [7] and the transcription factor Smad2/3 [18]. Of note, the gal-7/Bcl-2 interaction was shown to be independent of carbohydrate binding. Furthermore, non-glycosylated protein partners are also known to other galectins, such as Bcl-2 itself for gal-3 [19] and farnesyl-Ras for gal-1 [20]. In the latter case, the binding epitope on gal-1 was delineated by the construction of a mutant (gal-1 L11A) which retained normal carbohydrate-binding abilities, but inhibited Ras-GTP, giving rise to an attenuated pro-oncogenic activity of H-Ras(G12V) [21]. Taken together, all these data evidence the interacting capabilities of galectins beyond the carbohydrate recognition site.

Fragment-based drug discovery (FBDD) is a powerful approach for the identification of novel inhibitor scaffolds as well as novel binding pockets on protein targets [22]. It involves the screening of small libraries of low molecular weight compounds (typically <300 Da), whose small size and low chemical complexity results in low-affinity binders, thus requiring sensitive methods such as NMR for detection. Subsequently, fragment hits are developed or merged to yield inhibitors of higher potency.

Targeting gal-7 through FBDD, we found a set of compounds that weakly bind to gal-7, as screened by using saturation transfer difference (STD) experiments performed on fragment mixtures from a commercial library of 1,000 small organic compounds. Characterization of their binding epitopes on the gal-7 surface in ^{1}H-^{15}N HSQC experiments revealed that all fragment hits targeted either of two protein epitopes, none of which is the sugar binding site. These novel binding sites are the gal-7 dimerization interface, and a previously unreported site located at a shallow groove in a lateral face of the lectin (fig. 15).

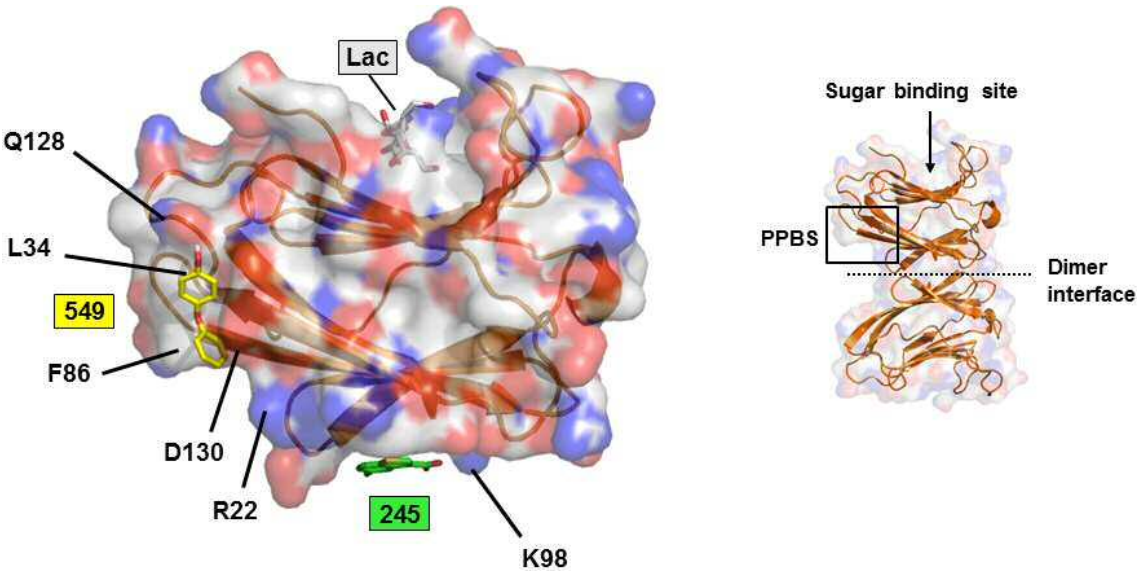


Figure 15: Left: Monomer of the X-ray structure of lactose-bound gal-7 (PDB: 4GAL) with fragments 245 and 549 bound in their respective sites (the dimer interface and the PPBS, respectively) as calculated with AutoDock. Several gal-7 residues important for fragment binding are labeled. Right: scheme of the different binding sites on the gal-7 dimer structure.

The latter binding site was targeted by hits bearing a phenoxyphenyl scaffold (fig. 16). This region, which we subsequently called the phenoxyphenyl-binding site (PPBS), is located at a shallow cleft near a larger groove close to the dimer interface. NMR titrations (fig. S5 in Appendix) revealed fragment 549 as the most potent of the series, with a K_D of 1.27 ± 0.09 mM. Docking calculations positioned 549's hydroxyl group at hydrogen bond distance of the Q128 side-chain, and a phenyl ring well-positioned to establish a CH-π interaction with the L24 side chain (fig. 15). Although the role of the PPBS on the gal-7 surface, if any, is unknown, the size of its adjacent groove, added to the performance of phenoxyphenyls as β-turn mimics [23-24] suggest that the PPBS could be involved in gal-7/protein interactions. Of note, the PPBS in gal-7 is structurally and sequence-homologous to an epitope of galectin-1 encompassing F30, L32 and I128. This region of gal-1 overlaps to the putative binding site of farnesyl-Ras [20-21], whose association to gal-1 is necessary for H-Ras and K-Ras activation [21].

A second set of fragments, exemplified by 245 (fig. 17), gave rise to STD signals, indicating binding to gal-7. However, and interestingly, no changes in the ^{1}H-^{15}N HSQC spectrum of gal-7 were detected upon addition of 245 (fig. S6 in Appendix). This was suggestive that its binding

site was hidden under those experimental conditions. Particularly, and since the only difference between the experimental setup for STD and HSQC experiments was the employed protein concentration (15 µM and 250 µM, respectively), we guessed that 245 bound to the dimer interface, an event that would be precluded by the dimerization process that takes place for higher gal-7 concentrations.

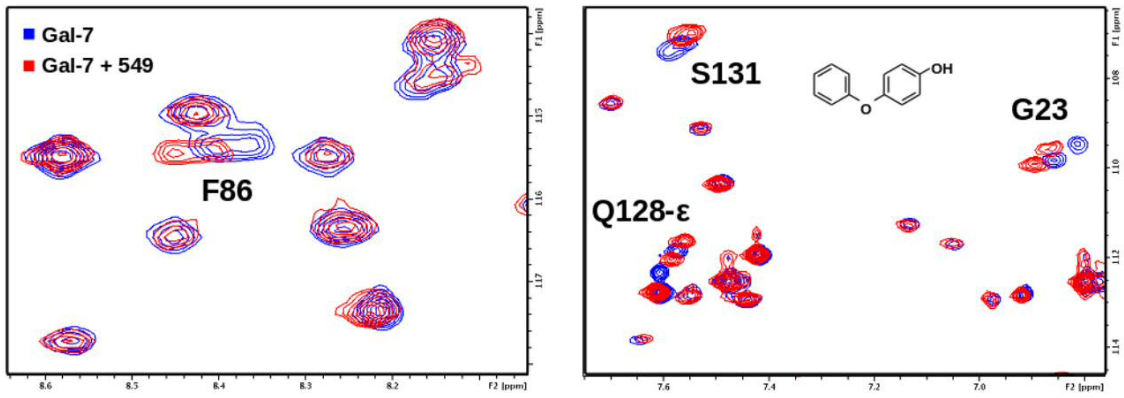


Figure 16: Details of the ^{1}H-^{15}N HSQC spectra of gal-7 in the absence (blue) and in the presence of 549 (red), showing selective perturbation of a number of peaks corresponding to residues located at the PPBS.

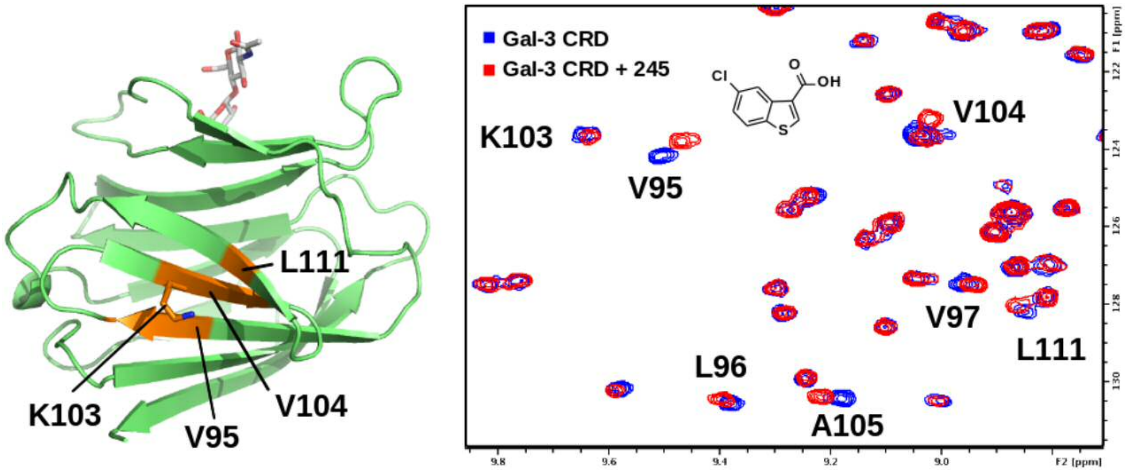


Figure 17: Details of the ^{1}H-^{15}N HSQC spectra of gal-3 CRD monomer in the absence (blue) and in the presence of 245 (red), showing perturbation of a number of peaks located at a region which is structurally homologous to the gal-7 dimer interface. On the left, the X-ray structure of LacNAc-loaded gal-3 CRD (PDB: 1A3K) is shown, highlighting the residues corresponding to the most perturbed peaks by 245 addition.

To corroborate this hypothesis, we used the CRD of gal-3 (gal-3 CRD), a monomer homolog of gal-7, obtained by truncation of the chimera-type full-length gal-3 [25]. ^{1}H-^{15}N experiments performed with ^{15}N-labeled gal-3 CRD revealed selective chemical shift perturbations of a number of peaks when 245 was added. The most shifted signals corresponded to amino acids located at a hydrophobic patch of the protein surrounding the K103 side chain, at the opposite face of the sugar binding site (fig. 17). This region is structurally homologous to that at the gal-7 dimer interface. Accordingly, docking calculations using the gal-7 dimer interface as input positioned 245's benzotiophene rings at CH-π distance of the A99 side-chain methyl, and its negatively charged carboxylate in contact with the K98 side-chain (figs. 15, 19B). This pose is in agreement with the ligand epitope mapping of 245 obtained by diffusion-ordered spectroscopy (DOSY) [26] (fig. 18), which indicates that protons adjacent to the carboxylic group are in tighter contact with the protein.

Importantly, binding of 245 at the dimer interface of gal-7 showed disruptive effects for dimerization. By performing DOSY experiments, we obtained direct evidence that fragment 245 disrupts gal-7 self-association, as shown by a significant change in the diffusion coefficient (*D*) of gal-7 in the presence of 245. By adding the fragment to a sample containing 15 µM gal-7, a concentration where it is predominantly dimeric, given its dimerization K_D of around 1.7 µM [27], the *D* value of the gal-7 signals shifted from -10.07 to -9.95 (fig. 18). Extrapolation of these values in a calibration line [28] (fig. S1 in Appendix) gave apparent molecular weights of 27 kDa and 13 kDa for gal-7 in the absence and in the presence of 245, respectively, which is in agreement with dimer dissociation of gal-7 upon addition of 245.

In addition, STD experiments showed that binding of 245 promoted the access of other small molecules to the interface (fig. 19A), where there is enough room for at least one additional fragment (fig. 19B). Similarly, binding of 245 to gal-7 enhanced further binding of 245 itself, as shown by STD response curves as a function of fragment concentration (fig. 20). These results not only confirm the ability of 245 to disrupt gal-7 dimerization, but also give good perspectives about the feasibility of obtaining stronger binders to the gal-7 dimer interface following a fragment merging approach. Looking at this possibility into detail, selective inter-ligand NOEs (iLOEs) [29] were detected between 205 and 245 (fig. 19B), thus providing information on the relative orientation of these fragments, when bound at the gal-7 dimer interface.

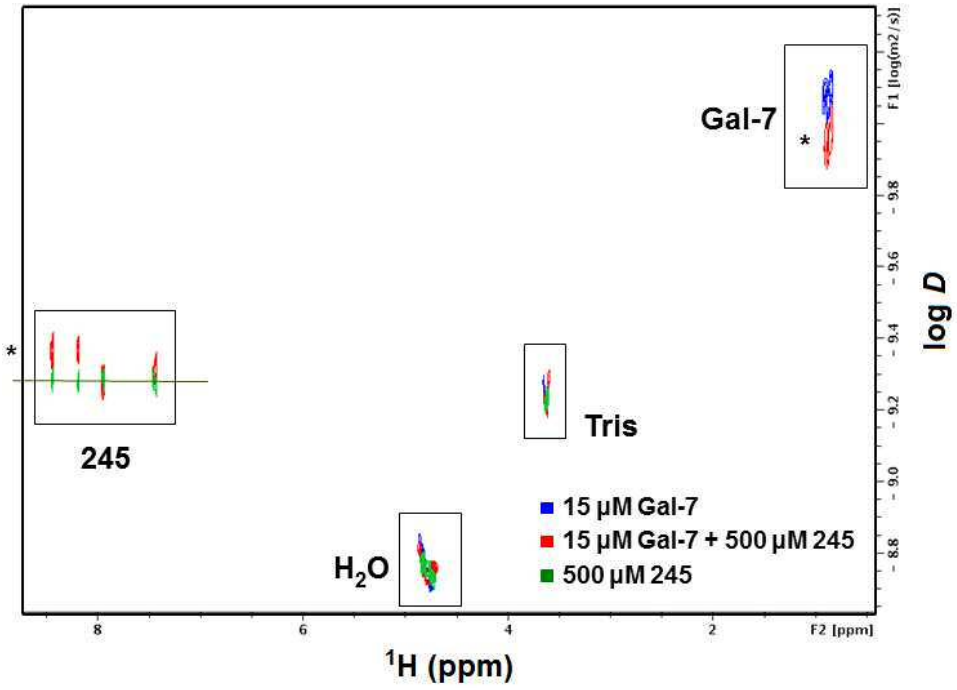


Figure 18: DOSY spectra of gal-7/245 samples, showing disruption of gal-7 dimerization upon addition of 245. Blue: gal-7; green: 245; red: gal-7 + 245. An epitope mapping effect [26] is observed at the 245 signals.

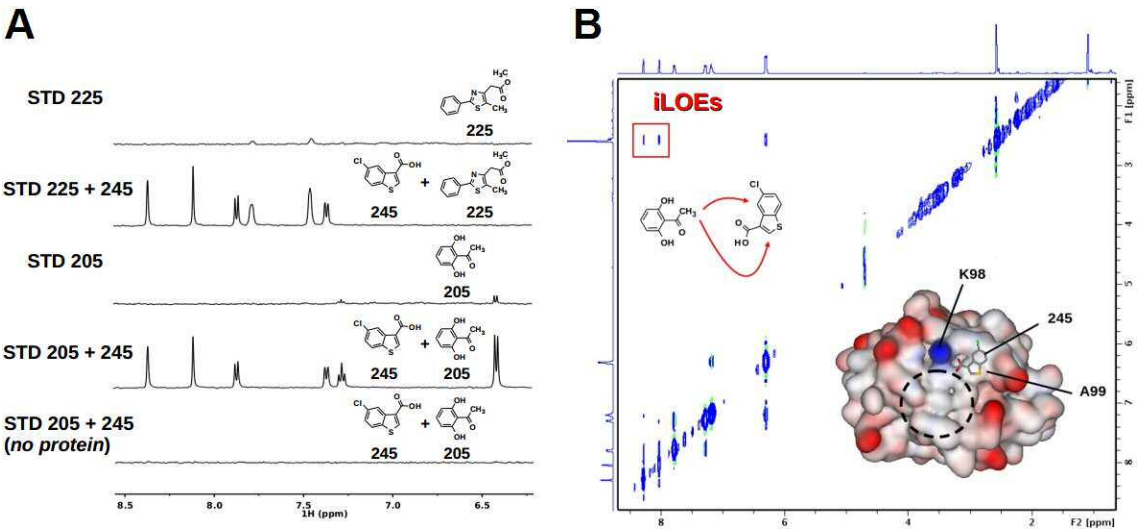


Figure 19: Simultaneous binding of multiple fragments to the gal-7 dimer interface. A: STD spectra of 500 µM (1) 205, (2) 205 + 245, (3) 235, (4) 235 + 245, in the presence of 15 µM gal-7; and (5) 235 + 245 in the absence of gal-7. B: NOESY experiment showing selective cross-peaks between the 205 methyl signals and two protons of 245 (iLOEs). Overlaid, docking structure of the gal-7 monomer with 245, highlighting a hydrophobic space belonging to the gal-7 dimer interface that could be accessed by additional fragments when gal-7 is found as a monomer.

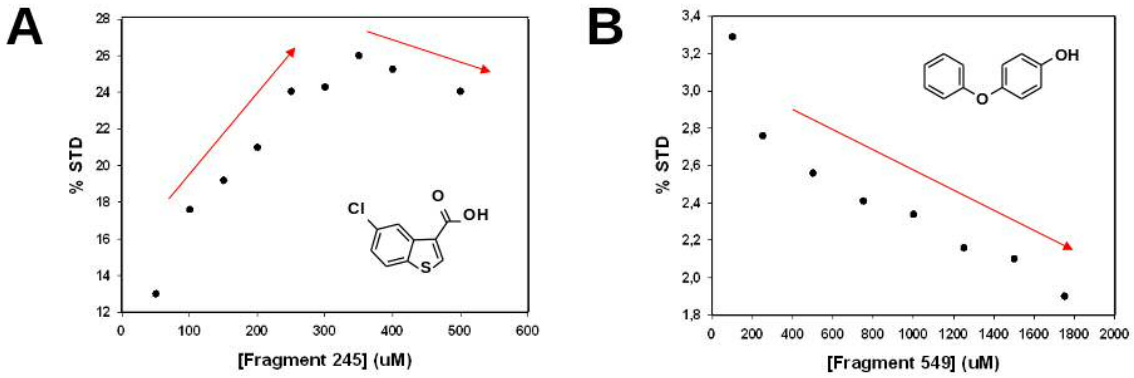


Figure 20: STD response curves as a function of fragment concentration. While it is expected that increasing ligand molar excesses result in diminished STD response ratios (calculated as $I_{STD}/I_{Off\text{-}resonance}$), the %STD grows at growing concentrations of 245 (A) up to 300 µM, when it starts to drop. This also gives support to the notion that binding of 245 inhibits gal-7 dimerization, subsequently triggering the binding of small molecules to secondary sites at the hydrophobic gal-7 dimer interface. In contrast, the STD rate of 549 (B), a fragment that binds at the phenoxyphenyl binding site on gal-7 dimers, diminishes as a function of the fragment's concentration.

To the best of our knowledge, these results represent the first example of disruption of the oligomeric state of a protein achieved through a fragment-based approach. This success could be due to the intrinsic nature of lectins as proteins recognizing highly polar molecules, and thus devoid of large hydrophobic clefts other than interfaces for oligomerization and other protein/protein interactions, and will inspire future approaches with other members of the lectin family.

It is known that the galectin oligomer state can drastically impact function [30]. For example, while both gal-1 dimers and monomers can bind glycans on various cells, only the Gal-1 dimeric form can induce signaling and promote macrophage-mediated phagocytosis vis-à-vis cell apoptosis [31]. In another study, it was shown that only a dimeric form of Gal-1 could potently induce apoptosis in murine thymocytes and mature T cells, whereas a monomer form of Gal-1 could not [32]. In addition, Miura et al. reported that while both dimer and monomer forms can promote axonal regeneration, only dimer gal-1 could effectively induce Jurkat cell death [33]. Other studies have reported galectin oligomer-mediated effects in cell agglutination studies [34-35]. These facts invite to use 245, or a more potent derivative thereof, to explore the possible biological effects of inhibiting dimerization of gal-7.

At the same time, phenoxyphenyl derivatives emerge as scaffolds for disrupting any potential, yet to be discovered gal-7/protein interactions taking place at the PPBS.

5.2.4 Lactose binding enhances the stability of the gal-7 homodimer

Although the binding site of fragment 245 was identified at the gal-7 dimerization interface, initial attempts to determine it yielded contradictory results.

Competition experiments using STD showed that addition of lactose to samples containing gal-7 and fragment 245 resulted in an apparent displacement of 245 by the disaccharide (fig. 21). While this was suggestive that lactose and 245 competed for the same binding site, the absence of perturbations in the gal-7 ^{1}H-^{15}N HSQC spectrum exerted by 245, and subsequent evidences for binding at the dimer interface ruled out this possibility. Instead, these observations, supported by other data, indicated that lactose allosterically competed with 245 by promoting dimerization of gal-7.

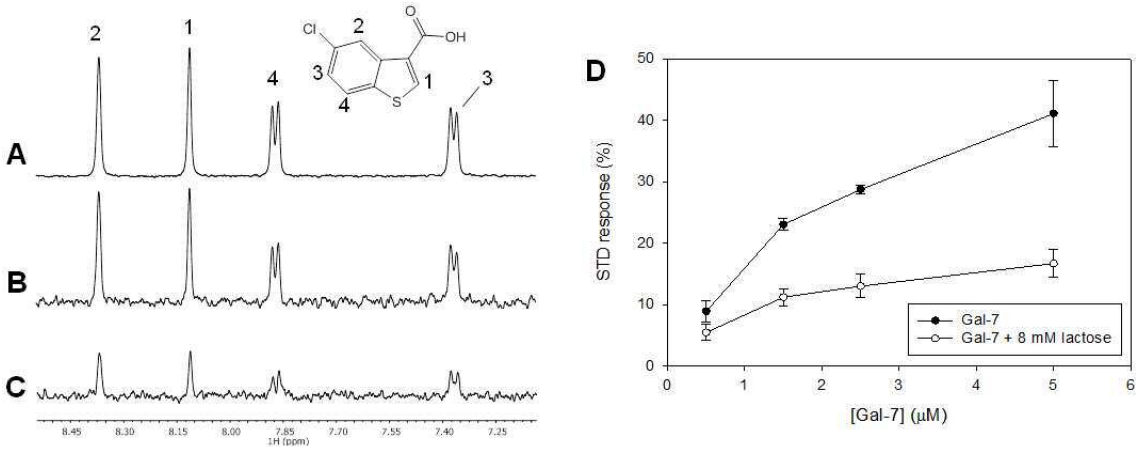


Figure 21: ^{1}H-NMR spectra of samples containing 500 µM 245 in the presence of 1.5 uM Gal-7. A: Reference spectrum (scaled 25%) with labels indicating the assignment of proton resonances. B: STD spectrum. C: STD spectrum in the presence of 8 mM lactose. D: Plots showing STD responses for 500 µM 245 vs. the concentration of gal-7, before (filled circles) and after (open circles) addition of lactose to reach the concentration of 8 mM.

Figure 21 shows a series of STD experiments performed using fragment 245 as a spy molecule, at four different gal-7 concentrations ranging from 500 nM to 5 µM, i.e. around its dimerization K_D, of 1.7, µM. In all instances, the addition of lactose to the analyzed samples significantly reduced the STD signal intensities of 245. Since the actual amount of monomer in fact increases as the total concentration of Gal-7 increases, the STD effect also increases with the Gal-7 concentration. Because STD signals are significantly reduced in the presence of

lactose at any concentration of gal-7, these data were interpreted in terms of an indirect competition over 245 exerted by carbohydrate binding, as shown in the scheme of fig. 22. In this model, the binding of sugar stabilizes the gal-7 dimer state, promoting dimerization of the lectin, and thus the dimer interface becomes no longer exposed, effectively preventing the spy molecule to access its binding site.

These observations, in concert with data from FPLC, circular dichroism, and molecular dynamics studies, revealed that carbohydrate binding to gal-7 induces long-range effects throughout the protein, consisting in minor conformational shifts and changes in structural dynamics, which result in stabilization of the dimer state [27].

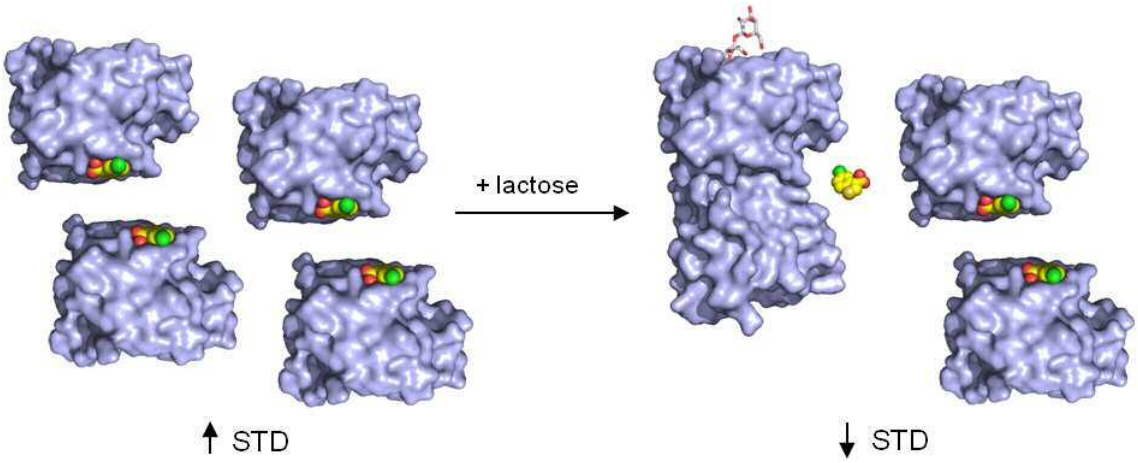


Figure 22: Model of allosteric competition exerted by lactose on 5-chloro-benzo[b]thiophene-3-carboxylic acid (fragment 245). At low gal-7 concentrations, the protein is predominantly found as a monomer, and the dimer interface is exposed. Under such condition, the probe molecule is able to access its binding site and becomes saturated in the STD experiments. When lactose binds to the lectin, dimerization is promoted and the dimer interface is no longer accessible. As a result, access of the probe molecule to its binding site is impaired, and its STD signal decreases.

5.3 Galectin-3: a chimera-type galectin

Gal-3 is structurally unique among galectins. Aside from its pan-galectin-conserved CRD, gal-3, the only chimera-type galectin, has an extended N-terminal domain (ND) of 115 residues composed of a 20 amino acid-long N-terminal peptide, followed by tandem repeats of short segments (mostly collagen-like, rich in Pro and Gly) connected to its C-terminal CRD (~135 residues) [36]. Sequence conservation of ND repeat units is strong among mammals, extending to homologues in fish, frog and chicken, that developed phylogenetically with a conserved structural design [37].

Although the gal-3 CRD β-sandwich structure binds glycans like other galectins, its unique N-terminal domain plays a significant role in gal-3 function. For example, it modulates sugar binding to its CRD, mediates cell signaling, and interacts with lipopolysaccharide (LPS) lipid A moiety to enhance LPS-mediated neutrophil activation. In addition, the N-terminal domain has three post-translational phosphorylation sites (Ser6, S12, Y107) with functional consequences.

Functionally, gal-3 is anti-apoptotic in various cells and against a diverse array of apoptotic stimuli [38-39], being involved in immune and inflammatory responses, and in tumor development and progression. [40-42]. Gal-3 is highly expressed in macrophages and immature dendritic cells, and its expression is up-regulated in activated macrophages and is down-regulated during dendritic cell maturation [43-44]. Gal-3 is also expressed in many types of epithelial and stromal cells, including fibroblasts [45-46]. Gal-3 is also heavily involved in innate immunity, by binding glycans at the surface of different types of pathogens. In addition to its aforementioned reactivity towards LPS, gal-3 can bind to Gal-β(1-3)-Gal, a *Leishmania* virulence factor [47-48], to Gal-α xenoantigen [49-50], and to LacdiNAc-glycans, which constitute a parasite pattern for gal-3-mediated immune recognition [51].

Although extracellular functions of gal-3, like galectins in general, mediate most of their known activities by binding to saccharide ligands on cells and extracellular matrix [30, 52-55], gal-3 can also function intracellularly [56] by e.g. inducing pre-mRNA splicing and regulating expression of certain genes like cancer-related genes cyclin D1, thyroid transcription factor 1 and mucin 2 [57], as well as regulating JNK1 [58] and directly binding to intracellular proteins like Bcl-2 [59].

Although the structure of the gal-3 CRD has been elucidated [60-61], that of full length gal-3 with its CRD and ND remains unknown. However, it has been suggested that residues 94-113 of the relatively flexible ND interact with the CRD [62-63]. A number of reports have

demonstrated that the ND is crucial to gal-3 function. For example, the presence of the ND enhances CRD binding to carbohydrates [64-65], and while the gal-3 CRD alone can bind cell surface glycans (albeit relatively weakly), it is otherwise functionally inactive without the ND [66].

The work presented in this unit addresses both the interaction properties of the gal-3 CRD with functionally relevant glycans, as well as structural aspects of its ND in the context of the full-length protein.

5.3.1 Full NMR assignment of galectin-3

Sequence specific NMR assignments for the gal-3 CRD had been previously reported [67]. However, emerging evidence on the broad physiological significance of the full-length protein, well beyond glycan binding, emphasized the need for NMR assignment of full-length gal-3.

To allow for detailed studies on gal-3 at atomic resolution, the ^{1}H, ^{13}C and ^{15}N resonances of the full-length protein were assigned, also as a joint effort with the laboratory of Prof. Mayo (University of Minnesota).

Most resonances in native gal-3 exhibit nearly the same chemical shifts as those in gal-3 CRD, thus facilitating the assignment of this domain in the full-length protein. The real challenge was assigning resonances from the ND, which has 115 residues, with about 20 from a short N-terminal segment and about 95 in the collagen α-like domain. This domain is PG-rich (26 Pro and 26 Gly residues), and has a number of repeat sequences. Most of the resonances of the ND fall within a narrow ^{1}H spectral region coincident with that for random coil sequences.

To aid in the assignment process, several truncation variants of the protein were made by permutationally removing tandem repeats. In addition, a series of peptides derived from the ND were chemically synthesized and their NMR resonances were assigned. Our contribution consisted in the full assignment of ^{1}H, ^{13}C and ^{15}N resonances of the 20-residues long N-terminal peptide (fig 23A), both unphosphorylated, and with biologically relevant phosphorylations on Ser6 and/or Ser12. To this end, homonuclear TOCSY and NOESY experiments were performed, in combination with heteronuclear ^{1}H-^{13} HSQC and ^{1}H-^{15}N HSQC experiments at natural isotopic abundance.

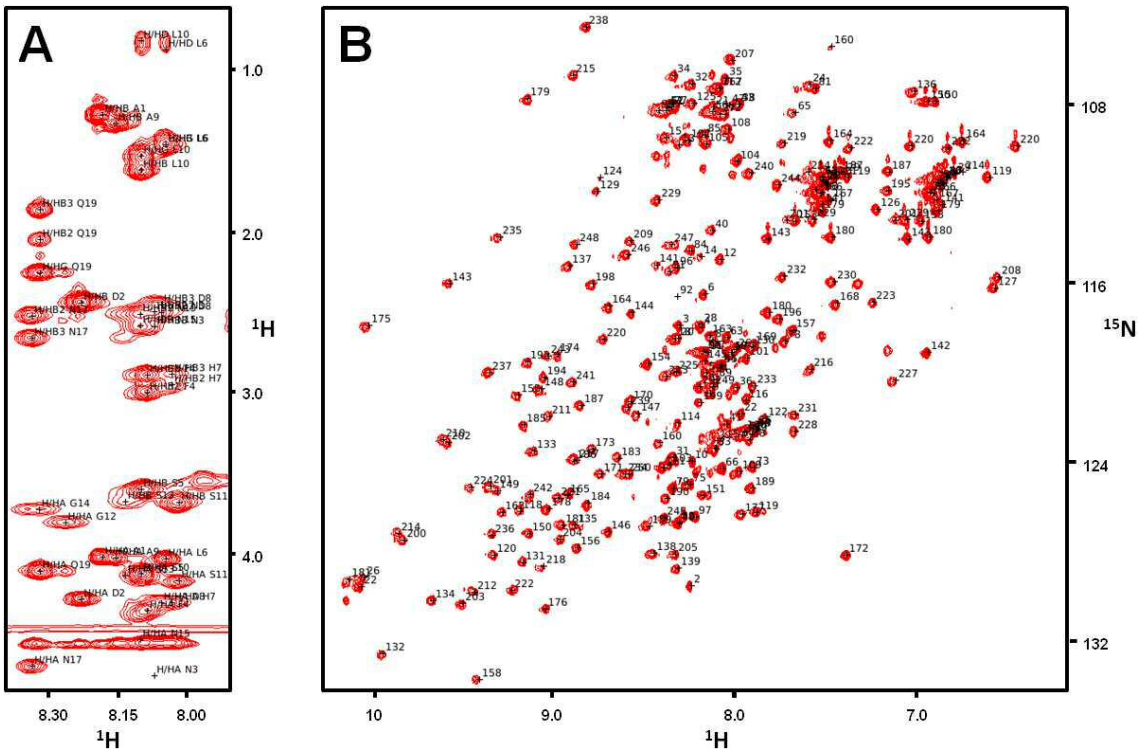


Figure 23: Assignment of full-length gal-3. A: Detail of the ^{1}H-^{1}H TOCSY of the 20 amino acids-long gal-3 N-terminal peptide. B: ^{1}H-^{15}N HSQC spectrum of full-length gal-3. Signals from the ND are poorly dispersed within a narrow ^{1}H spectral region coincident with that for random coil sequences (ca. 7.8-8.3 ppm).

In total, backbone resonances (^{1}HN, ^{15}N, ^{13}CO, $^{13}C_\alpha$ and $^1H_\alpha$) for all residues within the CRD and N-terminal tail were fully assigned (fig. 23B), with the exception of the fast exchangeable HN amides H64, G93, G152, N153, N166, S188, and K226, as well as CO of residues preceding prolines in the CRD. In contrast, resonances within the tandem-repeat section of the ND were difficult to uniquely assign, because most resonances have near-identical proton and backbone carbon chemical shifts. For example, the segment ^{67}PGAY^{70} could not be uniquely assigned. Similarly, ^{42}PGAYPGQAPP^{51} has the same amino acid sequence as segment ^{51}PGAYPGQAPP^{60}, leading to indistinguishable chemical shifts for resonances of each equivalent amino acid residue.

Overall, greater than 90% of side-chain resonances were assigned in both the CRD and N-terminal tail. Chemical shift assignments were deposited in the BioMagRes-Bank [8] under accession number 19491 and published in the *Biomol NMR Assign* journal [68].

This assignment report also represents the first step toward the 3D structural elucidation of native, full-length gal-3 in solution.

5.3.1 The interaction of oligosaccharides at the gal-3 CRD

Assignment of the full length protein allowed for the study of ligand-binding to the full, chimeric protein. Figure 24 shows a titration of lactose, the pan-galectin ligand, against full length gal-3. The largest chemical shift differences are concentrated in the CRD part of the protein. In contrast, signals from the ND remained essentially unaffected upon binding of the disaccharide. Accordingly, in many of our studies with small oligosaccharides, we made use of the truncated protein, i.e. the gal-3 CRD. As a first approach, we investigated the binding of this protein to several oligosaccharide ligands of relevance for innate immunity.

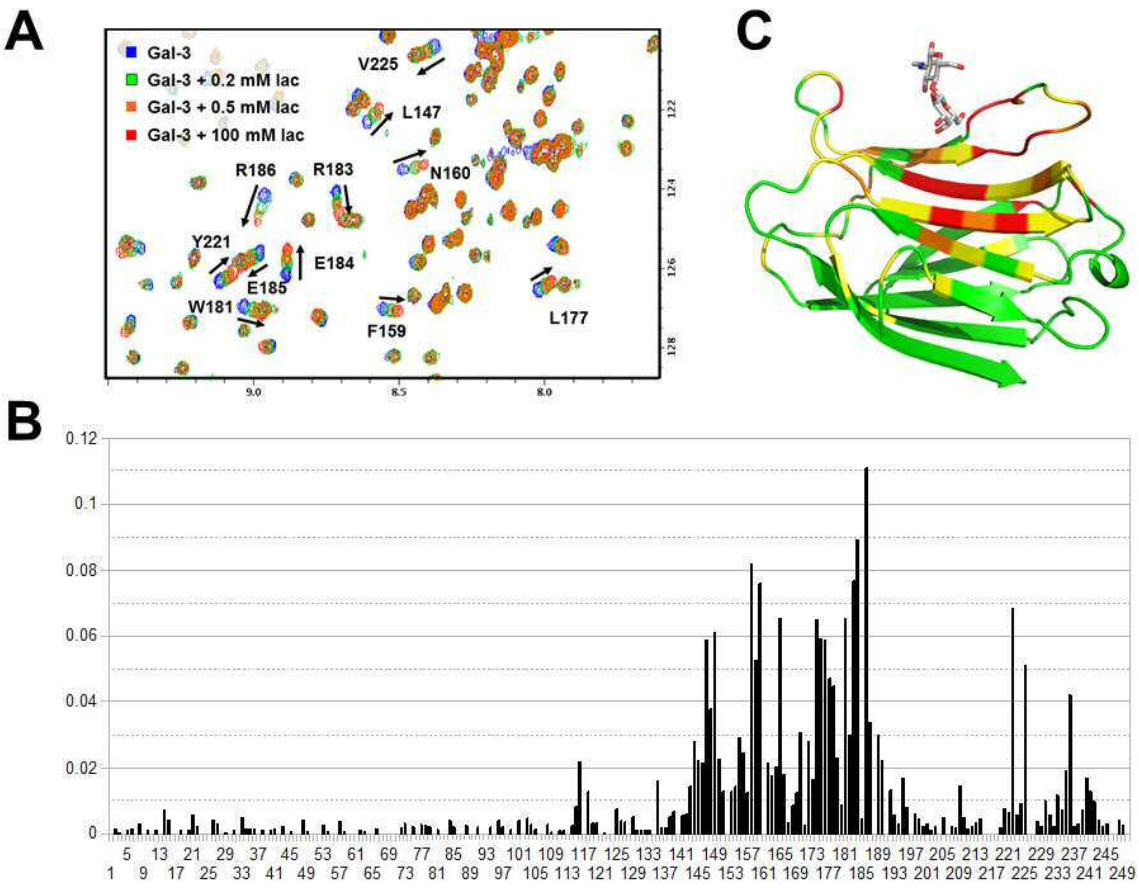


Figure 24: Lactose-binding titration on full-length gal-3. A: Overlaid ^{1}H-^{15}N HSQC spectra of full length gal-3 acquired at 800 MHz in the presence of lactose at different concentrations up to 100 mM, with labels indicating some of the most affected residues upon lactose binding. B: ^{1}H-^{15}N-weighted chemical shift differences for full-length gal-3 backbone NH signals upon addition of a saturating concentration of lactose. The largest chemical shift differences are concentrated in the CRD (residues 107-250), with changes in the resonances of the ND (residues 1-106) virtually non-existent. C: Structure of LacNAc-loaded gal-3 CRD (PDB: 1A3k), with residues highlighted in red for the most shifted resonances upon lactose binding, followed by orange and yellow.

5.3.1.1 The recognition of lacdiNAc by gal-3

LacdiNAc (GalNAc-β(1-4)GlcNAc) is structurally very similar to lactose, only differing in that it incorporates an *N*-acetyl group at 2-OH of both sugar rings of the disaccharide. LacdiNAc is present in vertebrates as constituents of *N*-glycans. Besides, it is highly abundant in helminths, and it has been proposed as a parasite pattern for gal-3- mediated immune response [51].

The interaction between lacdiNAc and gal-3 was studied by NMR. The addition of 10 equivalents of lacdiNAc to ^{15}N-labeled gal-3 CRD produced a very similar profile of the backbone chemical shift variations to those produced by lactose (fig. 25). In addition, a region of gal-3 comprising residues Arg162-Phe163-Asn164 became differentially perturbed by lacdiNAc. Docking calculations showed a pose of lacdiNAc very similar to that of lactose. The differential perturbation of residues 162-164 can be explained by their proximity to the additional *N*-acetyl group of lacdiNAc at the galactose residue (fig. 25D).

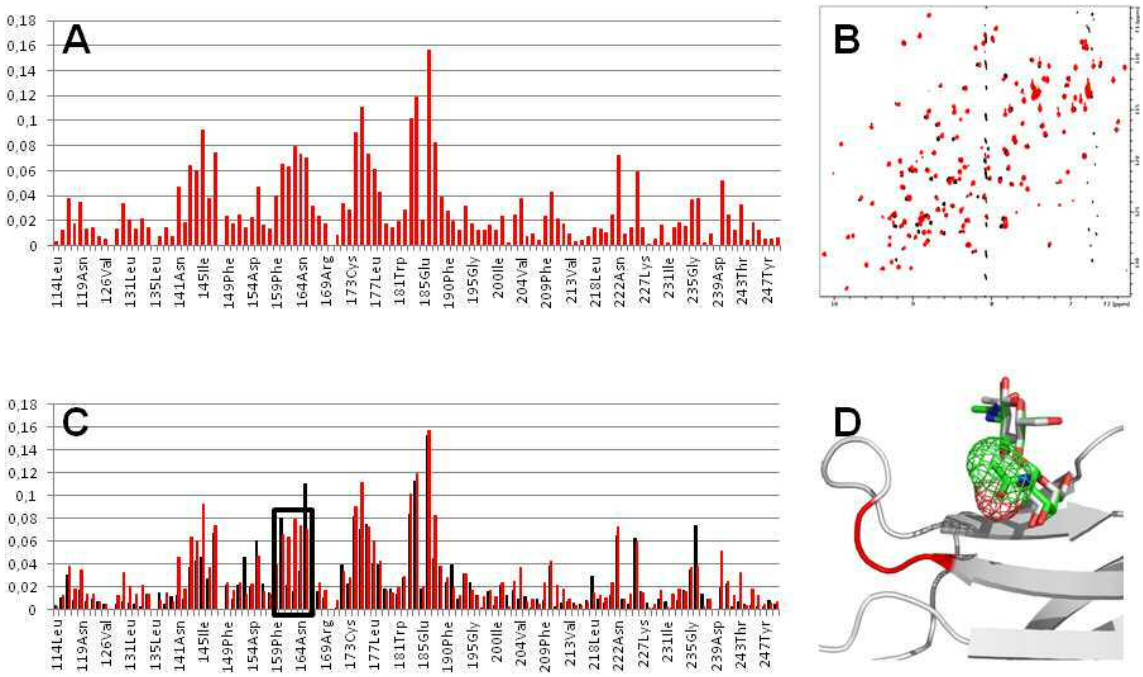


Figure 25: The interaction of lacdiNAc with gal-3. A: Chemical shift differences for backbone amides of gal-3 CRD upon addition of 10 equivalents lacdiNAc. B. Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 CRD (black) and in the presence of 20 equivalents lacdiNAc (red). C: Comparison between chemical shift differences for lacdiNAc (red) and lactose (black), highlighting a group of residues differentially perturbed by lacdiNAc (Arg162-Phe163-Asn164). D: Docked structure of lacdiNAc (green) to gal-3, superimposed to the X-ray structure of LacNAc-bound gal-3 (white). The additional N-acetyl of lacdiNAc is shown as mesh. Residues differently perturbed by lacdiNAc are highlighted in red.

5.3.1.2 The recognition of the α-Gal epitope by gal-3

The α-Gal epitope (Galα-1,3-Galβ-1,4-GlcNAc-R) is a major xenoantigen expressed on glycoconjugates of non-primate mammals. In turn, apes and humans produce a natural antibody (anti-Gal), which specifically binds the α-Gal epitope. Anti-Gal is produced as the most abundant antibody (1% of immunoglobulins) in all individuals, and represents an important barrier for xenotransplantation [69]. However, this antibody is not the only responsible for the rejection of pig xenograft organs in humans, since lectins have also been shown to be involved in that process. Specifically, gal-3 has been identified as the receptor for monocyte-dependent recognition of porcine endothelium via binding to the α-Gal epitope [49], playing an important role in delayed xenograft rejection [70].

Using NMR, we studied the interaction between gal-3 and two oligosaccharides containing the α-Gal epitope: the Galα-1,3-Galβ-1,4-GlcNAc trisaccharide, i.e. the minimal structure containing the α-Gal epitope, and a tetrasaccharide derivative thereof, namely Galα-1,3-Galβ-1,4-GlcNAcα-1,3-Gal. STD spectra of these oligosaccharides in the presence of gal-3 (figs. 26-27) show a differential saturation response for several ligand nuclei. Specifically, in the case of the trisaccharide, those of the galactose residues showed higher STD responses that of the glucosamine residue (fig.26). This result is similar to analogous experiments previously reported on lactose [71], in the sense that the glucosyl moiety receives less saturation than the galactose unit. This, in turn, is in agreement with the binding mode of this disaccharide to the CRD of gal-3, which exposes the glucose residue toward the solvent, as demonstrated both by X-ray crystallography [60] and NMR studies [72]. For the tetrasaccharide, the STD pattern of the digalactose portion was very similar to that obtained with the trisaccharide. However, slightly higher degrees of saturation were found at the GlcNAc moiety. Besides, the additional β(1-3)-linked galactosyl moiety also received some degree of saturation (fig. 27). This seems rather surprising, since the expected binding mode for the tetrasaccharide would leave this galactose residue as the one most exposed to the solvent. Thus, it is possible that transient interactions occur between this β-galactose residue and the canonical galactose-binding site of gal-3, explaining this phenomenon.

The addition of 10 equivalents of either molecule to ^{15}N-labeled gal-3 produced a similar profile of the backbone chemical shift variations to those exerted by lactose (fig. 28). This indicates that the binding mode of these molecules closely resembles that of lactose. Besides, the differential shifting of residues 148-149 in the presence of both sugars (and not with lactose) suggest that this region is proximal to the additional, α(1-3)linked galactosyl moiety.

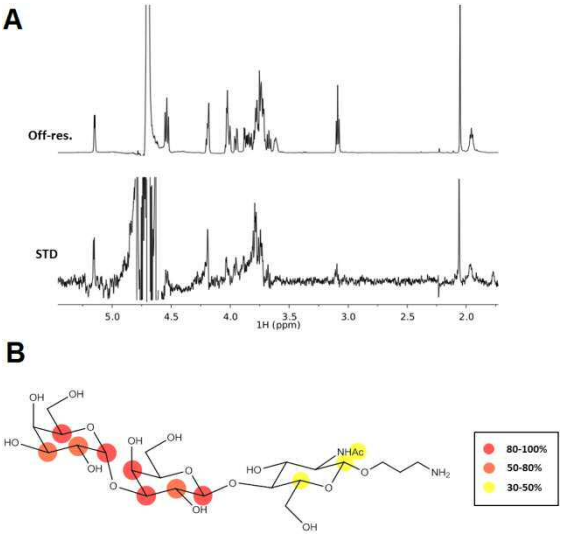


Figure 26: A: STD experiment on Galα-1,3-Galβ-1,4-GlcNAc in the presence of gal-3. Top spectrum: Off-resonance. Bottom: STD (up-scaled). B: Epitope mapping deduced from STD data, with colors showing relative STD intensities to the most saturated signal.

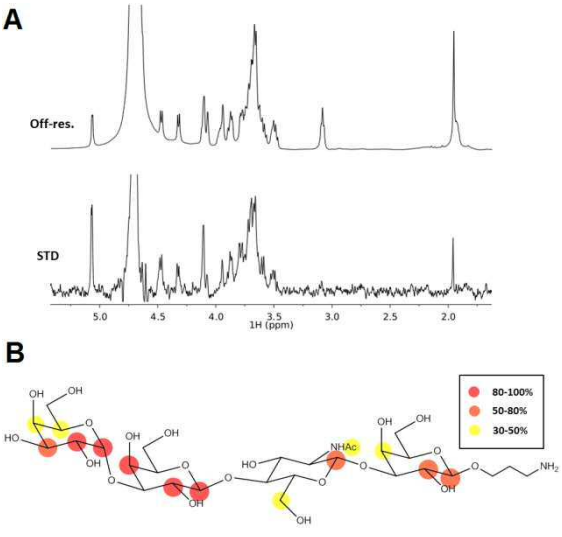


Figure 27: A: STD experiment on Galα-1,3-Galβ-1,4-GlcNAcα-1,3-Gal in the presence of gal-3. Top: Off-resonance. Bottom: STD (up-scaled). B: Epitope mapping deduced from STD data, with colors showing relative STD intensities to the most saturated signal.

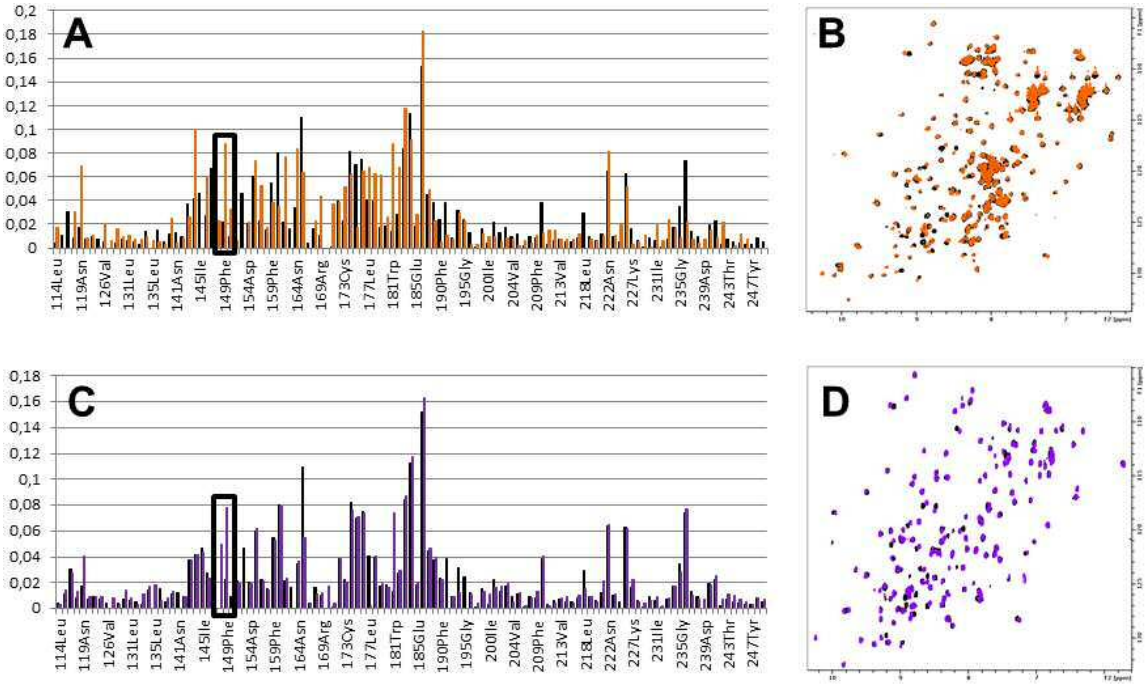


Figure 28: Interaction of α-Gal tri- and tetrasaccharide with gal-3. A: Comparison between chemical shift differences for the trisaccharide (orange) and lactose (black). B. Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 (black) and with 10 eq. trisaccharide (orange). C: Comparison between chemical shift differences for the tetrasaccharide (purple) and lactose (black). B. Overlaid ^{1}H-^{15}N HSQC spectra of gal-3 CRD (black) and with 10 eq. tetrasaccharide (purple).

In order to obtain a 3D view on the interaction, docking calculations were performed on the gal-3 CRD and the α-Gal trisaccharide. As experimental constraints, taking into account the STD results and the protein chemical shift perturbation analyses, we assumed that the lactose moiety of the tetrasaccharide was located at the lactose binding site. A structure from the most populated cluster of this docking analysis is shown in fig. 29. The predicted binding mode indicates that the interactions with the lactosyl moiety are kept, i.e. hydrogen bonds involving side chains of aminoacids His158, Arg162, Arg186 and Glu184, and a CH-π interaction between the Trp181 indole ring and the α-face of GalNAc. Besides, the docked structure predicts the positioning of the additional, α(1-3)-linked galactose moiety onto the vicinal strand, hydrogen-bonding with the Arg144 and Asp148 side-chains. These putative interactions are in good agreement with the selective chemical shift perturbations of these residue peaks, occurring in the presence of the tri- and the tetrasaccharide, but not with lactose.

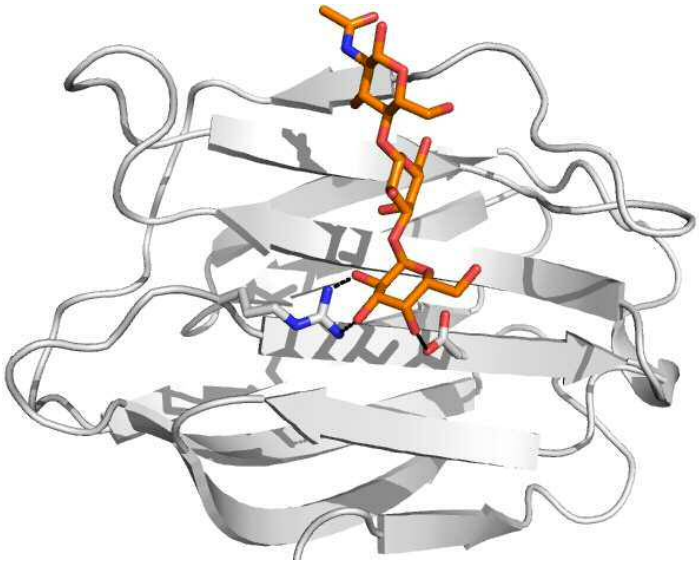


Figure 29: Docked structure from the most populated cluster of an AutoDock 4.2 calculation, using the gal-3 X-ray structure (PDB: 1A3K) treated as rigid and a flexible Galα-1,3-Galβ-1,4-GlcNAc molecule. The predicted binding pose of the LacNAc moiety closely resembles that of the crystallographic structure. Besides, new hydrogen bonding interactions putatively arise between the additional, α(1-3)-linked galactose moiety at the non-reducing terminus and the Arg144 and Asp148 side-chains.

As a general conclusion, the recognition of both lacdiNAc and α-Gal epitope-derived oligosaccharides, all of which can be essentially understood as substituted lactose molecules, show the same recognition features of the disaccharide. In addition, selective local perturbations detected by NMR, and interpreted by molecular modeling, permit to map the location of the additional substitution moieties.

5.3.1.4 The recognition of GAGs at the gal-3 CRD

Although the previous sections have focused on the interaction between gal-3 and small oligosaccharides, galectins perform many of their important biological functions by binding to macromolecular receptors.

Even though the recognition between galectins and higher molecular weight glycans has been extensively documented, attempts of exploring these interactions at atomic resolution have not been yet reported. As shown in this unit, we investigated the binding of two large glycosaminoglycans (GAGs) to full-length gal-3 by using protein-detected NMR experiments.

GAGs are a class of polysaccharides encompassing a variety of sulfated carbohydrates. They include heparin, heparan sulfate, hyaluronic acid and keratan sulfate, which perform important structural functions, and whose recognition by different receptors triggers a plethora of biological responses, such as inflammation, cell adhesion, and regulation of cell growth and proliferation [73-74]. Since some types of GAGs contain poly-LacNAc sequences, they are regarded as potential ligands for galectins.

In a previous work, the interaction of galectins with several types of GAGs was analyzed by frontal affinity chromatography [75]. Results showed that their interaction with keratan sulfate was strong if the galactose residue(s) remained unsulfated, while GlcNAc sulfation had relatively little effect. Gal-3, -7, and -9 also exerted significant affinity to desulfated, GalNAc-containing GAGs, i.e., chondroitin and dermatan, but not at all to hyaluronan and *N*-acetylheparosan. Together, these observations revealed that the integrity of the 6-OH group of β-Gal is important for galectin recognition of GAGs.

GAGs used in this study were keratan sulfate (KS) and desulfated keratan (K), which were produced from bovine cornea in the laboratory of Prof. Robert Linhardt (Rensselaer Polytechnic Institute, Troy, NY, USA). Both polymers bear β1,3-linked LacNAc repeats, which in the case of KS can be sulfated at 6-OH of galactose, GlcNAc residues, both, or none.

As can be seen in the DOSY spectrum of fig. 30, the size of the K and KS samples used in this study were rather large. As estimated through their diffusion coefficients, they behaved as particles of mean sizes 15 kDa and 145 kDa, respectively.

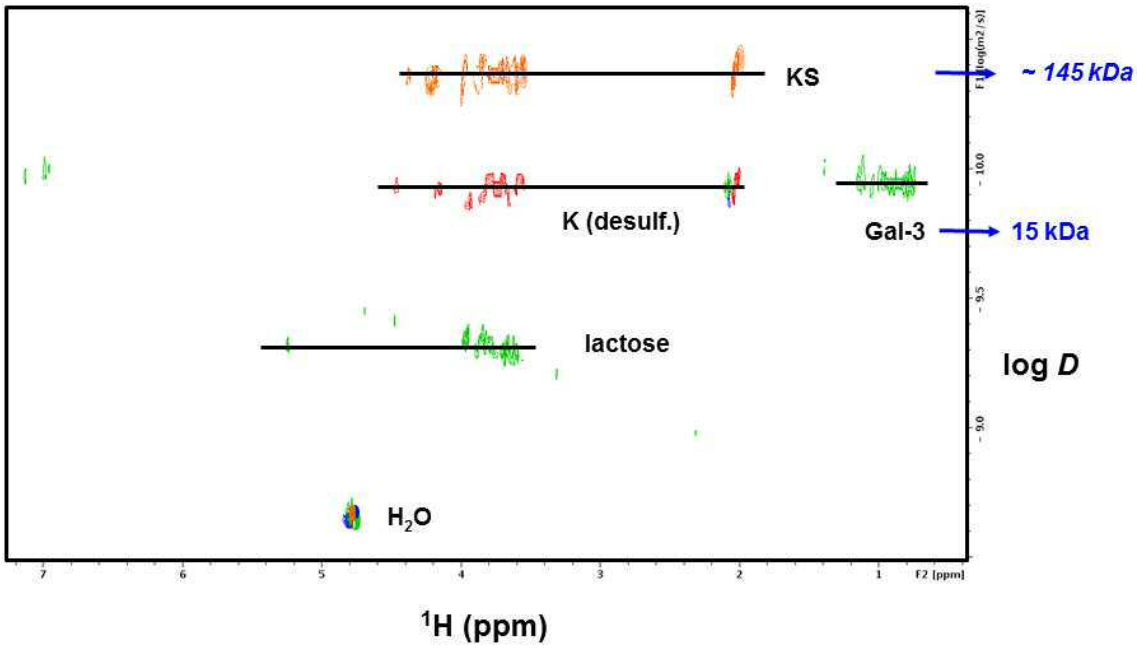


Figure 30: Overlaid DOSY spectra of glycans of different size. Orange: keratan sulfate; red: desulfated keratan; green: a gal-3 CRD/lactose mixture. A molecular weight estimate based on a calibration curve is indicated in blue.

Addition of these glycans to ^{15}N-labeled gal-3 samples caused a dramatic loss of the gal-3 HN signals. This was attributed to a sharp decrease in the correlation time of the protein upon binding to the heavy polysaccharide chains. As seen in fig. 31A, there is a selective perturbation of the CRD signals upon addition of KS, which broadens them beyond detection, leaving the signals from the ND comparatively much less affected. In principle, this permits to deduce that binding occurs at the CRD. When lactose is added to gal-3/KS mixtures, the CRD signals are recovered (fig. 31B), which indicates that the recognition of KS by gal-3 takes place at the canonical sugar binding site of its CRD.

Analogous studies were performed on the less heavy desulfated keratan polysaccharide. These analyses showed that, although the selective broadening of the CRD signals was still apparent, signals were still visible, which permitted to monitor chemical shifts perturbations to map the gal-3 regions involved in the interaction. Again, competition with lactose was shown to revert the CRD signals to a chemical shift pattern characteristic of lactose-bound gal-3 (fig. 32A). As seen in fig. 32C, the gal-3 interface of binding to K includes a large area of the concave face of the β-sandwich, as well as the most C-terminal β-strand in the opposite face. Computational studies with high order oligosaccharides are currently underway to provide a 3D view of the forces involved in these interactions.

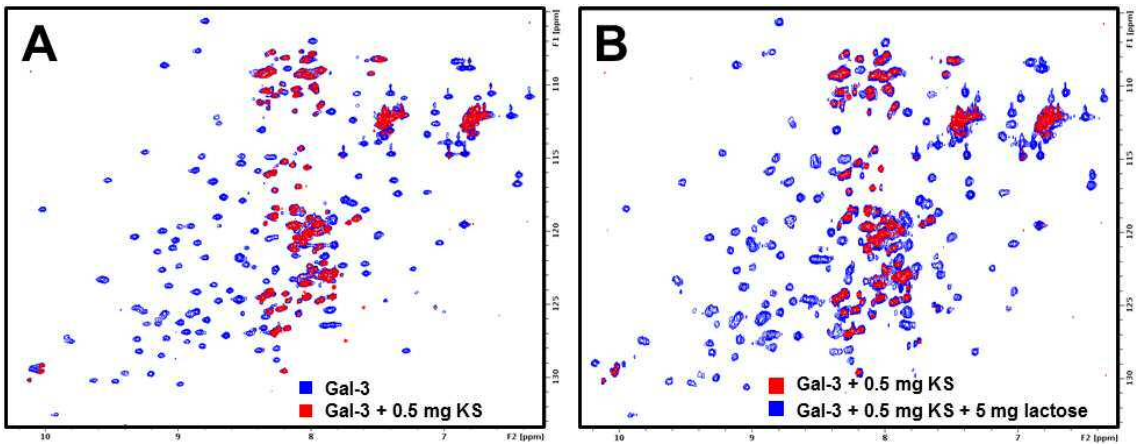


Figure 31: The interaction of the gal-3 CRD with KS. A: ^{1}H-^{15}N HSQC spectra of full length gal-3 before (blue) and after addition of KS (red). The CRD signals are selectively broadened beyond detection upon binding to the polysaccharide. B: ^{1}H-^{15}N HSQC spectra of a gal-3/KS mixture before (red) and after addition of a lactose excess. Addition of the disaccharide recovers the signals of the CRD, thus confirming that binding of gal-3 to KS takes place at the canonical sugar binding site of the gal-3 CRD.

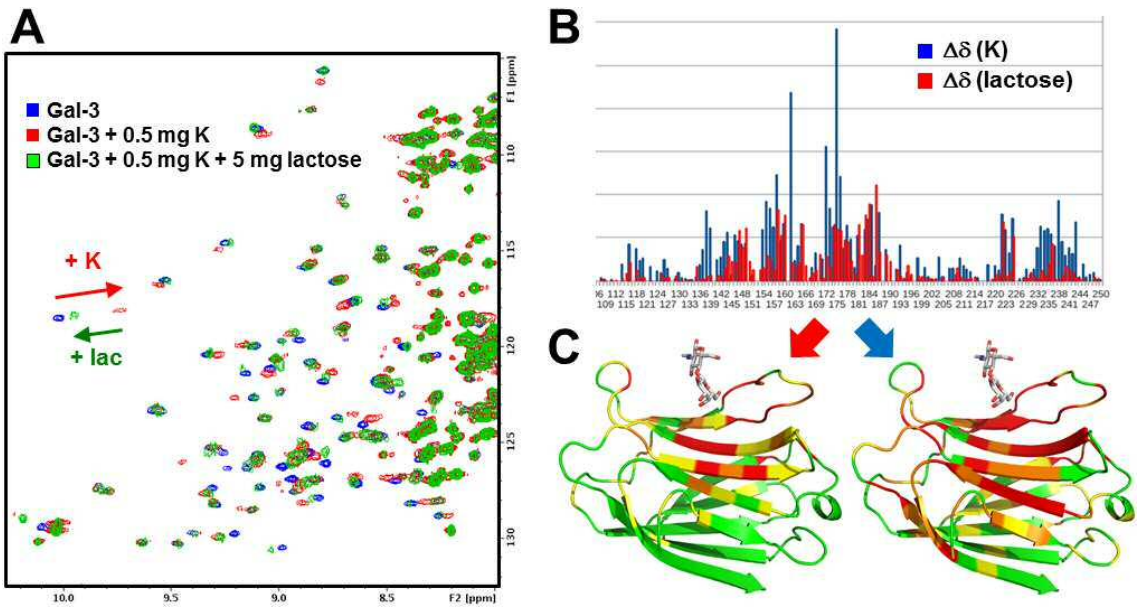


Figure 32: The interaction of gal-3 CRD with desulfated keratan (K). A: ^{1}H-^{15}N HSQC spectra of full length gal-3 alone (blue), after addition of K (red), in the presence of K and a lactose excess (green). B: Comparison between chemical shift differences exerted by K (blue) and those by lactose (red). C: Structures of LacNAc-loaded gal-3 CRD (PDB: 1A3K), with residues highlighted in red for the most shifted resonances upon ligand binding, followed by orange and yellow. Left: shifts in the presence of lactose; right: shifts in the presence of K.

5.3.2 The interaction of the ND of gal-3 with its CRD

In structural aspects, the modular design of galectin-3 (gal-3) is unique. It is composed of the common β-sandwich-type carbohydrate recognition domain (CRD) linked to collagen-like repeats (nine in human gal-3) and one N-terminal peptide with two sites for serine phosphorylation, the explanation why gal-3 is called a chimera-type galectin [37]. Since every animal species studied so far harbors gal-3 despite marked inter-species differences in the total number of galectin genes and the sequences in the two non-CRD regions appear conserved [2, 37, 76], the trimodular structure likely bears special physiological significance.

At present, the non-CRD portion is implicated in non-classical secretion [59, 77-78]; serine phosphorylation of the starting peptide (at S6 and S12 [78-80]) is known to play a role in nuclear export [59, 81-82]. Although an impact on binding lactose, ganglioside GM1 and N-glycans was excluded [12, 15, 16], this phosphorylation reduced gal-3 reactivity to laminin and asialomucin but enhanced both complex formation with the cell adhesion molecule L1 and its association with Thy-1-rich microdomains in neurons [83-84]. Equally intriguing, tyrosine phosphorylation by non-receptor kinases cAbl/Arg (at Y79, Y107 and Y118 [84-86]) appears to be a signal for gal-3 to reach the cell periphery with ensuing secretion in mouse embryonic fibroblasts [87]. It is an open question whether intramolecular recognition, influenced by phosphorylation status, may contribute to the structural basis of these processes.

Of note, the N-terminal portion appears to be rather flexible, although the transition temperature at 39 °C during thermal denaturation and the bimodal charge distribution in nano-ESI mass spectrometry provide evidence for an interconverting mixture of conformers, tentatively between extended and compact forms [4, 88]. Importantly, chemical shift analysis of hamster gal-3 suggested the presence of transient contacts with the CRD, for residues in the first part of the N-terminal tail (F5, W22, W26) and L109 [62]. The delineation of a nuclear export signal, probably interacting with a transporter such as nucleoporin Nup98 [89], in the distal section of the CRD, i.e. around 240-255 in murine gal-3 [90], in conjunction with these NMR-based observations and the switch-like impact of serine phosphorylation on nuclear export noted above, raises the intriguing possibility that serine phosphorylation can trigger so far unknown structural consequences, directly and/or indirectly.

As noted in previous sections, the assignments for the CRD part of full-length gal-3 were essentially the same as for the truncated protein, indicating that, in the overall structure, the CRD is folded in a very similar way as when truncated.

However, close inspection of a spectral overlay of both protein variants reveals selective differences in the chemical shifts of a number of signals. Among these, the most shifted ones correspond to residues covering a large area located on the convex side of the CRD β-sandwich, at the opposite face of the sugar binding site (fig. 33). These observations were strongly suggestive that, in the full-length protein, the ND establishes interactions with this region of the CRD, and motivated an investigation to unveil the structural details of this interaction by dissecting the contribution of each ND part.

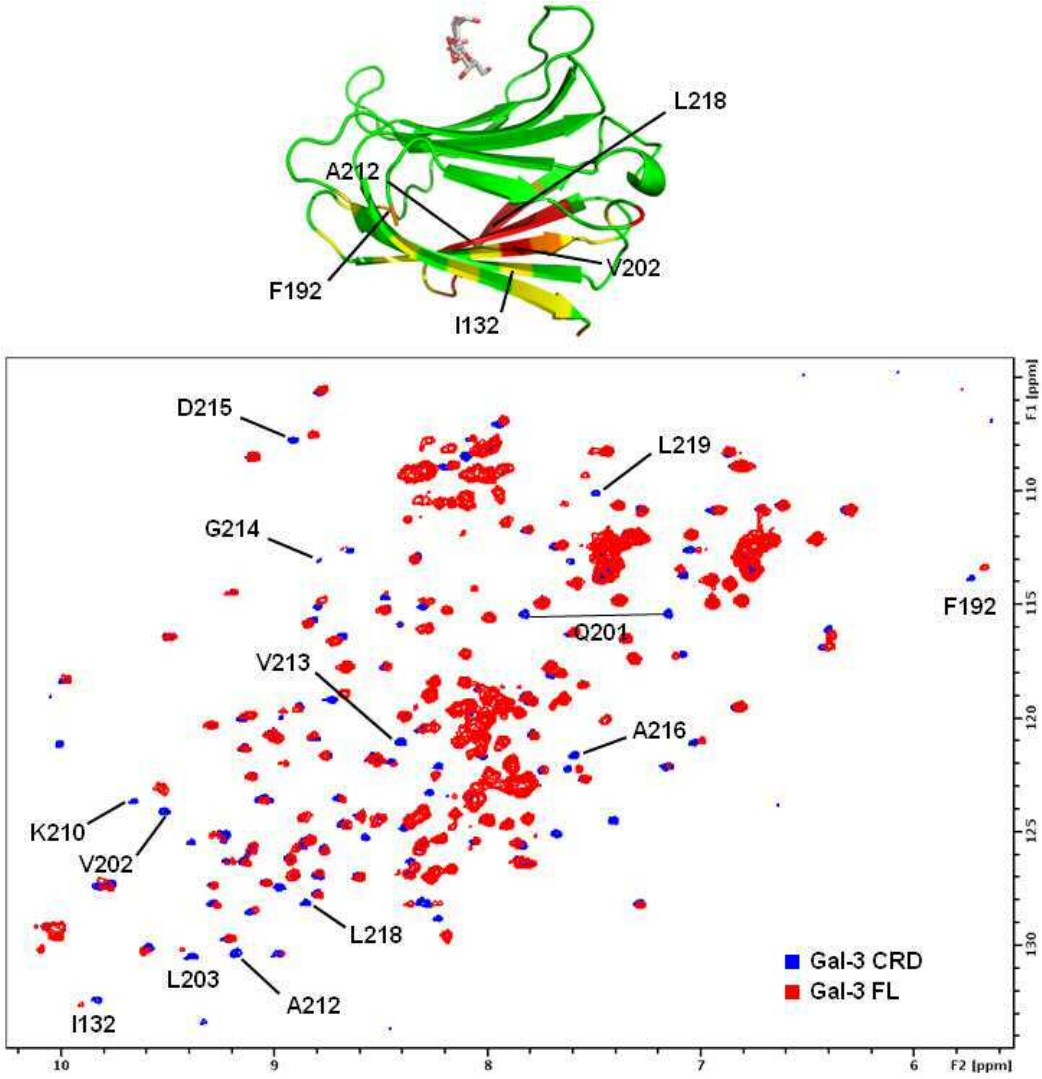


Figure 33: Bottom panel: superimposition of ^{1}H-^{15}N HSQC spectra of the gal-3 CRD (blue) and hGal-3 full-length (red), with labels indicating some of the most perturbed signals. Top panel: structure of the LacNAc-loaded gal-3 CRD (PDB: 1a3k), with residues highlighted in red for the most shifted resonances, followed by orange and yellow.

On this basis, we tested the hypothesis of the existence of interactions with synthetic peptides and ^{15}N-labeled CRD. Looking at secretion, tyrosine phosphorylation may play into this process by favoring a conformer, most suitably in extended form to make the tail fully accessible. Testing potential for peptide-CRD interactions followed the described strategy. To complete this study line, we also examined the chemical shift properties of the CRD in the presence of peptides comprising one and two collagen-like repeats. Peptides were synthesized in the laboratory of Prof. Dieter Kübler (German Cancer Research Center, Heidelberg, Germany).

5.3.2.1 Definition of peptides

To assess interactions between the gal-3 CRD and N-terminal tail, we used ^{1}H-^{15}N HSQC experiments and three sets of structurally-relevant tail-derived peptides as shown in Figure 34. The first set (P1-P3) represents the N-terminal peptide with and without Ser phosphorylation. The second set (P4-P7) encompasses sequences from the C-terminal part of the tail that contains phosphorylation sites Y107 and Y118. In the third set, we tested two peptides (P8, P9) from the collagen-like part of the tail.

5.3.2.2 N-terminal peptides

Although peptide P1 failed to show significant interactions with the CRD (fig. S7 in Appendix), its phosphorylated counterparts did (fig. 35). Here, the most markedly shifted resonances belonged to residues located at the backside of the lectin, opposite to the canonical sugar-binding site (fig. 36), which in turn leaves the β-galactoside (e.g. lactose) binding site unaffected. Significant chemical shift perturbations of NH cross peaks belonging to residues 207-211 suggests that K210 is likely the contact site for phosphate groups within the tail.

The use of two scrambled control phosphopeptides (SCR1 and SCR2) allowed us to conclude that these effects were specific. Despite the presence of two phosphate groups in SCR2, the NMR spectra of gal-3 CRD showed only very small chemical shift perturbations (fig. S9 in Appendix). When the distance between the two phosphates was maintained, as in SCR1, some minor chemical shift perturbations were observed (fig. S8 in Appendix), but not to the extent observed with the non-scrambled peptide. Overall, these results strongly suggest that the N-terminal peptide of human gal-3 can specifically interact with the CRD in a phosphorylation-dependent and sequence-specific manner.

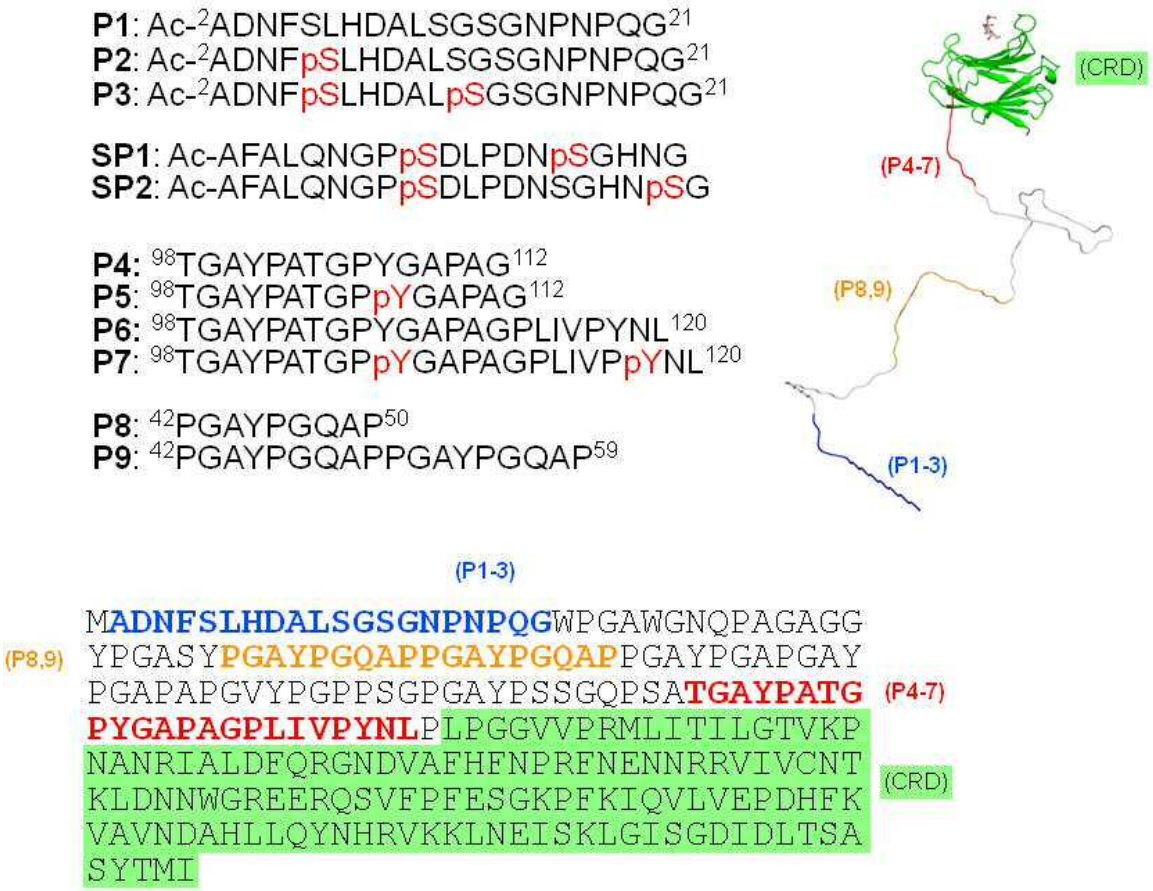


Figure 34: Top panel: sequence of the synthetic peptides used in this study, with phosphoserine (pS) and phosphotyrosine (pY) residues highlighted in red. Bottom panel: location of the peptides in the whole protein sequence. Red: peptides P1-P4; orange: peptides P5 and P6; blue: peptides P7-P9. The sequence of the CRD is indicated on a green background. On the right, a schematic representation of the crystal structure of the gal-3 CRD (green), with the ND (at scale) is shown.

5.3.2.3 Tyr-P peptides

Phosphorylation of Tyr107 and Tyr118 had the opposite effect. Whereas peptides P4-P7 elicited significant chemical shift perturbations of resonances from the CRD, their phosphorylated counterparts actually attenuated interactions with the CRD (figs. 37, 38). Once again, the most highly shifted signals belonged to residues located at the backside of the β-sandwich structure (fig. 36), and chemical shift differences were generally greater with the longer peptides P6 and P7. In any event, phosphorylation decreased the extent of chemical shift perturbations in the gal-3 CRD, indicating that affinity of these phosphorylated peptides towards the CRD is significantly reduced. Thus, phosphorylation of serine and tyrosine resiudes within the N-terminal tail appears to differentially modulate the interactions with the CRD.

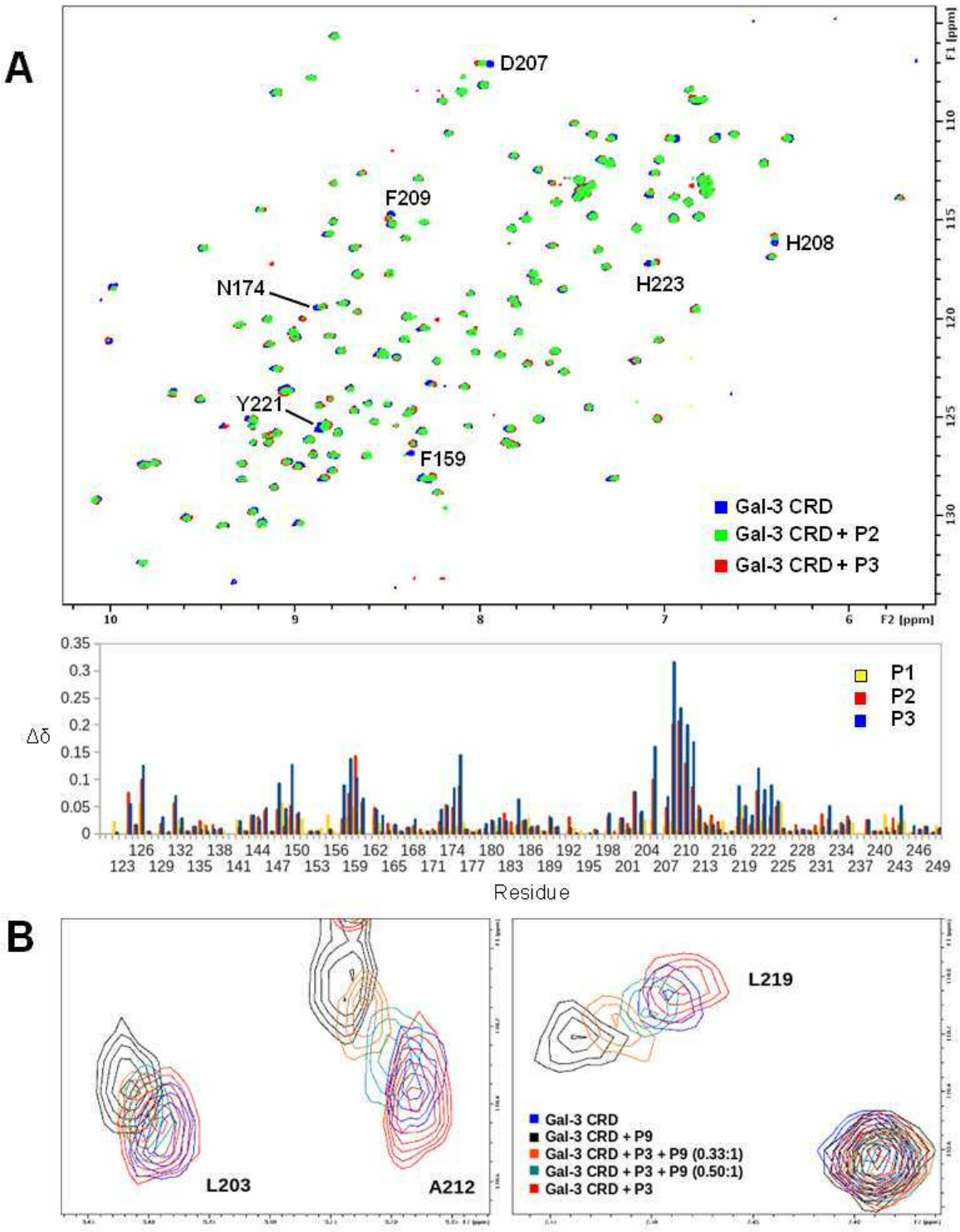


Figure 35: A: Top panel: superimposition of ^{1}H-^{15}N HSQC spectra of the gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of P2 (green) or P3 (red), with labels indicating some of the most perturbed signals. Bottom panel: ^{1}H-^{15}N-weighted chemical shift differences (Δδ) for gal-3 CRD backbone NH signals upon addition of 10 equivalents P1 (yellow), P2 (red) or P3 (blue). B: Superimposition of ^{1}H-^{15}N HSQC spectra of the gal-3 CRD with different P9/P3 ratios. Blue: gal-3 alone; black: in the presence of ten equivalents of P9; orange: after the addition of 3.3 equivalents P3; green: after the addition of 5 equivalents P3; red: in the presence of ten equivalents P3.

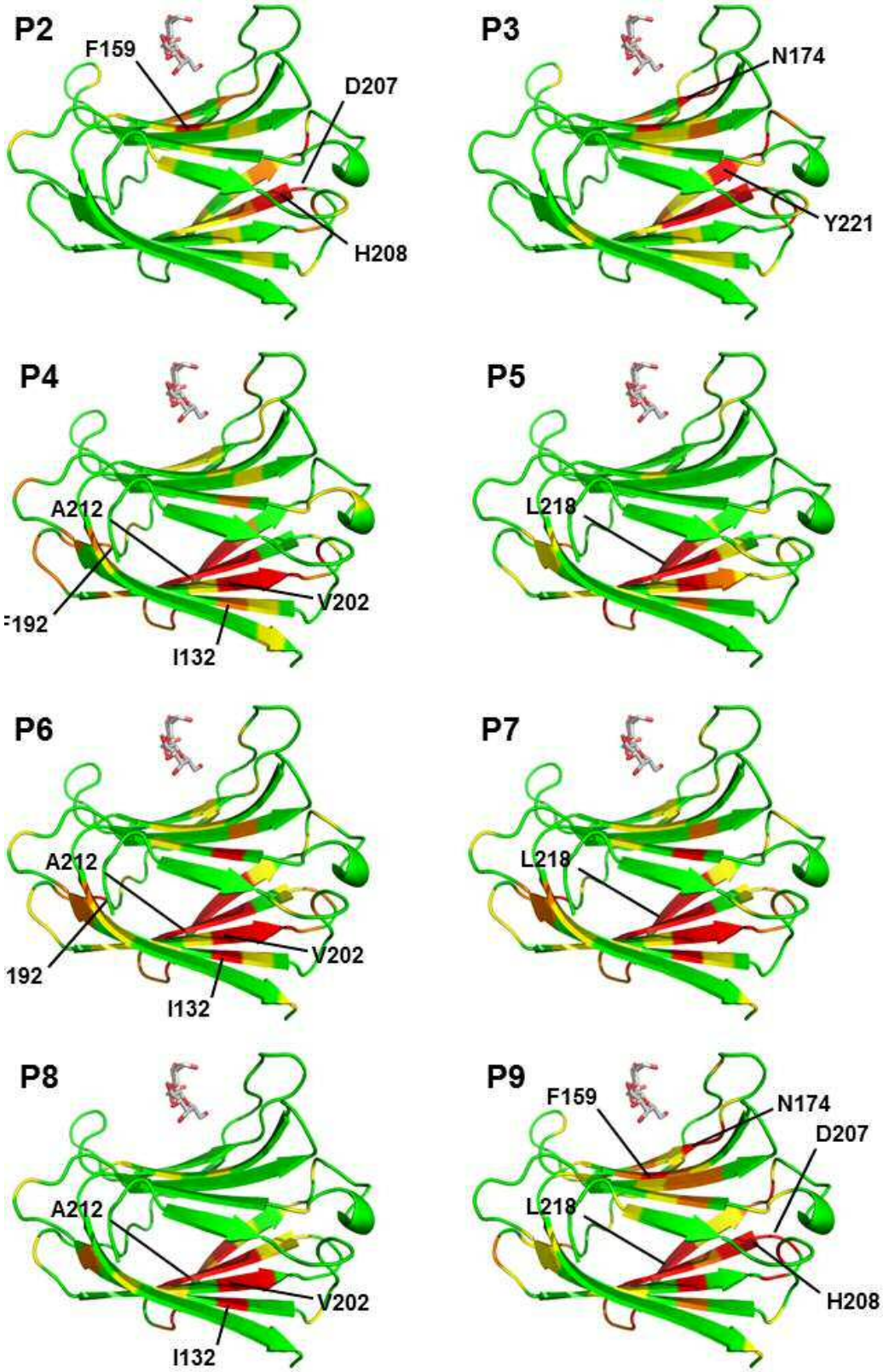


Figure 36: Mapping of the gal-3 CRD residues with the most shifted signals for a number of peptides. In all cases, the structure of the LacNAc-loaded human gal-3 CRD (PDB: 1A3K) is shown, with residues highlighted in red for the most shifted resonances, followed by orange and yellow.

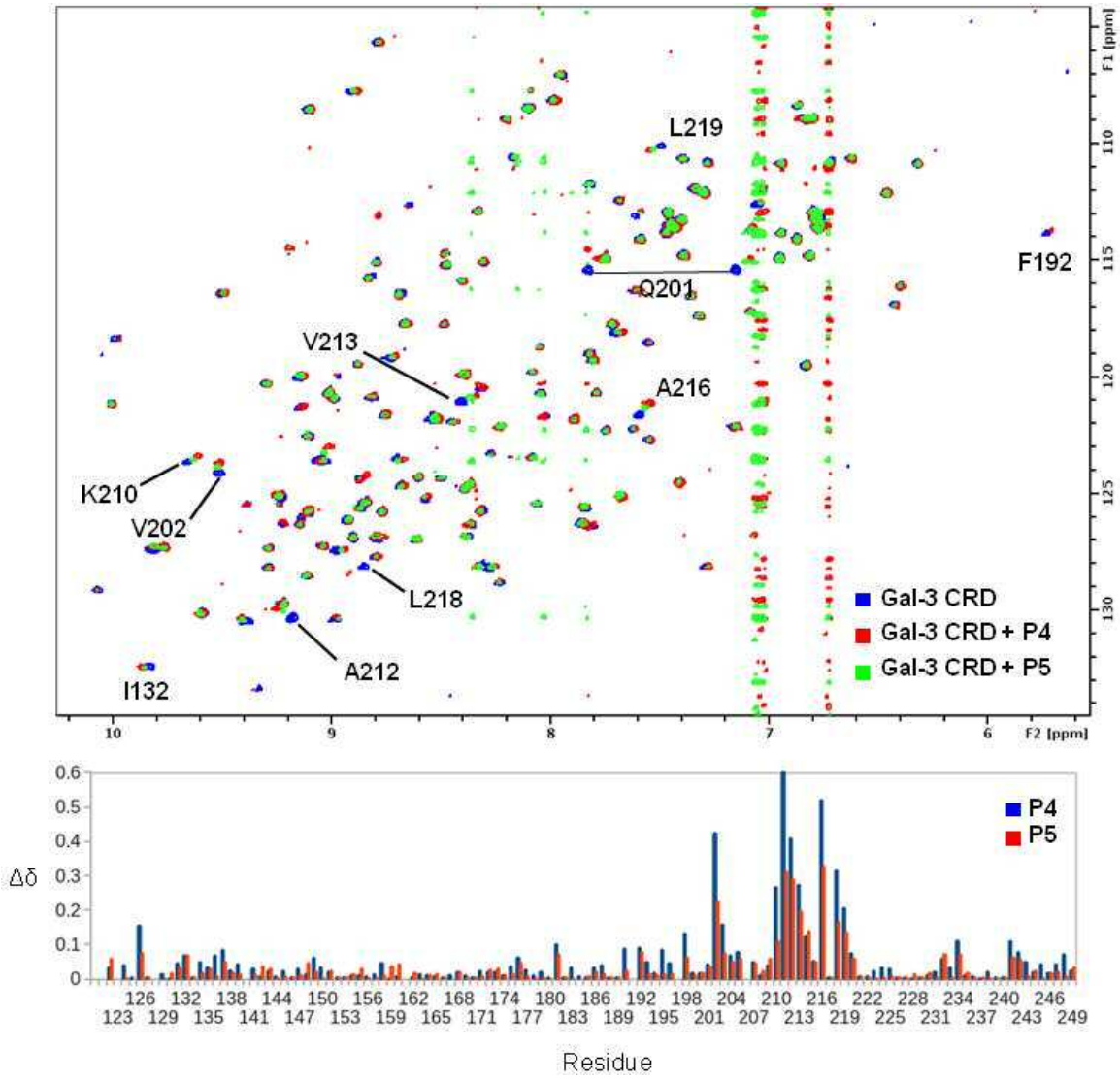


Figure 37: Top panel: superimposition of ^{1}H-^{15}N HSQC spectra of the gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of P4 (red) or P5 (green). Bottom panel: ^{1}H-^{15}N-weighted chemical shift differences (Δδ) for gal-3 CRD backbone NH signals upon addition of 10 equivalents P4 (blue) or P5 (red).

5.3.2.4 Pro/Gly-rich tandem-repeat peptides

To complete this study, we tested peptides containing one and two repeat sequences derived also from the tail, specifically one peptide comprising residues 42-50 (P8) and another one comprising residues 42-59 (P9). The observed chemical shift perturbations within the CRD were rather similar, compared to those exerted by peptides P4-P7, consistent with their sequence similarity (fig. 38, fig. 36). The profile of apparent contacts suggests that the region of interaction is somewhat larger for P9 than for the single-repeat variant peptide P8.

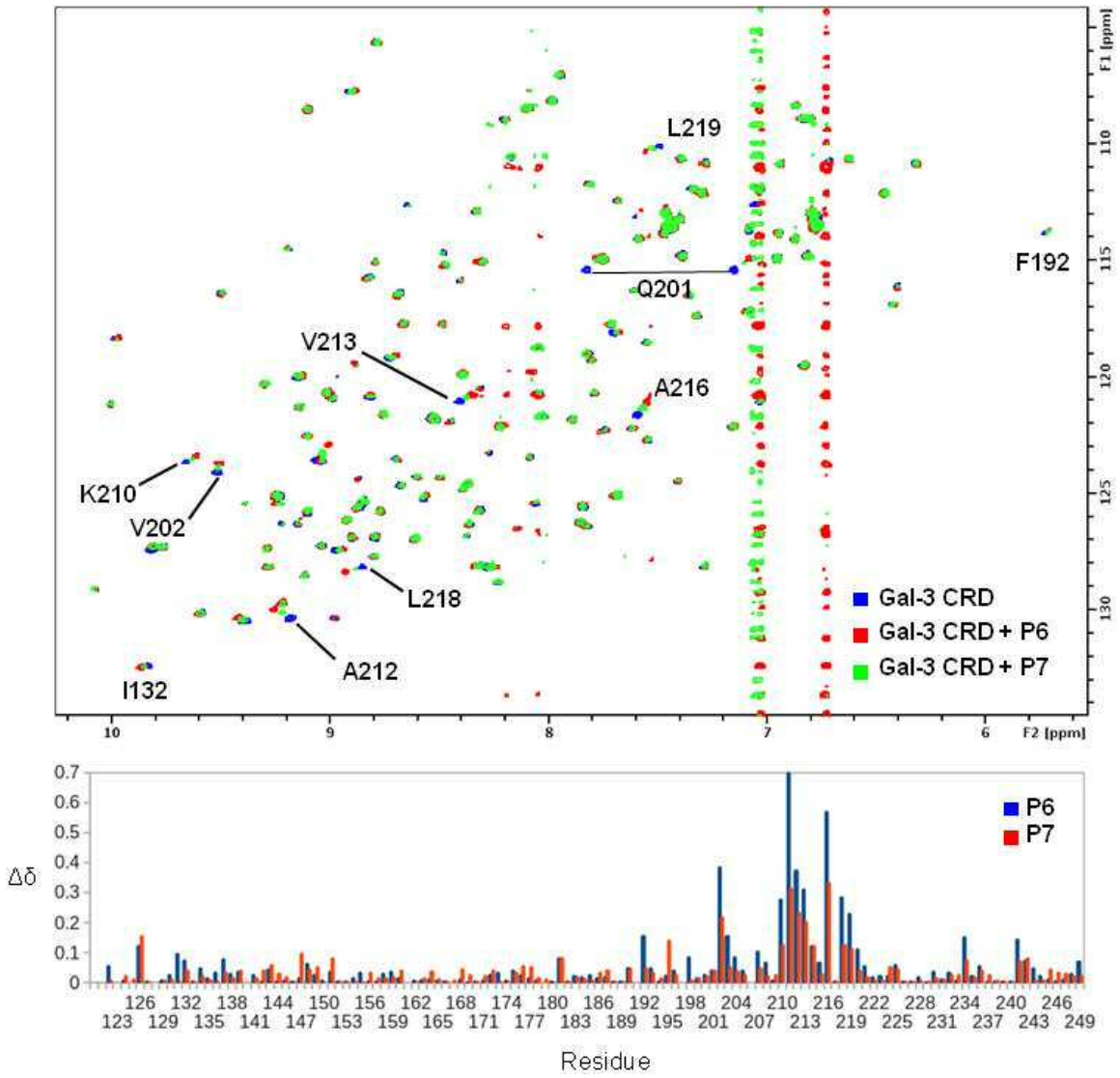


Figure 38: Top panel: superimposition of ^{1}H-^{15}N HSQC spectra of the gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of P6 (red) or P7 (green). Bottom panel: ^{1}H-^{15}N-weighted chemical shift differences (Δδ) for gal-3 CRD backbone NH signals upon addition of 10 equivalents P6 (blue) or P7 (red).

Because many of these peptides appeared to interact with the same region of the Gal-3 CRD, a competition experiment was performed to verify whether the phosphorylated N-terminal peptide P3 could displace peptide P9 from the CRD. The results of this experiment are illustrated in fig. 35B. Three signals (L203, A212 and L219) are significantly perturbed by the presence of P9, but not by that of P3. Addition of increasing amounts of P3 to the sample of gal-3 CRD + P9 reversed the chemical shift changes induced by P9, thus indicating that N-terminal phosphopeptide P3 can compete with the two tandem-repeat peptide P9 for binding to the CRD. These evidences suggest that phosphorylation of gal-3 at the N-terminus may alter the Gal-3 structure, potentially with recognition and/or functional consequences.

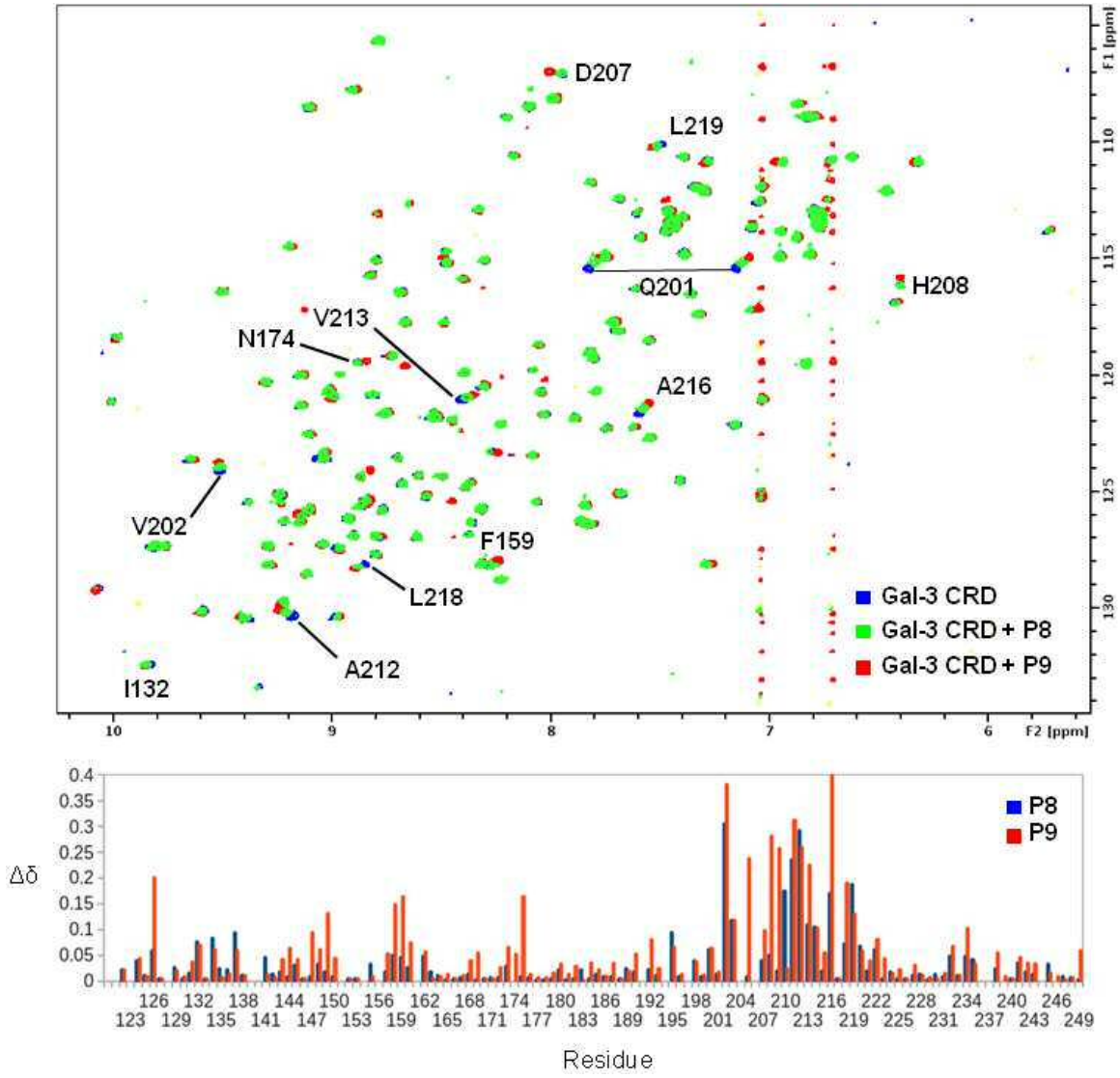


Figure 39: Top panel: superimposition of ^{1}H-^{15}N HSQC spectra of the Gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of P8 (green) or P9 (red). Bottom panel: ^{1}H-^{15}N-weighted chemical shift differences (Δδ) for Gal-3 CRD backbone NH signals upon addition of 10 equivalents P8 (blue) or P9 (red).

5.3.2.5 Conclusions

Interaction of the gal-3 CRD with synthetic (phospho)peptides derived from the gal-3 N-terminal tail was assessed using NMR HSQC experiments with ^{15}N-labeled CRD. Results show that distinct N-terminal peptides can interact at specific sites within the gal-3 CRD, suggesting that a similar behavior may take place in full length gal-3. Thus, the reported data can be relevant for association of tail-derived peptides (by proteolytic truncation of gal-3) with the lectin, as well as for intra- and intermolecular interactions, the latter in gal-3 aggregates

formed with multivalent ligands. Moreover, while phosphorylation of Ser6 and Ser12 in specific N-terminal peptides promotes association with the CRD, phosphorylation of Tyr107 and Tyr118 in other peptides attenuates this association. In this regard, serine and tyrosine phosphorylation of the N-terminal tail may act as on/off switches or modulators for various functions of Gal-3.

5.4 Methods

Section 5.2:

Protein assignment: 3D experiments for assignment of gal-7 were acquired in a Bruker AVANCE 800 MHz equipped with a cryoprobe. NMR data were processed in TopSpin 3.0. and analyzed with CARA.

Protein sequencing: the N-terminus of gal-7 was sequenced by standard Edman degradation chemistry in an Applied Biosystems 494 protein sequencer.

NMR experiments: ^{1}H-^{15}N HSQC spectra were recorded on gal-7 samples contained 0.2 mM uniformly ^{15}N-labeled gal-7 in 20 mM PBS buffer, pH 7.4 in H_2O, plus 10% D_2O for field-frequency lock. Data were collected at 298 K in a Bruker Avance 600 MHz spectrometer equipped with a cryogenically cooled HCN probe with z-axis gradients.

Peptide synthesis: A nonapeptide corresponding to the N-terminal sequence of human gal-7 (SNVPHKSSL) was synthesized using standard Fmoc solid phase synthesis in an Advanced Chemtech Peptide Synthesizer Model 348 Ω following manual attachment of the protected first amino acid (Leu-Fmoc) to 2-Cl-trityl chloride resin macrobeads (Peptides International, Louisville, KY).

Sequence alignment: Sequences of galectin-7 from different species were downloaded from the NCBI protein database and aligned with Clustal Omega. Alignments were manually edited with JalView to eliminate gaps and loaded with ClustalW2 to build the figures.

In silico modeling: A peptide model for the nonapeptide was built in xleap and subject to an energy minimization in Macromodel in the OPLS_2005 force field. A model for the gal-7 P4 *cis* conformer was generated by a manual 180º torsion of the V3-P4 peptide bond in the apo-gal-7 crystal structure (PDB: 1BKZ-A), followed by an energy minimization with Macromodel in the OPLS_2005 force field.

Fragment-based drug discovery: The Maybridge Ro3 fragment library (1000 compounds) was screened against gal-7. Mixes of 5 fragments (150 μM each) were screened by STD in the presence of 15 μM gal-7 in Tris-d_{11} buffer, p^{2}H 7.5, in a Bruker 500 MHz spectrometer. ^{1}H-^{15}N HSQC experiments were performed at 600 MHz, on 250 μM ^{15}N-labeled gal-7 or gal-3 CRD, using Tris buffer, pH 7.5 plus 10% D_2O at 298 K. Chemical shift perturbations were monitored, using the gal-7 and gal-3 ^{1}H and ^{15}N assignments. To obtain iLOEs, a NOESY 500 ms experiment was performed with 30 μM gal-7 plus 500 μM of fragments 205 and 245. Computational docking was performed with AutoDock 4.2 [91], treating gal-7 and gal-3 CRD X-ray structures (4GAL and 1A3K, respectively) as rigid. Ligand structures were retrieved from the manufacturer's database and treated as flexible.

Gal-7 dimer probing: Gal-7 dimer STD experiments were performed using 5-chloro-benzo[b]thiophene-3-carboxylic acid (i.e. fragment 245) as a spy molecule, at four different gal-7 concentrations around its dimerization K_D, ranging from 500 nM to 5 μM, in fully deuterated 20 mM phosphate buffer at pD 7.4. The effect of lactose was analyzed by repeating the experiments after the addition of 8 mM lactose to the samples.

Section 5.3

Protein assignment: Assignment of the gal-3-derived peptides was done by a combination of TOCSY, NOESY, and HSQC experiments performed in a Bruker AVANCE 600 MHz equipped with a cryoprobe.

NMR spectroscopy: All experiments were performed at 298 K in a Bruker AVANCE 600 MHz spectrometer equipped with a cryogenically-cooled z-gradient triple resonance probe. Samples for ^{1}H-^{15}N HSQC experiments contained ^{15}N-labeled human gal-3 (full length or CRD) at a concentration of 100 to 200 μM, with or without ten to 20 equivalents of the tested ligands. STD experiments were performed with 30-100 equivalents of ligands and 15 to 30 μM gal-3. Samples were dissolved in phosphate buffer in 90% H_2O/10% D_2O. Chemical shift perturbations were monitored using the sequence-specificassignments for the human gal-3 ^{1}H and ^{15}N resonances.

Modeling: The starting geometry for gal-3 CRD was the pdb 1A3K. The initial structure for the docking was built by manually docking the oligosaccharide structures into the lactose binding site. The docking was run in Autodock 4.2, with selective active torsions of the ligand, a grid box of 64×64×64 with a 0.375 Å spacing, maximum number of evaluations 25×10^7, maximum

number of generation 30000, 150 individual populations, using a Lamarckian genetic algorithm with 50 runs, and a clustering with an rms tolerance of 2.0 Å.

All proteins were produced in the laboratory of Prof. H.-J. Gabius (Ludwig Maximilian University of Munich, Germany).

5.5 References

[1] Gabius, H. J., Andre, S., Jimenez-Barbero, J., Romero, A., and Solis, D. (2011), *Trends Biochem Sci* **36**, 298-313.

[2] Kaltner, H., and Gabius, H. J. (2012), *Histol Histopathol* **27**, 397-416.

[3] Polyak, K., Xia, Y., Zweier, J. L., Kinzler, K. W., and Vogelstein, B. (1997), *Nature* **389**, 300-305.

[4] Kopitz, J., Andre, S., von Reitzenstein, C., Versluis, K., Kaltner, H., Pieters, R. J., Wasano, K., Kuwabara, I., Liu, F. T., Cantz, M., Heck, A. J., and Gabius, H. J. (2003), *Oncogene* **22**, 6277-6288.

[5] St-Pierre, Y., Campion, C. G., and Grosset, A. A. (2012), *Front Biosci (Landmark Ed)* **17**, 438-450.

[6] Saussez, S., and Kiss, R. (2006), *Cell Mol Life Sci* **63**, 686-697.

[7] Villeneuve, C., Baricault, L., Canelle, L., Barboule, N., Racca, C., Monsarrat, B., Magnaldo, T., and Larminat, F. (2011), *Mol Biol Cell* **22**, 999-1013.

[8] Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Kent Wenger, R., Yao, H., and Markley, J. L. (2008), *Nucleic Acids Res* **36**, D402-408.

[9] Nesmelova, I. V., Berbis, M. A., Miller, M. C., Canada, F. J., Andre, S., Jimenez-Barbero, J., Gabius, H. J., and Mayo, K. H. (2012), *Biomol NMR Assign* **6**, 127-129.

[10] Smith, S. P., Barber, K. R., and Shaw, G. S. (1997), *Protein Sci* **6**, 1110-1113.

[11] Winzor, D. J., and Jackson, C. M. (2006), *J Mol Recognit* **19**, 389-407.

[12] Lu, K. P., Finn, G., Lee, T. H., and Nicholson, L. K. (2007), *Nat Chem Biol* **3**, 619-629.

[13] Cierpicki, T., and Otlewski, J. (2001), *J Biomol NMR* **21**, 249-261.

[14] Byerly, D. W., McElroy, C. A., and Foster, M. P. (2002), *Protein Sci* **11**, 1850-1853.

[15] Zimmerman, J. M., Eliezer, N., and Simha, R. (1968), *J Theor Biol* **21**, 170-201.

[16] Ivanova, G., Yakimova, B., Angelova, S., Stoineva, I., and Enchev, V. (2010), *Journal of Molecular Structure* **975**, 330-334.

[17] Houzelstein, D., Goncalves, I. R., Fadden, A. J., Sidhu, S. S., Cooper, D. N., Drickamer, K., Leffler, H., and Poirier, F. (2004), *Mol Biol Evol* **21**, 1177-1187.

[18] Inagaki, Y., Higashi, K., Kushida, M., Hong, Y. Y., Nakao, S., Higashiyama, R., Moro, T., Itoh, J., Mikami, T., Kimura, T., Shiota, G., Kuwabara, I., and Okazaki, I. (2008), *Gastroenterology* **134**, 1180-1190.

[19] Yang, R. Y., Hsu, D. K., and Liu, F. T. (1996), *Proc Natl Acad Sci U S A* **93**, 6737-6742.

[20] Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E., and Kloog, Y. (2001), *Oncogene* **20**, 7486-7493.

[21] Rotblat, B., Niv, H., Andre, S., Kaltner, H., Gabius, H. J., and Kloog, Y. (2004), *Cancer Res* **64**, 3112-3118.

[22] Hajduk, P. J., and Greer, J. (2007), *Nat Rev Drug Discov* **6**, 211-219.

[23] Chianelli, D., Kim, Y. C., Lvovskiy, D., and Webb, T. R. (2003), *Bioorg Med Chem* **11**, 5059-5068.
[24] Wada, C. K., Holms, J. H., Curtin, M. L., Dai, Y., Florjancic, A. S., Garland, R. B., Guo, Y., Heyman, H. R., Stacey, J. R., Steinman, D. H., Albert, D. H., Bouska, J. J., Elmore, I. N., Goodfellow, C. L., Marcotte, P. A., Tapang, P., Morgan, D. W., Michaelides, M. R., and Davidsen, S. K. (2002), *J Med Chem* **45**, 219-232.
[25] Berbis, M. A., Andre, S., Canada, F. J., Pipkorn, R., Ippel, H., Mayo, K. H., Kubler, D., Gabius, H. J., and Jimenez-Barbero, J. (2014), *Biochem Biophys Res Commun* **443**, 126-131.
[26] Yan, J., Kline, A. D., Mo, H., Zartler, E. R., and Shapiro, M. J. (2002), *J Am Chem Soc* **124**, 9984-9985.
[27] Ermakova, E., Miller, M. C., Nesmelova, I. V., Lopez-Merino, L., Berbis, M. A., Nesmelov, Y., Tkachev, Y. V., Lagartera, L., Daragan, V. A., Andre, S., Canada, F. J., Jimenez-Barbero, J., Solis, D., Gabius, H. J., and Mayo, K. H. (2013), *Glycobiology* **23**, 508-523.
[28] Groves, P., Palczewska, M., Molero, M. D., Batta, G., Canada, F. J., and Jimenez-Barbero, J. (2004), *Anal Biochem* **331**, 395-397.
[29] Li, D., DeRose, E. F., and London, R. E. (1999), *J Biomol NMR* **15**, 71-76.
[30] Nesmelova, I. V., Dings, R. P. M., and Mayo, K. H. (2008), In *Galectins* (Klyosov, A. A., Witczak, Z. J.,Platt, D., Eds.), pp 33-69, John Wiley & Sons, Hoboken.
[31] Dias-Baruffi, M., Zhu, H., Cho, M., Karmakar, S., McEver, R. P., and Cummings, R. D. (2003), *J Biol Chem* **278**, 41282-41293.
[32] Battig, P., Saudan, P., Gunde, T., and Bachmann, M. F. (2004), *Mol Immunol* **41**, 9-18.
[33] Miura, T., Takahashi, M., Horie, H., Kurushima, H., Tsuchimoto, D., Sakumi, K., and Nakabeppu, Y. (2004), *Cell Death Differ* **11**, 1076-1083.
[34] Ahmad, N., Gabius, H. J., Sabesan, S., Oscarson, S., and Brewer, C. F. (2004), *Glycobiology* **14**, 817-825.
[35] Nishi, N., Abe, A., Iwaki, J., Yoshida, H., Itoh, A., Shoji, H., Kamitori, S., Hirabayashi, J., and Nakamura, T. (2008), *Glycobiology* **18**, 1065-1073.
[36] Herrmann, J., Turck, C. W., Atchison, R. E., Huflejt, M. E., Poulter, L., Gitt, M. A., Burlingame, A. L., Barondes, S. H., and Leffler, H. (1993), *J Biol Chem* **268**, 26704-26711.
[37] Cooper, D. N. (2002), *Biochim Biophys Acta* **1572**, 209-231.
[38] Hsu, D. K., and Liu, F. T. (2004), *Glycoconj J* **19**, 507-515.
[39] Hsu, D. K., Yang, R. Y., and Liu, F. T. (2006), *Methods Enzymol* **417**, 256-273.
[40] Cao, Z., Said, N., Amin, S., Wu, H. K., Bruce, A., Garate, M., Hsu, D. K., Kuwabara, I., Liu, F. T., and Panjwani, N. (2002), *J Biol Chem* **277**, 42299-42305.
[41] Gendronneau, G., Sidhu, S. S., Delacour, D., Dang, T., Calonne, C., Houzelstein, D., Magnaldo, T., and Poirier, F. (2008), *Mol Biol Cell* **19**, 5541-5549.
[42] Yang, R. Y., Rabinovich, G. A., and Liu, F. T. (2008), *Expert Rev Mol Med* **10**, e17.
[43] Dietz, A. B., Bulur, P. A., Knutson, G. J., Matasic, R., and Vuk-Pavlovic, S. (2000), *Biochem Biophys Res Commun* **275**, 731-738.
[44] Kim, K., Mayer, E. P., and Nachtigal, M. (2003), *Biochim Biophys Acta* **1641**, 13-23.
[45] Henderson, N. C., Mackinnon, A. C., Farnworth, S. L., Poirier, F., Russo, F. P., Iredale, J. P., Haslett, C., Simpson, K. J., and Sethi, T. (2006), *Proc Natl Acad Sci U S A* **103**, 5060-5065.
[46] Neidhart, M., Zaucke, F., von Knoch, R., Jungel, A., Michel, B. A., Gay, R. E., and Gay, S. (2005), *Annals of the rheumatic diseases* **64**, 419-424.
[47] Pelletier, I., and Sato, S. (2002), *J Biol Chem* **277**, 17663-17670.
[48] Pelletier, I., Hashidate, T., Urashima, T., Nishi, N., Nakamura, T., Futai, M., Arata, Y., Kasai, K., Hirashima, M., Hirabayashi, J., and Sato, S. (2003), *J Biol Chem* **278**, 22223-22230.

[49] Jin, R., Greenwald, A., Peterson, M. D., and Waddell, T. K. (2006), *J Immunol* **177**, 1289-1295.
[50] Greenwald, A. G., Jin, R., and Waddell, T. K. (2009), *Transplantation* **87**, 44-51.
[51] van den Berg, T. K., Honing, H., Franke, N., van Remoortere, A., Schiphorst, W. E., Liu, F. T., Deelder, A. M., Cummings, R. D., Hokke, C. H., and van Die, I. (2004), *J Immunol* **173**, 1902-1907.
[52] Davidson, P. J., Davis, M. J., Patterson, R. J., Ripoche, M. A., Poirier, F., and Wang, J. L. (2002), *Glycobiology* **12**, 329-337.
[53] Yu, F., Finley, R. L., Jr., Raz, A., and Kim, H. R. (2002), *J Biol Chem* **277**, 15819-15827.
[54] Yang, D., and Kay, L. E. (1996), *J Mol Biol* **263**, 369-382.
[55] Hughes, R. C. (1999), *Biochim Biophys Acta* **1473**, 172-185.
[56] Liu, F. T., Patterson, R. J., and Wang, J. L. (2002), *Biochim Biophys Acta* **1572**, 263-273.
[57] Nakahara, S., and Raz, A. (2007), *Cancer Metastasis Rev* **26**, 605-610.
[58] Chen, H. Y., Sharma, B. B., Yu, L., Zuberi, R., Weng, I. C., Kawakami, Y., Kawakami, T., Hsu, D. K., and Liu, F. T. (2006), *J Immunol* **177**, 4991-4997.
[59] Haudek, K. C., Spronk, K. J., Voss, P. G., Patterson, R. J., Wang, J. L., and Arnoys, E. J. (2010), *Biochim Biophys Acta* **1800**, 181-189.
[60] Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998), *J Biol Chem* **273**, 13047-13052.
[61] Saraboji, K., Hakansson, M., Genheden, S., Diehl, C., Qvist, J., Weininger, U., Nilsson, U. J., Leffler, H., Ryde, U., Akke, M., and Logan, D. T. (2012), *Biochemistry* **51**, 296-306.
[62] Birdsall, B., Feeney, J., Burdett, I. D., Bawumia, S., Barboni, E. A., and Hughes, R. C. (2001), *Biochemistry* **40**, 4859-4866.
[63] Morris, S., Ahmad, N., Andre, S., Kaltner, H., Gabius, H. J., Brenowitz, M., and Brewer, F. (2004), *Glycobiology* **14**, 293-300.
[64] Hirabayashi, J., Hashidate, T., Arata, Y., Nishi, N., Nakamura, T., Hirashima, M., Urashima, T., Oka, T., Futai, M., Muller, W. E., Yagi, F., and Kasai, K. (2002), *Biochim Biophys Acta* **1572**, 232-254.
[65] Dam, T. K., Gabius, H. J., Andre, S., Kaltner, H., Lensch, M., and Brewer, C. F. (2005), *Biochemistry* **44**, 12564-12571.
[66] Boscher, C., Dennis, J. W., and Nabi, I. R. (2011), *Curr Opin Cell Biol* **23**, 383-392.
[67] Umemoto, K., and Leffler, H. (2001), *J Biomol NMR* **20**, 91-92.
[68] Ippel, H., Miller, M. C., Berbis, M. A., Suylen, D., Andre, S., Hackeng, T. M., Canada, F. J., Weber, C., Gabius, H. J., Jimenez-Barbero, J., and Mayo, K. H. (2014), *Biomol NMR Assign*.
[69] Macher, B. A., and Galili, U. (2008), *Biochim Biophys Acta* **1780**, 75-88.
[70] Hancock, W. W. (1997), *World J Surg* **21**, 917-923.
[71] Canales, A., Mallagaray, A., Berbis, M. A., Navarro-Vazquez, A., Dominguez, G., Canada, F. J., Andre, S., Gabius, H. J., Perez-Castells, J., and Jimenez-Barbero, J. (2014), *J Am Chem Soc* **136**, 8011-8017.
[72] Diehl, C., Genheden, S., Modig, K., Ryde, U., and Akke, M. (2009), *J Biomol NMR* **45**, 157-169.
[73] Esko, J. D., Kimata, K., and Lindahl, U. (2009).
[74] Sasisekharan, R., Raman, R., and Prabhakar, V. (2006), *Annu Rev Biomed Eng* **8**, 181-231.
[75] Iwaki, J., Minamisawa, T., Tateno, H., Kominami, J., Suzuki, K., Nishi, N., Nakamura, T., and Hirabayashi, J. (2008), *Biochem Biophys Res Commun* **373**, 206-212.
[76] Kaltner, H., Kubler, D., Lopez-Merino, L., Lohr, M., Manning, J. C., Lensch, M., Seidler, J., Lehmann, W. D., Andre, S., Solis, D., and Gabius, H. J. (2011), *Anat Rec (Hoboken)* **294**, 427-444.
[77] Gong, H. C., Honjo, Y., Nangia-Makker, P., Hogan, V., Mazurak, N., Bresalier, R. S., and Raz, A. (1999), *Cancer Res* **59**, 6239-6245.

[78] Menon, R. P., and Hughes, R. C. (1999), *Eur J Biochem* **264**, 569-576.
[79] Cowles, E. A., Agrwal, N., Anderson, R. L., and Wang, J. L. (1990), *J Biol Chem* **265**, 17706-17712.
[80] Huflejt, M. E., Turck, C. W., Lindstedt, R., Barondes, S. H., and Leffler, H. (1993), *J Biol Chem* **268**, 26712-26718.
[81] Tsay, Y. G., Lin, N. Y., Voss, P. G., Patterson, R. J., and Wang, J. L. (1999), *Exp Cell Res* **252**, 250-261.
[82] Takenaka, Y., Fukumori, T., Yoshii, T., Oka, N., Inohara, H., Kim, H. R., Bresalier, R. S., and Raz, A. (2004), *Mol Cell Biol* **24**, 4395-4406.
[83] Mazurek, N., Conklin, J., Byrd, J. C., Raz, A., and Bresalier, R. S. (2000), *J Biol Chem* **275**, 36311-36315.
[84] Diez-Revuelta, N., Velasco, S., Andre, S., Kaltner, H., Kubler, D., Gabius, H. J., and Abad-Rodriguez, J. (2010), *J Cell Sci* **123**, 671-681.
[85] Balan, V., Nangia-Makker, P., Jung, Y. S., Wang, Y., and Raz, A. (2010), *Biochim Biophys Acta* **1803**, 1198-1205.
[86] Li, X., Ma, Q., Wang, J., Liu, X., Yang, Y., Zhao, H., Wang, Y., Jin, Y., Zeng, J., Li, J., Song, L., Li, P., Qian, X., and Cao, C. (2010), *Cell Death Differ* **17**, 1277-1287.
[87] Menon, S., Kang, C. M., and Beningo, K. A. (2011), *Biochem Biophys Res Commun* **410**, 91-96.
[88] Agrwal, N., Sun, Q., Wang, S. Y., and Wang, J. L. (1993), *J Biol Chem* **268**, 14932-14939.
[89] Funasaka, T., Balan, V., Raz, A., and Wong, R. W. (2013), *Biochem Biophys Res Commun* **434**, 155-161.
[90] Li, S. Y., Davidson, P. J., Lin, N. Y., Patterson, R. J., Wang, J. L., and Arnoys, E. J. (2006), *Glycobiology* **16**, 612-622.
[91] Goodsell, D. S., Morris, G. M., and Olson, A. J. (1996), *J Mol Recognit* **9**, 1-5.

CHAPTER 6

NEW DEVELOPMENTS: APPLICATION OF PARAMAGNETISM TO PROBE SUGAR CONFORMATION AND INTERACTION WITH RECEPTORS

6.1 Introduction

Currently, conformational analyses of carbohydrates by NMR make extensive use of nuclear Overhauser enhancements (NOE) and scalar couplings. However, despite the important information that can be extracted from these sources of information, additional parameters are usually required to define the global shape and the dynamics of complex glycans [1-3]. Indeed, NOEs and *J*-couplings only provide local information, and it is common to find oligosaccharides in extended shapes, with no long-range NOE contacts. Also, the flexibility of glycans often precludes the detection of NOEs with structural information [4]. This is aggravated by the inherent difficulty of carbohydrates due to signal overlapping problems.

In addition, studies on protein-carbohydrate interactions by NMR from the receptor perspective are often hampered by the weak signal perturbations that weakly-binding, neuter sugars exert on protein NMR spectra, and the scarcity of intermolecular NOE information. For these reasons, new tools are needed that boost the effect of sugar binding on the protein signals and introduce additional restraints for structural elucidation of protein-carbohydrate complexes by NMR.

The exploitation of paramagnetic effects, especially those exerted by lanthanide ions, offers a potential to tackle these challenges, thanks to a rich source of long-range structural restraints that they provide, including residual dipolar couplings, relaxation enhancement and pseudocontact shifts. The use of paramagnetic restraints has traditionally been restricted to proteins and nucleic acids [5-6]. In this chapter, we show how they can be used as key parameters in the structural characterization of carbohydrates and carbohydrate-protein complexes.

The work presented in this chapter was done in close collaboration with Dr. Angeles Canales (Complutense University of Madrid and CIB-CSIC), and Dr. Alvaro Mallagaray of Prof. Javier Pérez-Castells lab (University San Pablo CEU, Madrid).

6.1.1 Theoretical background

Paramagnetism arises from the presence of unpaired electrons in atomic or molecular orbitals of certain chemical species, notably metal ions and organic radicals. Among the latter, nitroxide spin labels are notable due to their stability in aqueous solutions. For their part, metal ions with paramagnetic properties include Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} and most lanthanide (III) cations (Ln^{3+}).

6.1.1.1 Isotropic and anisotropic magnetic susceptibility

The paramagnetism that is originated in a paramagnetic center can be described by its magnetic susceptibility tensor χ, which in turn can be expanded in its three spatial components χ_x, χ_y and χ_z.

The χ tensor is anisotropic if the magnetic moment of the paramagnetic center varies as a function of the orientation of the molecule inside the main magnetic field, B_0; otherwise, the tensor is defined as isotropic. Gd^{3+} is the only lanthanide ion with an isotropic χ tensor, while the rest of paramagnetic lanthanide (III) ions possess strongly anisotropic χ tensors.

Since the major part of paramagnetic effects is accounted for by the deviation of the χ tensor from spherical symmetry only, this tensor is usually decomposed in its isotropic (χ_{iso}) and anisotropic (the $\Delta\chi$ tensor) components. The latter is in turn characterized by its axial and rhombic components:

$$\Delta\chi_{ax} = \chi_z - \frac{(\chi_x + \chi_y)}{2} \quad (1)$$

$$\Delta\chi_{rh} = \chi_x - \chi_y \quad (2)$$

Thus, the $\Delta\chi$ tensor is null when $\chi_x = \chi_y = \chi_z$. The χ tensor anisotropy depends on the coordination environment.

6.1.1.2 Types of paramagnetic effects

While unpaired electron spins of nitroxide radicals and some cations (e.g. Gd^{3+}, Mn^{2+} and Cu^{2+}) are characterized by long relaxation times (in the nanosecond range), metals with strongly anisotropic χ tensors (most lanthanide ions) display very short (sub-picosecond) relaxation times [7]. This difference is accounted for by different relaxation mechanisms, which are ultimately responsible for the different types of paramagnetic effects observed by NMR [5].

6.1.1.2.1 Paramagnetic relaxation enhancement (PRE)

PREs arise from dipolar interactions between the unpaired electrons and the nuclei in its vicinity. Thus, they are dependent on the nuclear spin (γ_I) and its distance to the paramagnetic center (r). Specifically, PREs due to dipolar relaxations operating via the Solomon mechanism (long relaxation times, like in nitroxide radicals, Mn^{2+} and Gd^{3+} ions) or the Curie mechanism (short relaxation times, like in most Ln^{3+} ions) show the same r^{-6} dependence. For dipolar relaxation:

$$R_{2,S}^{PRE} \propto \frac{\gamma_I^2 \gamma_S^2 \tau_c}{r^6} \tag{3}$$

Where $1/\tau_c = 1/\tau_r + 1/\tau_e$ (τ_r being rotational correlation time and τ_e the electron spin state halftime), and γ_S is the electron spin. For Curie-spin relaxation:

$$R_{2,C}^{PRE} \propto \frac{\gamma_I^2 B_0^2 \left[g_e S_e (S_e + 1)\right]^2 \tau_r}{r^6} \tag{4}$$

Where g_e and S_e are the electronic factor g and the spin, respectively, T is temperature and B_0 is the strength of the main magnetic field. The inverse sixth power distance dependence means that PREs are strong close to the paramagnetic center, where they cause severe broadening of the NMR signals beyond its detection limit, but are minimized at long distances. Thus, their study is suitable around regions close enough to the paramagnetic center to be measured, but where they are not too strong to not be detected.

Today, the use of PREs is widely extended in structural studies of biological macromolecules [8]. In the context of carbohydrate-protein interactions, it is worth mentioning the application of nitroxide spin-labels for mapping the carbohydrate binding sites on protein surfaces, thanks to PREs arisen from the nitroxide moiety. Perturbations in the ^{1}H-^{15}N HSQC spectrum of galectin-3 (gal-3) were monitored and used to identify protein residues proximate to the binding site for *N*-acetyllactosamine [9].

6.1.1.2.2 Paramagnetic shifts

NMR signals affected by the presence of unpaired electrons in its vicinity can experience an external net contribution to their chemical shift called hyperfine shift. This contribution can be due to two different mechanisms:

6.1.1.2.2.1 Contact shifts

If the unpaired electrons are delocalized over the resonant nucleus due to chemical bonding, the NMR signal of the latter will experience a shift called Fermi contact shift (δ^{c}). The magnitude of δ^{c} contains structural information, but it is very difficult to extract, although successful attempts to use them in structural calculations have been reported [10-12].

6.1.1.2.2.2 Pseudo-contact shifts (PCS)

If the magnetic moment of the unpaired electrons is anisotropic, there is an additional spatial dipolar coupling causing a hyperfine shift on vicinal nuclei called pseudo-contact shift (PCS). The net hyperfine shift of every nucleus in a paramagnetic molecule is a sum of contact and pseudo-contact shifts. However, in the practice, when the number of chemical bonds between the resonant nucleus and the paramagnetic center is above four, and none of them are π-bonds, hyperfine shifts can be considered as entirely caused by PCS.

PCSs are proportional to the anisotropy of the χ tensor ($\Delta\chi$) and independent of the isotropic component of the χ tensor. It is therefore sufficient to describe the PCS as a function of the axial and rhombic components of the $\Delta\chi$ tensor, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, as seen in eq. 5, [13] where $\Delta\delta^{PCS}$ is the difference in chemical shifts measured between diamagnetic and paramagnetic

samples, *r* is the distance between the metal ion and the nuclear spin, and θ and ϕ are the polar coordinates describing the position of the nuclear spin with respect to the principal axes of the Δχ tensor.

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}\left(3\cos^2\theta - 1\right) + \frac{3}{2}\Delta\chi_{rh} sen^2\theta\cos 2\varphi\right] \quad (5)$$

PCS give rise to large changes in chemical shifts that decrease with the distance between the metal ion and the nuclear spin with a r^{-3} dependence. Therefore, PCS can provide long-range structural information. PCSs were first described in 1958 [14] and successfully used for structural studies on small nucleotides in 1971 [15]. Today, PCSs are widely used for calculation and refinement of the structure of biomolecules, especially proteins [5-6].

As shown in fig. 1, PCSs are usually represented as isosurfaces superimposed to the 3D structure of the paramagnetic molecule. These isosurfaces represent the spatial coordinates for which eq. 5 predicts the same PCS values.

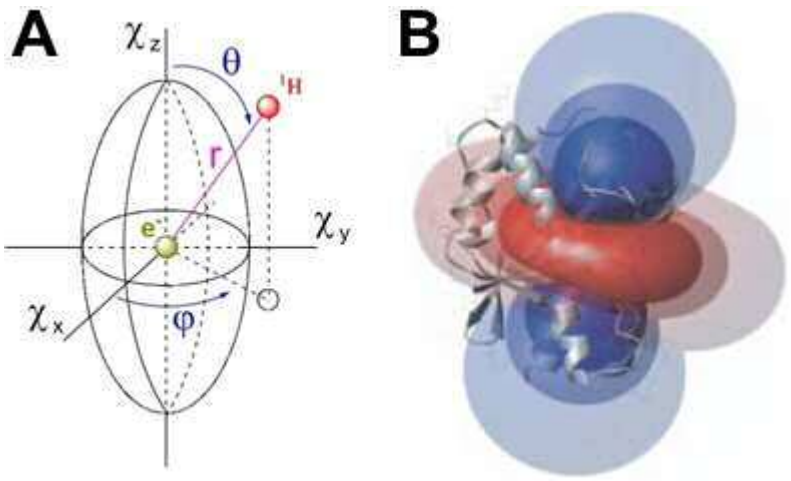


Figure 1: Pseudo-contact shifts (PCSs). A: An electron-proton system showing the definition of *r* and angular parameters θ and ϕ. B: PCS isosurfaces plotted on a paramagnetic-labeled protein. Figure from [16].

6.1.1.2.3 Residual dipolar couplings (RDC)

The 1-bond scalar coupling costant ^1J of a given nuclear pair (e.g. ^{1}H-^{13}C, ^{1}H-^{1}H, ^{1}H-^{15}N) may have a dipolar contribution (*D*) due to partial alignment of the biomolecule in the presence of

the external magnetic field B_0. This contribution, called residual dipolar coupling (RDC), causes splitting in the signal of each nucleus involved. The magnitude of this splitting depends on two geometrical parameters: the distance *r* between both nuclei and the angle θ between the vector connecting these nuclei and the direction of B_0. For a given pair of nuclei, for instance a ^{1}H-^{13}C pair in a carbohydrate, the dipolar interaction is given by:

$$D_{(^{1}H-^{13}C)} = \frac{\mu_0 \gamma_H \gamma_C h}{\left\langle 2\pi r_{HC} \right\rangle^3} \left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \quad (6)$$

Where the terms enclosed in brackets correspond to time-averaged distance $\langle r_{HC}^{-3} \rangle$ and time-averaged angle $\langle \cos^2 \theta \rangle$. Normally, solutes tumble isotropically; thus, $\langle \cos^2\theta \rangle$ is zero, leading to null dipolar splitting. In contrast, when a solute is weakly oriented in an alignment medium, there is a net dipolar contribution *D* to ^{1}J. For this reason, RDC measurement requires partial alignment of the molecules under study [17-18]. Several alignment media have been tested for carbohydrates, using both external alignment systems [18-20] and paramagnetic tags [21].

From the experimental viewpoint, RDCs simply add to splitting due to *J*-coupling. Thus, for a given NMR signal, the magnitude of the RDC is determined from the difference in splitting between the sample under alignment (RDC + *J*) and the isotropic non-aligned sample (*J*).

The great advantage of RDCs is that they permit to correlate the geometrical orientation between structural fragments (e.g. pyranose rings in the carbohydrate field) no matter how apart they are in the molecule [22]. Thus, RDCs are particularly useful to deduce the geometry of the glycosidic linkages, which has a major influence on the conformation of carbohydrates and their overall shape [23].

RDCs have also been successfully employed for the study of protein-carbohydrate complexes. In this regard, RDCs have been used to determine the carbohydrate conformation in the bound state [24], the location of the carbohydrate binding site and the structure of the complex, in combination with PCS-derived information [25].

6.1.2 Paramagnetic labeling of biomolecules

Paramagnetic effects have been extensively exploited for structural characterization of biological macromolecules, mainly proteins [6, 26]. Since most biological molecules lack natural paramagnetic centers, these must be artificially incorporated.

A possible strategy consists in the replacement of naturally binding ions (e.g. in calcium-binding proteins) by paramagnetic (e.g. lanthanide) ions [27-28]. However, since Ln^{3+} ions have an additional charge and a higher coordination number than Ca^{2+}, this method does not always provide satisfactory results. Similar approaches have been reported with proteins that bind Mg^{2+} or Mn^{2+} [16, 29-30].

However, since a majority of proteins do not bind metals, the most extended strategy involves covalent labeling of biomolecules with paramagnetic tags. Labeling is selectively done at reactive functional groups of proteins (e.g. solvent-accessible cysteine thiols or lysine amines), which can exist in the natural molecule or be mutagenically incorporated. Paramagnetic tags can incorporate either nitroxide radicals or lanthanide chelators (fig. 2), and many of them are available in a commercial basis [8].

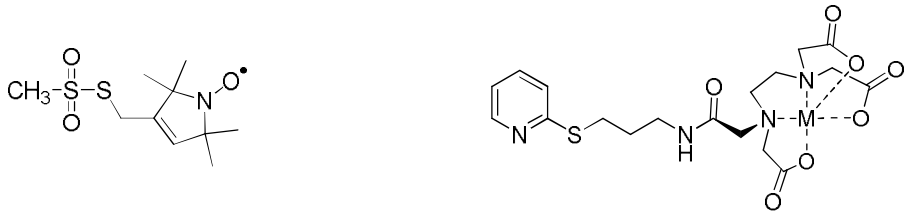


Figure 2: Two examples of cysteine-reactive paramagnetic tags, one incorporating a nitroxide spin label (left), and one incorporating a metal chelator (right).

6.1.2.1 Lanthanide ions

In lanthanide ions (Ln^{3+}), unpaired electrons are located in the non-reactive *f* orbitals. Since interactions involving these metals and their ligands are generally ionic in nature, these electrons do not tend to delocalize by forming chemical bonds, as it is usual, for instance, in hemo-iron paramagnetic complexes. As a result, contact shifts (the paramagnetic equivalents of scalar couplings) are restricted to nuclei in direct contact with the lanthanide ion, and the majority of paramagnetic effects that they exert can be explained by PREs and PCSs.

Trivalent cations of the different lanthanide elements differ substantially in their paramagnetic properties (fig. 3), namely their χ tensor values and their associated anisotropy. Looking at their $\Delta\chi$ tensors, lanthanide ions can be classified as strongly paramagnetic (Dy^{3+}, Tb^{3+}), moderately paramagnetic (e.g. Er^{3+}, Yb^{3+}), and weakly paramagnetic (e.g. Eu^{3+}, Ce^{3+}, Sm^{3+}). Gd^{3+} represents a special case, because it has an isotropic χ tensor. In addition, it induces strong PREs without detectable PCSs. Of note, there also exist diamagnetic lanthanides (La^{3+}, Lu^{3+}), which are used as diamagnetic references for their paramagnetic counterparts, together with other diamagnetic trivalent cations of similar radii like Y^{3+} and Sc^{3+}.

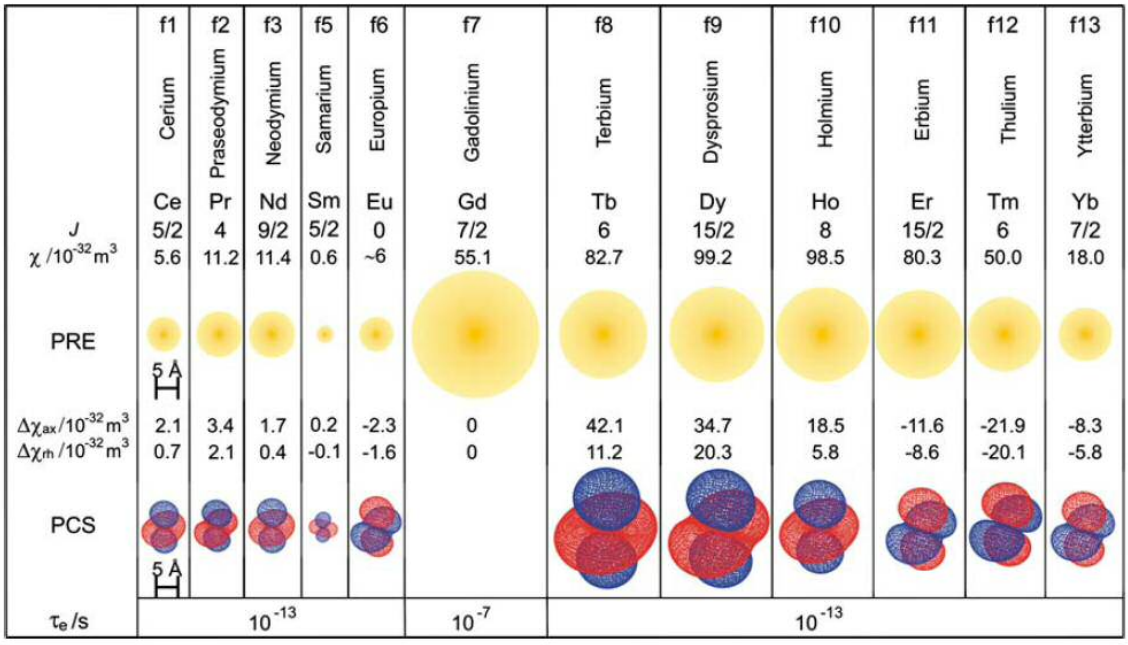


Figure 3: Paramagnetic properties of Ln^{3+} ions, including χ tensors and the axial and rhombic components of their anisotropic $\Delta\chi$ tensors, magnitude of PRE and isosurface representation of their PCSs, and typical electronic relaxation times (τ_e) at 800 MHz (figure from [31]).

6.2 Results

6.2.1 Assessing sugar conformation in solution by PCS measurement

The synthesis and evaluation of a molecule of chitobiose covalently attached to a lanthanide binding tag was previously reported by our laboratory, in collaboration with that of Prof.

Pérez-Castells [32], as the first lanthanide-chelating molecule in the realm of carbohydrates. Its design consisted in a 1-β-aminochitobiose derivative connected to an ethylenediamine-tetraacetic acid (EDTA) chelator unit by a rigid biphenyl linker. The role of the linker was to position the sugar nuclei beyond the distance range at which PREs lead nuclear signals to became broadened beyond detection. Large PCSs (of up to 0.22 ppm) were measured in the sugar moiety [32].

Next, a second generation of lanthanide binding tags (LBT) was designed, incorporating a phenylendiaminetetraacetic acid (PhDTA)-based chelator, which presented multiple advantages. First, it lacked stereogenic centers, thus facilitating its synthesis. Second, its structure gained in rigidity, boosting paramagnetic effects thanks to minimizing structure averaging in solution.

Using a lactose derivative bound to a biphenyl-DTA (Lac-biPhDTA), PCSs of the lactose moiety were measured by comparing the ^{13}C and ^{1}H-NMR chemical shifts in the presence of the diamagnetic reference (La^{3+}) and two different paramagnetic ions (Tb^{3+}, Dy^{3+}) through standard ^{1}H-^{13}C HSQC experiments. All sugar signals were detected in the spectrum, and only one set NMR of signals was observed for the molecule in the presence of stoichiometric amounts of these metals.

A comparison of the spectra of Lac-biPhDTA coordinated with either La^{3+} or Tb^{3+} is shown in fig. 4, with very large PCSs in the sugar moiety in the presence of the paramagnetic ion. These satisfactory results prompted us to explore the use of this molecule to assess the conformation of lactose by PCS measurement.

The goal was two-fold: first, to demonstrate that the shape of the molecule was not affected by the presence of both the biphenyl linker and the DTA chelator unit. Second, to provide a proof-of-principle example that these derivatives can be useful to determine carbohydrate conformation in solution.

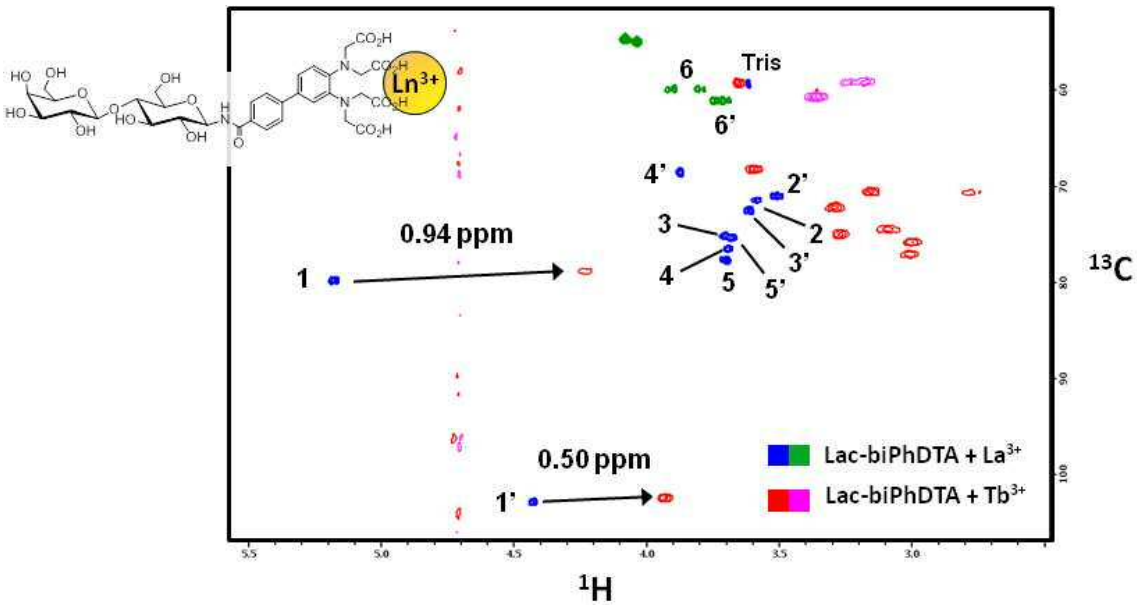


Figure 4: Formula and ^{1}H-^{13}C HSQC spectra of lac-biPhDTA complexed with a diamagnetic lanthanide ion (La^{3+}, blue/green) and a paramagnetic lanthanide ion (Tb^{3+}, red/pink).

The conformation of lactose has been previously studied in detail, using NMR and molecular mechanics [23, 33-34]. These studies have shown that lactose exists in an ensemble of conformations in solution, with a predominance of the typical $^{4}C_{1}$ chair pucker in both pyranose rings, with a major conformation (> 95%) around the glycosidic bond (*syn*Φ/*syn*Ψ) in coexistence with two other minor conformations (*syn*Φ/*anti*Ψ and *anti*Φ/*syn*Ψ, > 5%).

Calculations were performed with MSpin [35] software to deduce the theoretical PCS values using the geometries of these three lactose conformations. PCSs predicted for the major conformer (*syn*Φ/*syn*Ψ) were in excellent agreement with the experimental data (fig. 5). The Cornilescu's quality factor *Q* was 0.050 for Tb^{3+} and 0.006 for Dy^{3+}, for ^{1}H PCSs (the closer the *Q* factor to zero, the better the given trial structure fits the empirical data). The corresponding data for both *anti* conformers were much worse, (*Q* = 3.727 for the *syn*Φ/*anti*Ψ conformer and *Q* = 1.859 for the *anti*Φ/*syn*Ψ conformer).

These results show that PCS measurement is a valid approach to probe carbohydrate conformation in solution, paving the way for analogous analyses using larger, more complex glycans. In this regard, a similar approach has recently afforded the determination, for the first time, of the conformation of a biantennary *N*-glycan nonasaccharide by our group [36].

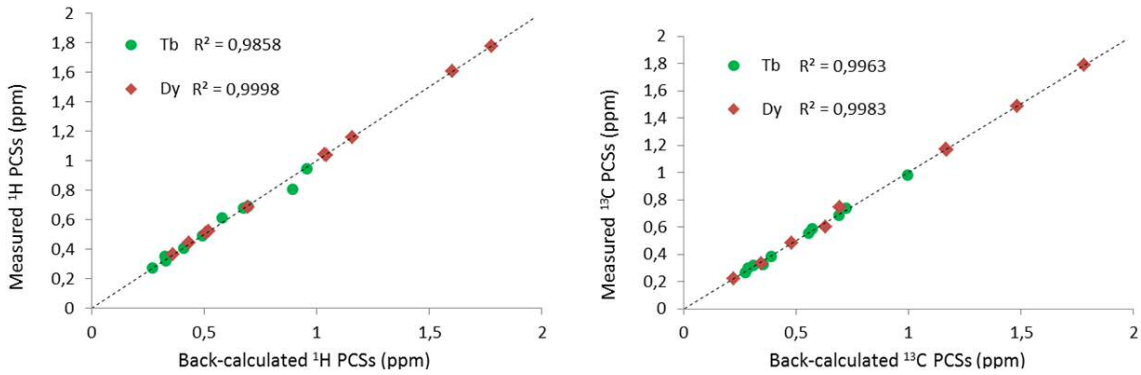


Figure 5: Plots correlating experimentally obtained and back-calculated ^{1}H and ^{13}C PCSs for the geometry of Lac-biPhDTA in the major lactose conformation, in the presence of either Tb^{3+} or Dy^{3+}.

6.2.2 Application to molecular recognition

The detection of large PCSs in the sugar moiety of Lac-biPhDTA prompted us to explore the applications of this technology to molecular recognition.

Since PCSs are originated by dipolar interactions which depend on distance, such effects could, in principle, be detected at the receptor signals, transferred through-space from the non-covalently binding paramagnetic sugar, to the protein nuclei. We looked into this using a well-known example of biomedical interest: carbohydrate binding to human gal-3, using Lac-biPhDTA and ^{15}N-labeled gal-3 CRD.

First, a control ^{1}H-^{15}N HSQC experiment with Lac-biPhDTA loaded with La^{3+} was performed, indicating the chemical shift perturbations induced in the gal-3 signals were basically identical to those obtained in the presence of just lactose, without the LBT moiety. (fig. 6). These observations were supported by the assessment that the modified ligand was equally recognized as lactose, using epitope mapping by STD (fig. 7). Thus, the binding modes of Lac-biPhDTA and lactose to gal-3 are equivalent.

Next, we compared ^{1}H-^{15}N HSQC spectra of gal-3 in the presence of Lac-biPhDTA loaded with La^{3+} (i.e. in diamagnetic conditions) and loaded with different paramagnetic lanthanides: Dy^{3+}, Tb^{3+} and Tm^{3+}. In all cases, PCSs were observed for many protein signals, as a result of the interaction between the protein and the carbohydrate analogue (fig. 8). Both positive and negative chemical shift changes were observed in the ^{1}H-^{15}N signals of gal-3 upon addition of Lac-biPhDTA loaded with the different paramagnetic metals, compared with the control

experiment in the presence of Lac-biPhDTA:La^{3+} (fig. 9). In addition, depending on the employed metal, some ^{1}H-^{15}N cross peaks became severely broadened, with some of them even disappearing, probably due to PREs. This effect was especially apparent in the case of Dy^{3+}, as expected from its stronger paramagnetic properties (fig., 3).

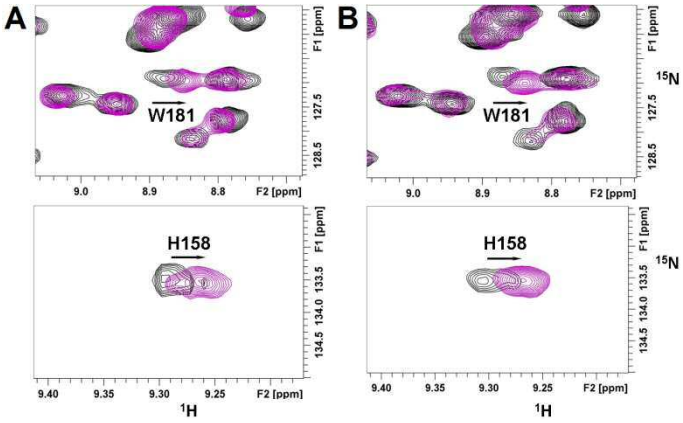


Figure 6: Right panels: ^{1}H-^{15}N HSQC (acquired at 700 MHz) for the gal-3:lactose (molar ratio 1:12) sample. Left panels: ^{1}H-^{15}N HSQC (acquired at 600 MHz) for the sample containing gal-3 and the lactose derivative 5, loaded with La^{3+} (molar ratio 1:12).

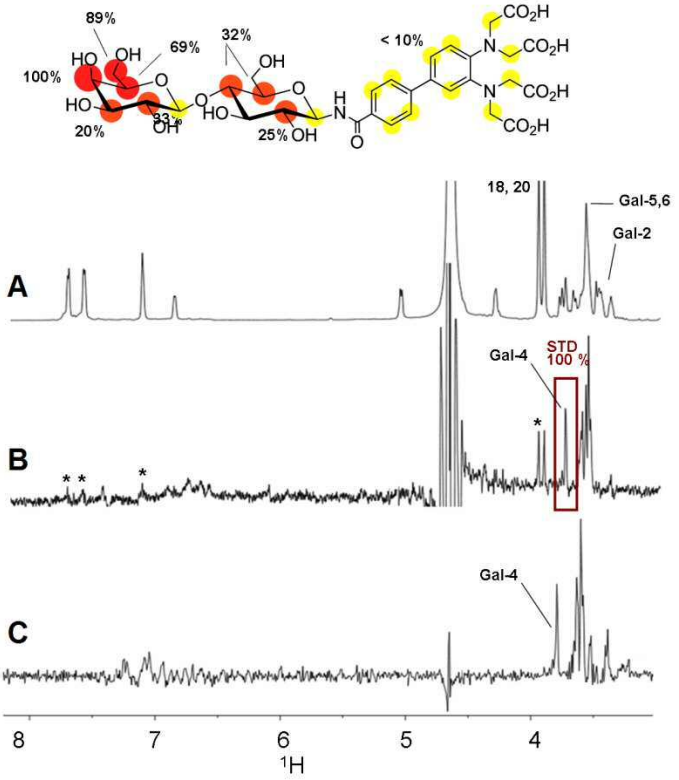


Figure 7: Lower panel: STD experiments performed on either Lac-biPhDTA or lactose in the presence of gal-3 CRD. A: Lac-biPhDTA, with labels (off-resonance); B: Lac-biPhDTA (STD); c: lactose (STD). Upper panel: ligand epitope mapping of the recognition of Lac-biPhDTA by human gal-3 CRD as inferred from STD experiments. Protons from the biPhDTA linker (labelled with asterisks, i.e. biphenyl protons and those of the chelating arms) give rise to much weaker STD responses than those of the lactose moiety, thus indicating that the recognition takes place at the level of the sugar.

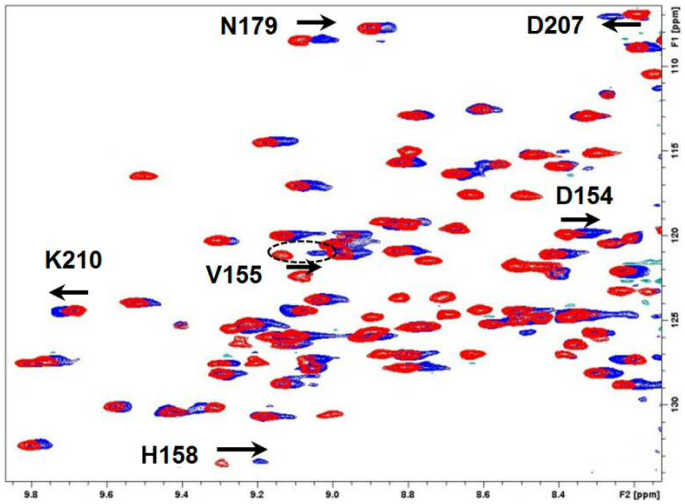


Figure 8: Superimposition of ^{1}H-^{15}N HSQC spectra of gal-3 in complex with 2.5 eq. Lac-biPhDTA acquired under isotropic conditions (red, loaded with La^{3+}) and anisotropic conditions (blue, loaded with Tb^{3+}). Some of the residues directly involved in the lactose-gal-3 interaction or in close proximity have been highlighted.

These results were analyzed on the basis of a three-dimensional model of the complex using the X-ray coordinates of lactose-bound gal-3 (PDB: 2NN8) [37]. This structure was used as a template to generate the putative complex with compound Lac-biPhDTA, by attaching the biPhDTA fragment to the lactose moiety, as described in the experimental section. Plot in fig. 9 shows the gal-3 residues experiencing significant low field and high field PCSs, as well as those broadening beyond detection. The latter always correspond to NH pairs located at closer distance from the metal (around 30.0 Å for both Dy^{3+} and Tb^{3+}). However, in the case of Tm^{3+} no signal was broadened beyond detection. PCSs were measured in most gal-3 NH signals, even in residues further than 40 Å apart from the Ln^{3+} metal core.

Looking at the effects of each individual metal, PCS data obtained for Dy^{3+} and Tb^{3+} showed the same pattern of positive and negative values, as expected from the corresponding isosurface shapes. In addition, the spectrum of the sample loaded with Tm^{3+} (with the lowest PRE of the three paramagnetic metals employed) proved very valuable to extract additional information. In this case, it was possible to observe NH cross-peaks belonging to residues lying 38.6 Å apart from the metal. These signals disappeared when Lac-biPhDTA was loaded with the other metals (fig. 9, e.g. around 163-169). Thus, the combination of data with different metals allows for the gathering of complementary information.

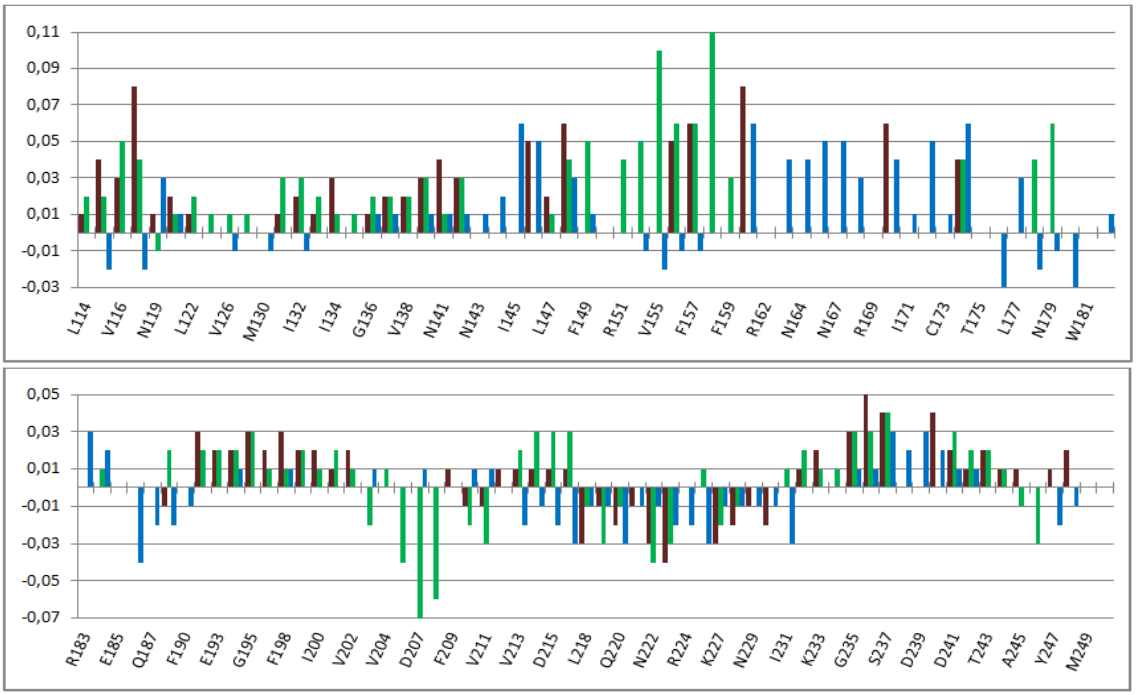


Figure 9: Plot of the positive and negative variations of ΔPCSs along the same complex, loaded with three different lanthanide metals (brown for Dy^{3+}, green for Tb^{3+} and blue for Tm^{3+}), taking as blank the corresponding data obtained with La^{3+}.

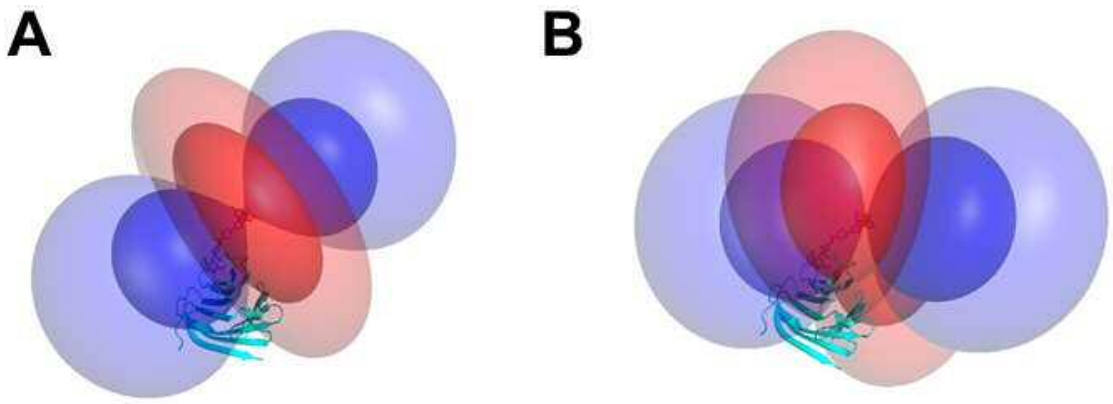


Figure 10: Structures of gal-3 complexed with Lac-biPhDTA, showing PCSs isosurfaces calculated from combined Δχ-tensors of Dy^{3+} (A), and Tm^{3+} (B), obtained through PCS data on the protein signals (figs. 8-9). Figures were built using Numbat [38].

6.2.3 Disulfide-based sugar-lanthanide conjugates

The synthetic route to Lac-biPhDTA took 10 steps in total [39]. Its synthesis, carried out at the laboratory of Prof. Pérez-Castells, was planned in a convergent way, with a final amide ligation between 1-β-aminolactose and the biPhDTA fragment. Although the synthesis is straightforward and provides the final product in good yields, it is out of the synthetic capabilities of most NMR groups.

In order to provide an alternative solution with less synthetic requirements, we conceived a strategy for the synthesis of sugar-paramagnetic derivatives via a disulfide linkage. This procedure allows for the exploitation of a diversity of thiol-targeting tags originally conceived for proteins, which are commercially available. In addition, one can take advantage of the many thio-saccharides available in the market. Altogether, this strategy can be practiced without complex synthetic requirements, since the final disulfide coupling between the thiosugar and the paramagnetic tag occurs instantaneously and stoichiometrically in water.

Coupling of 1-β-thiogalactose (Gal-1*S*) with a biPhDTA tag activated as a methanethiosulfonate proceeded and stoichiometricly in water. Large PCS were detected in the sugar signals (fig. 11). Of note, the non-planarity of the disulfide bond yielded a different PCS pattern from that obtained from the rigid conjugate Lac-biPhDTA. In Gal-1*S*-biPhDTA, H6 protons were the ones displaying the largest PCSs.

Another difference is that two distinct conformers of Gal-1*S*-biPhDTA arise from the two possible C—S—S—C torsion angles about the disulfide bond of +90 and -90 [40] (fig. 12). In our conjugate, PCSs predicted for the +90 conformer in the presence of Tb^{3+} were in better agreement with the experimental data (fig. 12), with Q = 0.122, in contrast with the value of Q = 0.225 obtained for the -90 conformer. This seems to indicate that the +90 conformer is the major species in solution. Nevertheless, Q factors obtained with this molecule were worse than those obtained with the rigid conjugates. This is probably accounted for by the coexistence of these two conformers, and the greater conformational freedom around the disulfide linkage vis-à-vis the amide bond.

Next, we also explored the application of these disulfide-based conjugates for the detection and study of their recognition by macromolecular receptors. Since the minimal recognition unit for galectins is the residue of galactose in its β-configuration, Gal-1(β)*S*-biPhDTA is also a valid tool to probe its binding to galectins through monitoring of paramagnetic effects from the protein perspective.

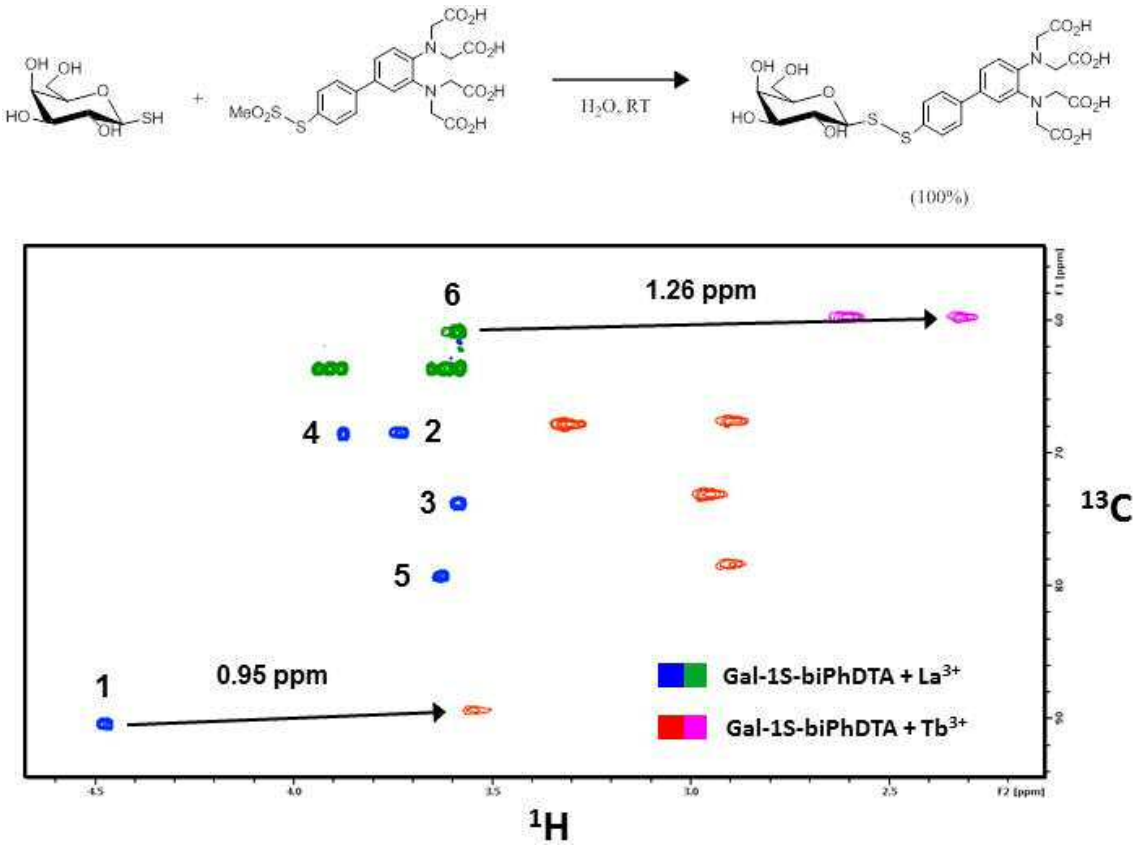


Figure 11: Gal-1*S*-biPhDTA. Upper panel: synthesis. Lower panel: overlaid ^{1}H-^{13}C HSQC spectra of Gal-1*S*-biPhDTA loaded with a diamagnetic lanthanide (La^{3+}, blue) and a paramagnetic lanthanide (Tb^{3+}, red).

^{1}H-^{15}N HSQC spectra of gal-3 recorded in the presence of 1 equivalent Gal-1*S*-biPhDTA loaded with Tb^{3+} showed dramatic effects in terms of PCSs and signal broadening beyond detection (fig. 13). These effects were in fact superior to those exerted by Lac-biPhDTA (fig. 8), if one takes into account the lower excess of ligand used in the case of the galactose derivative and its poorer affinity (in the mM range [41]) compared with the disaccharide. This is likely due to the fact that the paramagnetic center of Gal-1*S*-PhDTA is bent toward the protein surface, owing to the non-planarity of the disulphide bond. These results underline the sensitiveness of disulfide-based paramagnetic ligand derivatives for the detection of protein-binding events.

Of note, the -90 conformer of Gal-1*S*-PhDTA is the only one possible in the galectin-bound state, given the steric clashes that would arise between the +90 conformer and the loop encompassing residues R162-V170. Thus, there is a conformational selection of the minor conformer by protein binding. This fact is supported by the absence of signal loss, due to PREs,

around that region of the protein in the presence of Tb^{3+} and other paramagnetic lanthanides (fig. 13B, 13C).

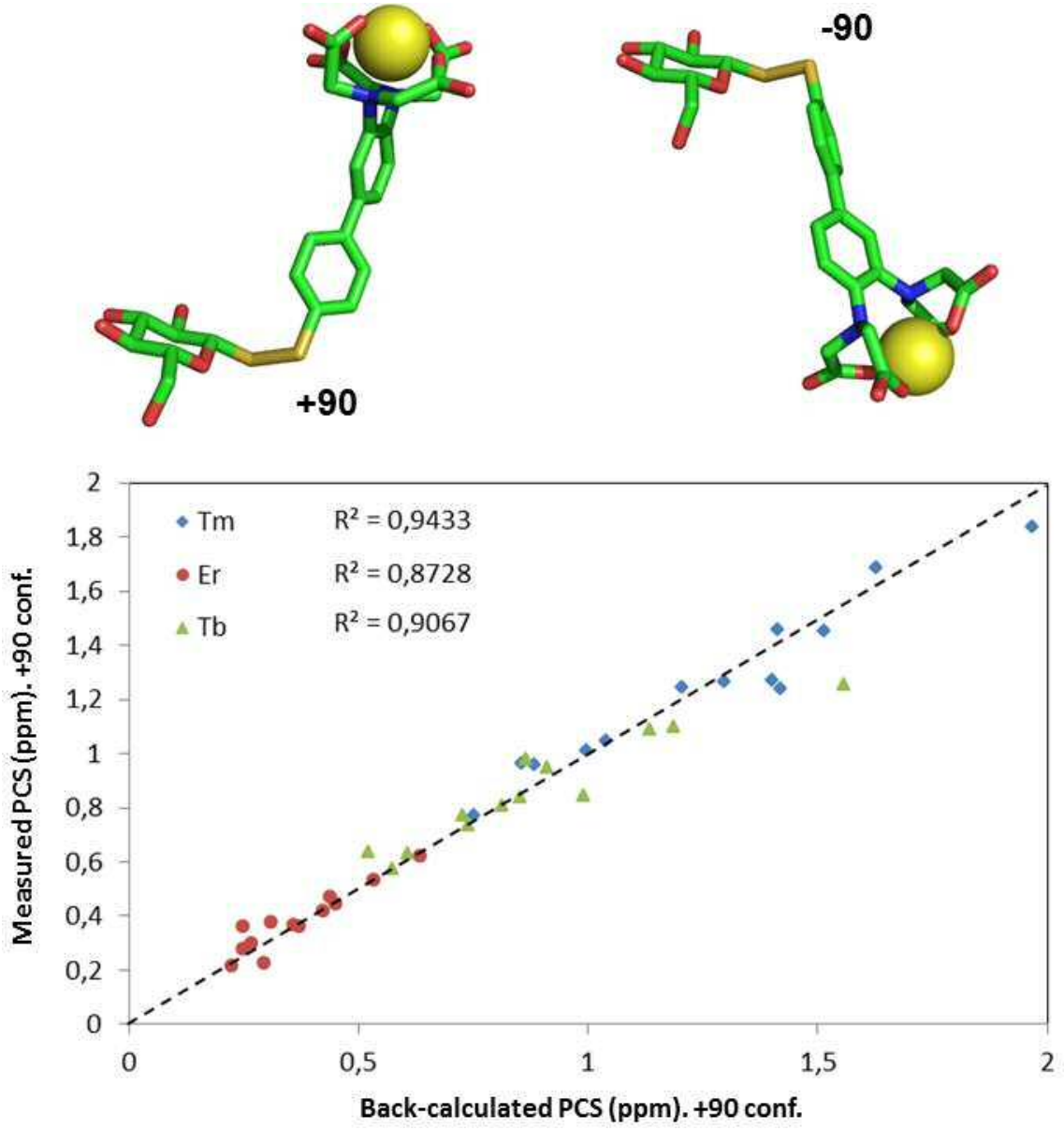


Figure 12: Upper panel: two possible conformers of Gal-1*S*-biPhDTA by rotation of the angle around the disulfide bond. Lower panel: Plots correlating experimentally obtained and back-calculated ^{1}H and ^{13}C PCSs for the geometry of Gal-1*S*-biPhDTA in the +90 conformation, in the presence of either Tm^{3+}, Er^{3+} or Tb^{3+}.

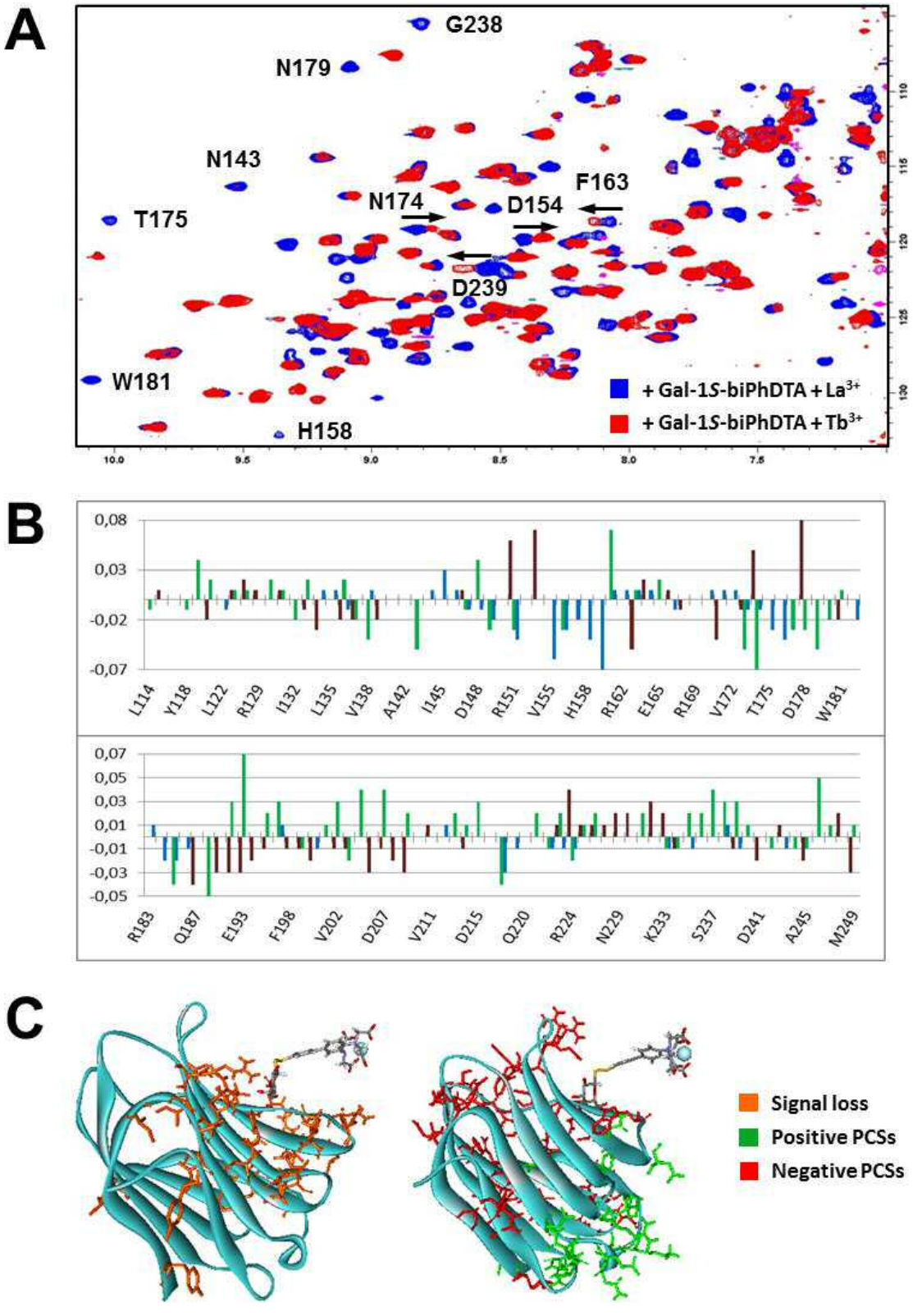


Figure 13: A: ^{1}H-^{15}N HSQC spectra of gal-3 in complex with 1 eq. Gal-1*S*-biPhDTA loaded with La^{3+} (blue) or Tb^{3+} (red). B: Plot showing PCSs for the same complex, when loaded with three different metals (brown: Tb^{3+}, green: Tm^{3+}; blue: Yb^{3+}). C: Model of the gal-3/Gal-1*S*-biPhDTA complex highlighting residues whose cross-peaks show positive (green), negative (red) PCSs or become broadened beyond detection (orange) in the presence of Tb^{3+}.

6.2.4 Fluorine-detected paramagnetism

It has previously been explained (Chapter 1) how the virtues of fluorine-19 have prompted the development of a wide variety of ^{19}F-detected NMR experiments devoted to study ligand-receptor interactions. Among them, we should emphasize the good sensitivity of ^{19}F by NMR and the lack of fluorine in virtually all biological molecules.

Obviously, a requisite is that the studied species contain at least one atom of fluorine. In the case of fluoroproteins, and since natural amino acids involved in protein synthesis of course lack fluorine, two different approaches have been used to circumvent this issue: i) biosynthetic introduction of fluorinated amino acid analogs in the expression systems and ii) post-translational labeling at reactive amino acids with fluorine-containing reagents. The first approach is much more work-consuming, and is susceptible to a number of challenges. Thus, the facile incorporation of fluorine tags via directed labeling is often regarded as more convenient, and allows for the use of a large number of commercially available fluorine-containing probes.

In this section, we show the application of a rationally designed ^{19}F-spin label of proteins to paramagnetic-detection of molecular recognition processes, which can also be employed for a number of other potential applications.

6.2.4.1 Design rationale of a trifluoroacetylcysteine methanethiosulfonate (TFAM) tag

Most protein-targeting tags are directed toward reactive amino acids, such as Cys, Lys, Arg and, to a lesser extent, Ser and Thr. We designed our tag as a methanethiosulfonate for highly reactive targeting of Cys residues, since there are usually only a few reactive Cys residues per molecule, thus rendering simplified ^{19}F-NMR spectra and minimize spectral overlapping in protein mixtures. As it is common among fluorine probes, a trifluoromethyl (CF_3) group was chosen as the fluorine carrier, in order to maximize sensitivity and take advantage of its relatively narrow peaks due to rapid rotation along the methyl axis. The CF_3 group was separated from the rest of the tag by an amide bond, in order to avoid ^{19}F-^{1}H cross-relaxation pathways leading to signal broadening, and to contribute to overall rigidity needed for high sensitivity to paramagnetic effects. Finally, and in order to overcome solubility problems often associated with fluorine-containing probes, the design of the tag was based on a cysteine backbone, in order to attain the polarity needed to work at high concentrations.

The synthesis of TFAM was carried out at the CIB-CSIC, by trifluoroacetylation of L-cysteine methanethiosulfonate with trifluoroacetic anhydride in CF_3COOH (fig. 14).

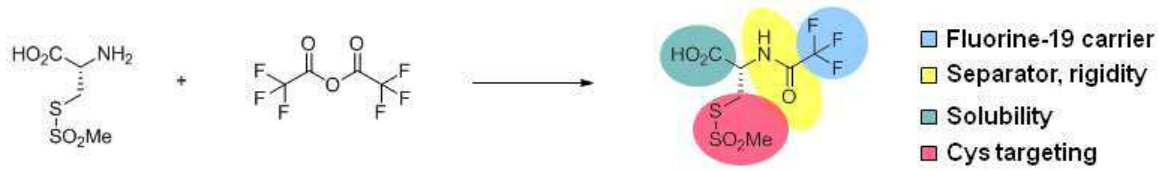


Figure 14: Synthesis and design rationale of the trifluoroacetylcysteine methanethiosulfonate (TFAM) protein-binding tag.

6.2.4.2 Labeling of gal-7 with TFAM

We deliberately chose to test this approach using Lac-PhDTA and gal-7, which has a single Cys residue (Cys38). Although, in all crystallographic structures solved, the C38 sulfhydryl group appears buried by R117, the flexibility of its side-chain allows for the tagging with TFAM. Tagging proceeded instantaneously and stoichiometrically in water. Completeness of the reaction could be monitored by ^{1}H-NMR, by the shifting of the methanethiosulfonate signal from around 3 ppm (when bound to the rest of the molecule) to around 2.1 ppm (when released upon protein tagging) (fig. S10 in Appendix).

Fig. 15 shows the overlay of ^{1}H-^{15}N HSQC spectra of gal-7 before and after tagging with TFAM. Several peaks corresponding to residues located in the vicinity of the modified sulfhydryl, such as C38 itself, G39, Q66, R117 and V118 showed considerable shifting. Covalent attachment of the tag was also confirmed by mass spectrometry (fig. 15B). Lactose binding to the labeled protein was confirmed by ^{1}H-^{15}N HSQC (fig. S11 in Appendix), thus excluding that tagging with TFAM would lead to loss of ligand recognition by gal-7.

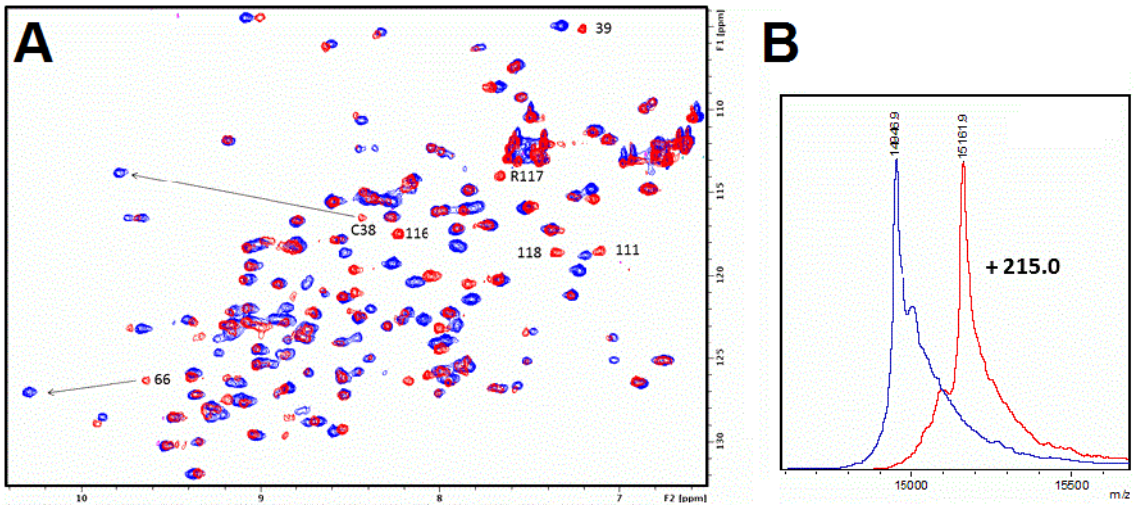


Figure 15: Labeling of gal-7 by TFAM at Cys38, as monitored by ^{1}H-^{15}N HSQC (A) and MALDI (B).

6.2.4.3 Detection of PCS in TFAM-labeled proteins

Addition of four equivalents Lac-PhDTA loaded with Tb^{3+} to a sample of TFAM-labeled gal-7 induced PCS of around 0.04 ppm at the ^{19}F signal of the labeled protein (fig. 16). According to the model shown in fig. 16B, the paramagnetic center lied around 27 Å apart from the fluorine nuclei. The measured PCSs are thus comparable with those obtained in previous studies using ^{1}H-^{15}N HSQC experiments.

Similar results were obtained using a system involving bovine serum albumin (BSA) and a paramagnetic derivative of tryptophan, a natural ligand thereof [42]. BSA is a 67 kDa protein, which has 35 Cys residues in total, of which 34 form intrachain disulfide bonds, and only one (Cys34) is reduced and solvent-accessible [42-43]. In this case, addition of two equivalents L-Trp-PhDTA:Dy^{3+} to TFAM-labeled BSA at C34 produced a ^{19}F-PCS of around -0.034 ppm, with the distance between the paramagnetic center and the fluorine nuclei estimated around 37.6 Å (fig. 17B).

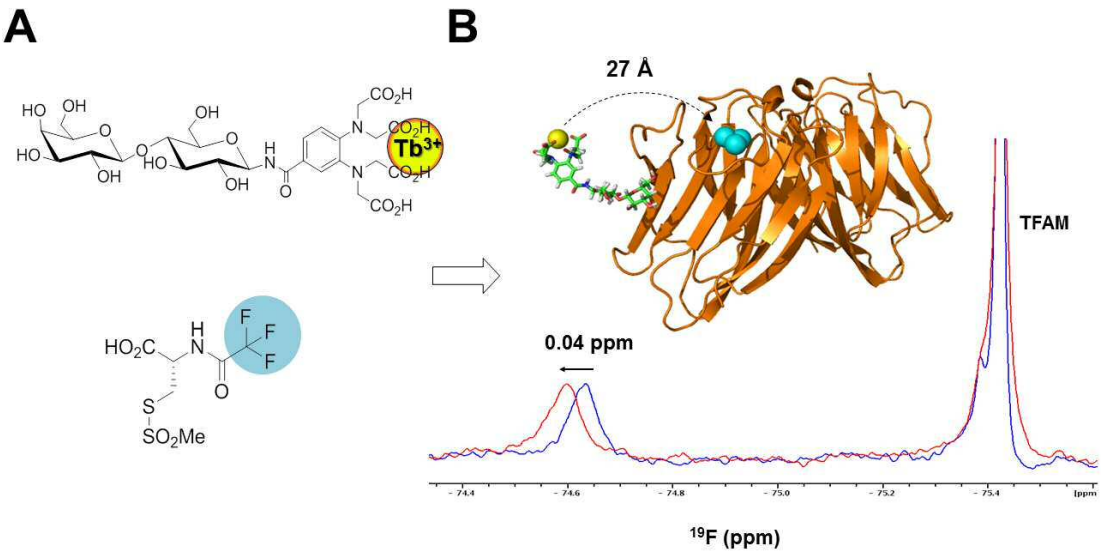


Figure 16: A: structures of Lac-PhDTA and TFAM. B: ^{19}F-NMR spectra of TFAM-labeled gal-7, in the presence of 4 equivalents Lac-PhDTA:La^{3+} (diamagnetic, blue) or Lac-PhDTA:Tb^{3+} (paramagnetic, red). The excess of TFAM is shown as an internal chemical shift reference. On top, a model of TFAM-labeled gal-7 complexed with Lac-PhDTA is shown, highlighting the distance measured between the paramagnetic center and the CF_3 moiety (blue balls).

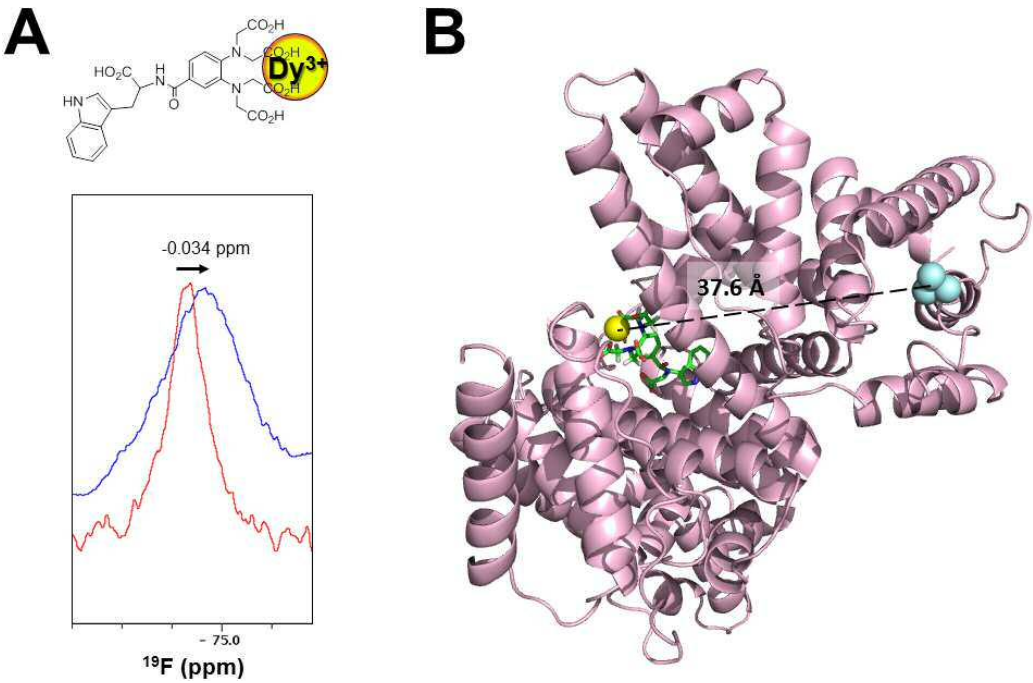


Figure 17: A: structure of Trp-PhDTA and ^{19}F-NMR spectra of TFAM-labeled BSA, in the presence of 2 eq. Trp-PhDTA:La^{3+} (diamagnetic, red) or Trp-PhDTA:Dy^{3+} (paramagnetic, blue). B: Model of TFAM-labeled BSA complexed with L-Trp-PhDTA, highlighting the distance between the paramagnetic center and the CF_3 moiety (blue balls).

In previous sections, we have shown how paramagnetic ligands can be used as tools for the study of structural aspects of their recognition by macromolecular receptors. Using uniformly ^{15}N-labeled galectin, we detected and measured specific perturbation of proton NH signals beyond the binding site to a distance up to 40 Å from the ligand.

Of course, the latter approach is only possible if recombinant production, isotopic labeling and acquisition of high-resolution 2D spectra of the protein of interest can be done. Thus, tagging of the protein by a fluorine-containing probe such as TFAM and monitoring of paramagnetic perturbations, exerted by ligands or other modified proteins, of the fluoroprotein signals in 1D ^{19}F-NMR spectra can be a useful way to deal with those instances where the latter approach is not possible (e.g., proteins isolated from natural sources or of very high molecular weight).

6.3 Methods

Experimental PCSs for lanthanide-chelating carbohydrate derivatives were calculated as the difference between the chemical shift of each signal in an ^{1}H-^{13}C HSQC spectrum acquired in the presence of a diamagnetic metal ion (La^{3+}) and one spectrum acquired in the presence of a paramagnetic metal ion (e.g. Tb^{3+}, Dy^{3+}). Spectra were acquired at 600 MHz and 298 K. Samples were prepared in deuterated Tris buffer (D_2O, Tris-d_6, 50 mM, pH 7.8) at a final concentration of the carbohydrate derivative 5 of 2.5 mM. CH_3CN was added as chemical shift reference.

Once the experimental values were obtained, MSpin [35] software was used to back-calculate the expected PCSs from the different lactose conformations. The quality of the fitting between experimental and back-calculated values is given by the Q factor, which is defined by the following expression:

$$Q = \sqrt{\frac{\sum (PCS_{calc} - PCS_{\exp})^2}{\sum PCS_{\exp}^{\ 2}}} \qquad (7)[35]$$

For molecular recognition applications, gal-3 PCSs were measured from ^{1}H-^{15}N HSQC experiments acquired in the presence of the lanthanide chelating carbohydrate conjugates loaded either with diamagnetic (La^{3+}) or with paramagnetic ions (Tm^{3+}, Tb^{3+}, Dy^{3+}). Spectra

were recorded at 700 MHz and 298 K. For experiments involving disulfide-based conjugates, samples were prepared in sodium acetate buffer (100 mM in H_2O/10% D_2O) at pH 6.5, because the tag activated as a methanethiosulfonate readily decomposed at basic pH. Protein concentration was 270 μM and the lactose derivative/protein molar ratio was set to 2.5:1 with Lac-biPhDTA, and 1:1 with Gal-1*S*-biPhDTA. Gal-1*S* was purchased from Sigma (St. Louis, MO).

For the analysis of protein PCSs, Numbat (New Userfriendly Method Built for Automatic χ-Tensor determination) [38] was used to back-calculate the expected PCSs for the geometry reported in the X-ray crystallographic structure of the galectin-3/lactose complex (PDB: 2NN8) [37]. The lactose molecule of the X-ray structure was replaced by the lactose conjugate bearing the lanthanide-binding tag. The geometry of the PhDTA unit was taken from the reported coordinates of this moiety crystallized in the presence of Fe(III) [44].

Synthesis of TFAM was done by trifluoroacetylation of L-cysteine methanethiosulfonate with trifluoroacetic anhydride in CF_3COOH. The reaction was carried out at 278 K and real-time monitored by ^{1}H-NMR and ^{19}F-NMR until completion. The product, obtained in excellent yields (98%) was isolated by coevaporation of the solvent with toluene under vacuum. Proteins were labeled with TFAM in water, using 1:1 proportions. PCSs were measured at the ^{19}F-NMR signals upon addition of paramagnetic ligands at different proportions, using a 500 MHz spectrometer equipped with a ^{19}F probe.

6.4 References

[1] Martin-Pastor, M., and Bush, C. A. (2001), *J Biomol NMR* **19**, 125-139.
[2] Azurmendi, H. F., Martin-Pastor, M., and Bush, C. A. (2002), *Biopolymers* **63**, 89-98.
[3] Martin-Pastor, M., and Bush, C. A. (2000), *Carbohydr Res* **323**, 147-155.
[4] Tian, F., Al-Hashimi, H. M., Craighead, J. L., and Prestegard, J. H. (2001), *J Am Chem Soc* **123**, 485-492.
[5] Otting, G. (2010), *Annu Rev Biophys* **39**, 387-405.
[6] Su, X. C., and Otting, G. (2010), *J Biomol NMR* **46**, 101-112.
[7] Alsaadi, B. M., Rossotti, F. J. C., and Williams, R. J. P. (1980), *Journal of the Chemical Society, Dalton Transactions*, 2147-2150.
[8] Clore, G. M., and Iwahara, J. (2009), *Chem Rev* **109**, 4108-4139.
[9] Jain, N. U., Venot, A., Umemoto, K., Leffler, H., and Prestegard, J. H. (2001), *Protein Sci* **10**, 2393-2400.
[10] Machonkin, T. E., Westler, W. M., and Markley, J. L. (2005), *Inorg Chem* **44**, 779-797.

[11] Machonkin, T. E., Westler, W. M., and Markley, J. L. (2004), *J Am Chem Soc* **126**, 5413-5426.
[12] Banci, L., Bertini, I., Cavallaro, G., and Luchinat, C. (2002), *J Biol Inorg Chem* **7**, 416-426.
[13] Bertini, I., Luchinat, C., and Parigi, G. (2002), *Progress in Nuclear Magnetic Resonance Spectroscopy* **40**, 249-273.
[14] McConnell, H. M., and Robertson, R. E. (1958), *The Journal of Chemical Physics* **29**, 1361-1365.
[15] Barry, C. D., North, A. C., Glasel, J. A., Williams, R. J., and Xavier, A. V. (1971), *Nature* **232**, 236-245.
[16] Pintacuda, G., Keniry, M. A., Huber, T., Park, A. Y., Dixon, N. E., and Otting, G. (2004), *J Am Chem Soc* **126**, 2963-2970.
[17] Thiele, C. M. (2007), *Concepts in Magnetic Resonance Part A* **30A**, 65-80.
[18] Annila, A., and Permi, P. (2004), *Concepts in Magnetic Resonance Part A* **23A**, 22-37.
[19] Martin-Pastor, M., Canales-Mayordomo, A., and Jimenez-Barbero, J. (2003), *J Biomol NMR* **26**, 345-353.
[20] Pham, T. N., Hinchley, S. L., Rankin, D. W., Liptaj, T., and Uhrin, D. (2004), *J Am Chem Soc* **126**, 13100-13110.
[21] Erdelyi, M., d'Auvergne, E., Navarro-Vazquez, A., Leonov, A., and Griesinger, C. (2011), *Chemistry* **17**, 9368-9376.
[22] Valafar, H., and Prestegard, J. H. (2004), *J Magn Reson* **167**, 228-241.
[23] Martin-Pastor, M., Canales, A., Corzana, F., Asensio, J. L., and Jimenez-Barbero, J. (2005), *J Am Chem Soc* **127**, 3589-3595.
[24] Zhuang, T., Leffler, H., and Prestegard, J. H. (2006), *Protein Sci* **15**, 1780-1790.
[25] Zhuang, T., Lee, H. S., Imperiali, B., and Prestegard, J. H. (2008), *Protein Sci* **17**, 1220-1231.
[26] Yamamoto, S., Yamaguchi, T., Erdelyi, M., Griesinger, C., and Kato, K. (2011), *Chemistry* **17**, 9280-9282.
[27] Pidcock, E., and Moore, G. R. (2001), *J Biol Inorg Chem* **6**, 479-489.
[28] Allegrozzi, M., Bertini, I., Janik, M. B. L., Lee, Y.-M., Liu, G., and Luchinat, C. (2000), *Journal of the American Chemical Society* **122**, 4154-4161.
[29] Brautigam, C. A., Aschheim, K., and Steitz, T. A. (1999), *Chem Biol* **6**, 901-908.
[30] Frey, M. W., Frey, S. T., Horrocks, W. D., Jr., Kaboord, B. F., and Benkovic, S. J. (1996), *Chem Biol* **3**, 393-403.
[31] Otting, G. (2008), *J Biomol NMR* **42**, 1-9.
[32] Mallagaray, A., Canales, A., Dominguez, G., Jimenez-Barbero, J., and Perez-Castells, J. (2011), *Chem Commun (Camb)* **47**, 7179-7181.
[33] Asensio, J. L., and Jimenez-Barbero, J. (1995), *Biopolymers* **35**, 55-73.
[34] Asensio, J. L., Martin-Pastor, M., and Jimenez-Barbero, J. (1995), *Int J Biol Macromol* **17**, 137-148.
[35] Navarro-Vazquez, A. (2012), *Magn Reson Chem* **50 Suppl 1**, S73-79.
[36] Canales, A., Mallagaray, A., Perez-Castells, J., Boos, I., Unverzagt, C., Andre, S., Gabius, H. J., Canada, F. J., and Jimenez-Barbero, J. (2013), *Angew Chem Int Ed Engl* **52**, 13789-13793.
[37] Collins, P. M., Hidari, K. I., and Blanchard, H. (2007), *Acta Crystallogr D Biol Crystallogr* **63**, 415-419.
[38] Schmitz, C., Stanton-Cook, M. J., Su, X. C., Otting, G., and Huber, T. (2008), *J Biomol NMR* **41**, 179-189.
[39] Canales, A., Mallagaray, A., Berbis, M. A., Navarro-Vazquez, A., Dominguez, G., Canada, F. J., Andre, S., Gabius, H. J., Perez-Castells, J., and Jimenez-Barbero, J. (2014), *J Am Chem Soc* **136**, 8011-8017.
[40] Richardson, J. S. (1981), *Adv Protein Chem* **34**, 167-339.

[41] Saraboji, K., Hakansson, M., Genheden, S., Diehl, C., Qvist, J., Weininger, U., Nilsson, U. J., Leffler, H., Ryde, U., Akke, M., and Logan, D. T. (2012), *Biochemistry* **51**, 296-306.
[42] Friedrichs, B. (1997), *Food / Nahrung* **41**, 382-382.
[43] Simons, S. S., Jr. (1987), *J Biol Chem* **262**, 9669-9675.
[44] Mizuno, M., Funahashi, S., Nakasuka, N., and Tanaka, M. (1991), *Inorganic Chemistry* **30**, 1550-1553.

CONCLUSIONS

Carbohydrate-protein systems of biological relevance have been studied using nuclear magnetic resonance (NMR) spectroscopy.

- The interacting capabilities of a set of viral protein fibers toward small oligosaccharides have been studied, at atomic scale, by using STD experiments.

- The interaction features of a monoclonal antibody toward glycogen have been characterized, from the ligand's perspective, using STD and STDD experiments glycogen and a few related oligosaccharides. Our results point towards a recognition of the inner, α(1-4)-linked Glc units by the antibody, rather than the branching points or the reducing ends.

- On-cell NMR experiments were performed on both a wild-type and a *glgB::Tn* mutant strain of *Mycobacterium smegmatis*. Results indicate that the α-glucan of the *glgB::Tn* mutant is primarily formed by a linear, unbranched α(1-4)glucose polysaccharide, lacking α(1-6) branches. We hypothesize that binding of the mAb to *glgB::Tn* mutants is precluded due to strongly reduced availability of the α(1-4) linked α-glucan epitope caused by insolubility and aggregation of the unbranched α-glucan polymers.

- Using STD experiments, we have gained insight into the catalytic mechanism of the UDP-glucose pyrophosphorylase (UGP) of *Streptococcus pneumoniae*. Our results point toward a sequential ordered Bi-Bi catalytic mechanism, in which UTP binds to the enzyme first, followed by Glc-1-P binding to the enzyme/nucleotide complex.

- An NMR-based assay has been developed to measure UGP activity. This assay represents a direct approach to monitor their activity, because substrate interconversion is directly observed. Thus, it is advantageous over the currently most used one for measuring nucleotidyltransferase activity, which relies on its coupling to NADH/$NAHD^+$ conversion by UDP-Glc 6-dehydrogenase.

- Targeting of the UGP of *S. pneumoniae* trough fragment-based drug discovery has provided a set of structurally similar small molecules binding with mM affinities to the enzyme. These studies represent the first step toward the targeting of pneumococcal UGP, a therapeutic target candidate.

- Full NMR assignment of galectin-7 (gal-7) has been accomplished, which revealed the coexistence of two conformations of gal-7. The origin of this conformational duality was found to be the *cis-trans* isomerization equilibrium of the Val3-Pro4 peptide bond.

- A fragment-based screening was done targeting gal-7. A set of fragment hits were found to bind with moderate affinities to two distinct epitopes of gal-7 involved, or presumably involved in protein-protein interactions, rather than the sugar recognition site: a previously unreported binding site targeted by phenoxyphenyl-derivatives, and the gal-7 dimerization interface. Fragment binding to the latter site was shown to disrupt self-association of gal-7 monomers.

- Allosteric competition between lactose and fragments binding to the gal-7 dimer interface has shown that sugar binding to gal-7 results in stabilization of the gal-7 dimer state.

- Complete NMR assignment of full-length galectin-3 (gal-3) has been done.

- The interaction features of a series of glycan ligands by the gal-3 carbohydrate recognition domain (CRD) have been studied by both ligand- and protein-detected NMR methods, in combination with molecular modeling.

- The Interaction of the gal-3 CRD with synthetic (phospho)peptides derived from the gal-3 N-terminal tail was assessed. Results show that distinct N-terminal peptides can interact at specific sites within the gal-3 CRD, suggesting that a similar behavior may take place in the full-length protein. Specifically, phosphorylation of Ser6 and Ser12 in specific N-terminal peptides was shown to promote association with the CRD, while phosphorylation of Tyr107 and Tyr118 was shown to attenuate this association.

- A novel strategy has been developed for the application of paramagnetic sugar conjugates to the study of carbohydrate conformation and recognition features with receptors. As a proof-of-principle example, the conformation of lactose has been assessed using pseudo-contact shift (PCS) information. In addition, these conjugates have been shown to be useful for the detection and characterization of recognition events by receptors.

- A method for facile synthesis of disulfide-based paramagnetic sugar derivatives has been proposed, also with applications to the study of molecular recognition processes.

- A fluorine-spin label reagent targeted to protein thiols has been synthesized. The application of this reagent to paramagnetism-based detection of ligand binding has been studied, using paramagnetic ligand derivatives.

CONCLUSIONS

Carbohydrate-protein systems of biological relevance have been studied using nuclear magnetic resonance (NMR) spectroscopy.

- The interacting capabilities of a set of viral protein fibers toward small oligosaccharides have been studied, at atomic scale, by using STD experiments.

- The interaction features of a monoclonal antibody toward glycogen have been characterized, from the ligand's perspective, using STD and STDD experiments glycogen and a few related oligosaccharides. Our results point towards a recognition of the inner, α(1-4)-linked Glc units by the antibody, rather than the branching points or the reducing ends.

- On-cell NMR experiments were performed on both a wild-type and a *glgB::Tn* mutant strain of *Mycobacterium smegmatis*. Results indicate that the α-glucan of the *glgB::Tn* mutant is primarily formed by a linear, unbranched α(1-4)glucose polysaccharide, lacking α(1-6) branches. We hypothesize that binding of the mAb to *glgB::Tn* mutants is precluded due to strongly reduced availability of the α(1-4) linked α-glucan epitope caused by insolubility and aggregation of the unbranched α-glucan polymers.

- Using STD experiments, we have gained insight into the catalytic mechanism of the UDP-glucose pyrophosphorylase (UGP) of *Streptococcus pneumoniae*. Our results point toward a sequential ordered Bi-Bi catalytic mechanism, in which UTP binds to the enzyme first, followed by Glc-1-P binding to the enzyme/nucleotide complex.

- An NMR-based assay has been developed to measure UGP activity. This assay represents a direct approach to monitor their activity, because substrate interconversion is directly observed. Thus, it is advantageous over the currently most used one for measuring nucleotidyltransferase activity, which relies on its coupling to NADH/NAD^+ conversion by UDP-Glc 6-dehydrogenase.

- Targeting of the UGP of *S. pneumoniae* trough fragment-based drug discovery has provided a set of structurally similar small molecules binding with mM affinities to the enzyme. These studies represent the first step toward the targeting of pneumococcal UGP, a therapeutic target candidate.

- Full NMR assignment of galectin-7 (gal-7) has been accomplished, which revealed the coexistence of two conformations of gal-7. The origin of this conformational duality was found to be the *cis-trans* isomerization equilibrium of the Val3-Pro4 peptide bond.

- A fragment-based screening was done targeting gal-7. A set of fragment hits were found to bind with moderate affinities to two distinct epitopes of gal-7 involved, or presumably involved in protein-protein interactions, rather than the sugar recognition site: a previously unreported binding site targeted by phenoxyphenyl-derivatives, and the gal-7 dimerization interface. Fragment binding to the latter site was shown to disrupt self-association of gal-7 monomers.

- Allosteric competition between lactose and fragments binding to the gal-7 dimer interface has shown that sugar binding to gal-7 results in stabilization of the gal-7 dimer state.

- Complete NMR assignment of full-length galectin-3 (gal-3) has been done.

- The interaction features of a series of glycan ligands by the gal-3 carbohydrate recognition domain (CRD) have been studied by both ligand- and protein-detected NMR methods, in combination with molecular modeling.

- The Interaction of the gal-3 CRD with synthetic (phospho)peptides derived from the gal-3 N-terminal tail was assessed. Results show that distinct N-terminal peptides can interact at specific sites within the gal-3 CRD, suggesting that a similar behavior may take place in the full-length protein. Specifically, phosphorylation of Ser6 and Ser12 in specific N-terminal peptides was shown to promote association with the CRD, while phosphorylation of Tyr107 and Tyr118 was shown to attenuate this association.

- A novel strategy has been developed for the application of paramagnetic sugar conjugates to the study of carbohydrate conformation and recognition features with receptors. As a proof-of-principle example, the conformation of lactose has been assessed using pseudo-contact shift (PCS) information. In addition, these conjugates have been shown to be useful for the detection and characterization of recognition events by receptors.

- A method for facile synthesis of disulfide-based paramagnetic sugar derivatives has been proposed, also with applications to the study of molecular recognition processes.

- A fluorine-spin label reagent targeted to protein thiols has been synthesized. The application of this reagent to paramagnetism-based detection of ligand binding has been studied, using paramagnetic ligand derivatives.

CONCLUSIONS

Se ha estudiado una serie de sistemas carbohidrato-proteína de interés biológico mediante espectroscopía por resonancia magnética nuclear (RMN).

- Se ha estudiado, a escala atómica, las capacidades de interacción de un conjunto de proteínas fibrilares virales con oligosacáridos pequeños.

- Se ha estudiado la interacción de un anticuerpo monoclonal con glucógeno y una serie de oligosacáridos relacionados, desde el punto de vista del ligando, mediante experimentos STD y STDD. Los resultados apuntan a un reconocimiento preferente de las unidades internas de glucosa en cadenas lineales α(1-4)-enlazadas, frente a las ramificaciones α(1-6) o los extremos reductores.

- Se han llevado a cabo experimentos de RMN en células de *Mycobacterium smegmatis* procedentes de cepas silvestres y mutantes *glgB::Tn*. Los resultados indican que el α-glucano del mutante *glgB::Tn* está formado principalmente por un polisacárido de glucosa lineal unido mediante enlaces α(1-4), carente de ramificaciones α(1-6). Se ha hipotetizado que el reconocimiento del α-glucano del mutante *glgB::Tn* por parte del anticuerpo monoclonal se ve impedido por la insolubilidad y agregación de los polímeros lineales de α(1-4)-glucosa.

- Utilizando experimentos STD, se ha obtenido información sobre el mecanismo catalítico de la UDP-glucosa pirofosforilasa (UGP) de *Streptococcus pneumoniae*. Los resultados apuntan a un mecanismo catalítico secuencial ordenado Bi-Bi, en el que el UTP se une a la enzima primero, seguido de la unión de Glc-1-P al complejo enzima/nucleótido.

- Se ha desarrollado un ensayo basado en RMN para medir la actividad UGP. Este ensayo representa una estrategia directa para el seguimiento de la actividad UGP, y es ventajoso sobre el método más utilizado actualmente, basado en la medición indirecta de la actividad UGP mediante su acoplamiento con la oxidación de la UDP-glucosa seguida por procedimientos espectrofotométricos.

- Se ha llevado a cabo un cribado de fragmentos contra spUGP, encontrándose una serie de compuestos estructuralmente relacionados que se unían con moderada afinidad al subsitio de unión de UTP de spUGP.

- Se ha llevado a cabo la asignación completa de las señales de RMN de galectina-7 (gal-7). Dicha asignación reveló la coexistencia de dos conformaciones de gal-7. Se ha determinado que el origen de esta dualidad conformacional reside en la isomerización *cis-trans* del enlace peptídico Val3-Pro4.

- Se ha llevado a cabo un cribado de fragmentos contra gal-7, hallándose una serie de fragmentos que se unen con moderadas afinidades a dos epítopos distintos de gal-7 involucrados, o presumiblemente involucrados en interacciones proteína-proteína, y distintos del sitio de unión a carbohidratos: un sitio de unión previamente desconocido hacia el que tienen afinidad derivados de fenoxifenilo, y la interfaz de dimerización de gal-7. La unión de fragmentos a dicha interfaz es capaz de competir con la autoasociación de los monómeros de gal-7.

- La competición alostérica entre lactosa y fragmentos con afinidad por la interfaz de dimerización de gal-7 ha puesto de manifiesto que la unión de azúcar a gal-7 resulta en la estabilización del estado dimérico de gal-7.

- Se ha llevado a cabo la asignación completa de las señales de RMN de galectina-3 (gal-3).

- Se ha estudiado la interacción de una serie de ligandos sacarídicos con el dominio de reconocimiento de carbohidratos (CRD) de gal-3, mediante experimentos de RMN basados en la observación de ligando y métodos basados en la observación de las señales de la proteína, en combinación con modelado molecular.

- Se ha estudiado la interacción del CRD de gal-3 con (fosfo)péptidos sintéticos derivados del dominio N-terminal de gal-3. Los resultados muestran la capacidad de distintos péptidos de interaccionar con el CRD de gal-3, lo que sugiere que en la proteína nativa podrían darse análogos contactos de manera intramolecular. La fosforilación de Ser6 y Ser12 en péptidos N-terminales mostró ser esencial para la interacción con el CRD, mientras que la fosforilación de Tyr107 y Tyr118 reveló atenuar dicha asociación.

- Se ha desarrollado una nueva estrategia basada en la aplicación de conjugados paramagnéticos sacarídicos para estudiar la conformación de los carbohidratos y su interacción con receptores. Como prueba de concepto, se ha comprobado la conformación de lactosa mediante información procedente de desplazamientos de pseudo-contacto (PCSs). Además, estos conjugados han revelado ser útiles para la detección y caracterización de su reconocimiento por receptores.

- Se ha propuesto un método para la síntesis sencilla de derivados sacarídicos paramagnéticos basados en enlaces disulfuro. Dichos derivados también han mostrado utilidad para el estudio de procesos de reconocimiento molecular.

- Se ha sintetizado un reactivo de tioles etiquetado con flúor-19. Se ha estudiado la aplicación de este reactivo a la detección de eventos de reconocimiento basada en efectos paramagnéticos, utilizando derivados paramagnéticos de distintos ligandos.

APPENDIX

SUPPLEMENTARY FIGURES

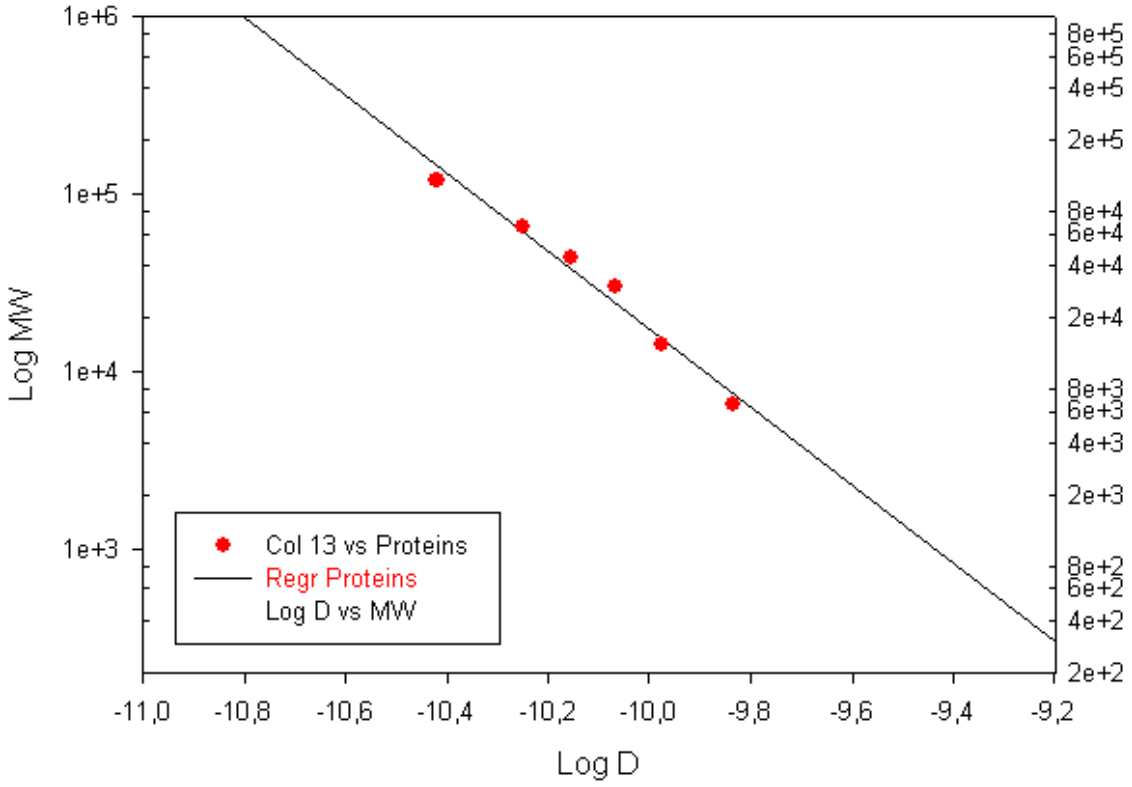


Figure S1: DOSY calibration line [1]built with 6 proteins with molecular weights ranging from 6.5 kDa to 120 kDa. Patterns were: aprotinin (6.5 kDa), α-lactalbumin (14.2 kDa), carbonic anhydrase (30 kDa), ovoalbumin (44.3 kDa), bovine albumin (66 kDa) and viscumin (120 kDa).

Calculated chemical shifts

α-d-Glc^{xii}(1→4)α-d-Glc^{xi}(1→4)α-d-Glc^{x}(1→4)α-d-Glc^{ix}(1→4)[α-d-Glc^{viii}(1→4)α-d-Glc^{vii}(1→4)α-d-Glc^{vi}(1→6)]α-d-Glc^{v}(1→4)α-d-Glc^{iv}(1→4)α-d-Glc^{iii}(1→4)α-d-Glc^{ii}(1→4)α-d-Glc^{i}

→4)α-d-Glc^{i}	1	2	3	4	5	6	6
Expected Calc. Error: 0.17	92.79	72.26	74.03	78.47	70.97	61.76	
	5.24	3.59	3.97	3.63	3.95	3.80	3.87
→4)α-d-Glc^{ii}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
→4)α-d-Glc^{iii}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
→4)α-d-Glc^{iv}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
→4,6)α-d-Glc^{v}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.50	100.50	72.47	74.35	78.19	70.71	66.76	
	5.37	3.66	3.95	3.74	4.01	3.80	4.02
→4)α-d-Glc^{vi}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	98.65	72.22	74.36	78.43	71.37	61.57	
	4.97	3.62	3.99	3.65	3.84	3.81	3.89
→4)α-d-Glc^{vii}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
α-d-Glc^{viii}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.17	100.65	72.72	73.89	70.44	73.60	61.59	
	5.35	3.59	3.70	3.43	3.73	3.77	3.86
→4)α-d-Glc^{ix}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
→4)α-d-Glc^{x}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
→4)α-d-Glc^{xi}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.33	100.45	72.51	74.14	78.20	72.20	61.51	
	5.36	3.64	3.95	3.64	3.84	3.81	3.89
α-d-Glc^{xii}(1→	1	2	3	4	5	6	6
Expected Calc. Error: 0.17	100.65	72.72	73.89	70.44	73.60	61.59	
	5.35	3.59	3.70	3.43	3.73	3.77	3.86

Figure S2: Chemical shift prediction of a glycogen-related oligosaccharide model (sequence shown on top) using CASPER [2].

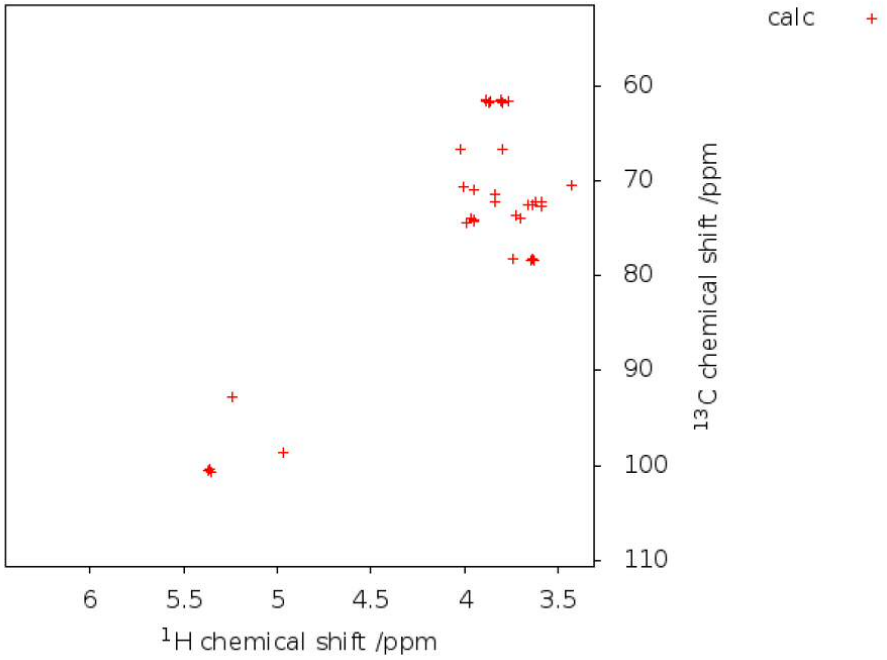


Figure S3: ^{1}H-^{13}C chemical shift prediction for an α(1-6)-branched, α(1-4) linear oligosaccharide using CASPER [2]. Plot built with the predicted chemical shifts shown in fig. S2, akin to a simulated ^{1}H-^{13}C HSQC spectrum.

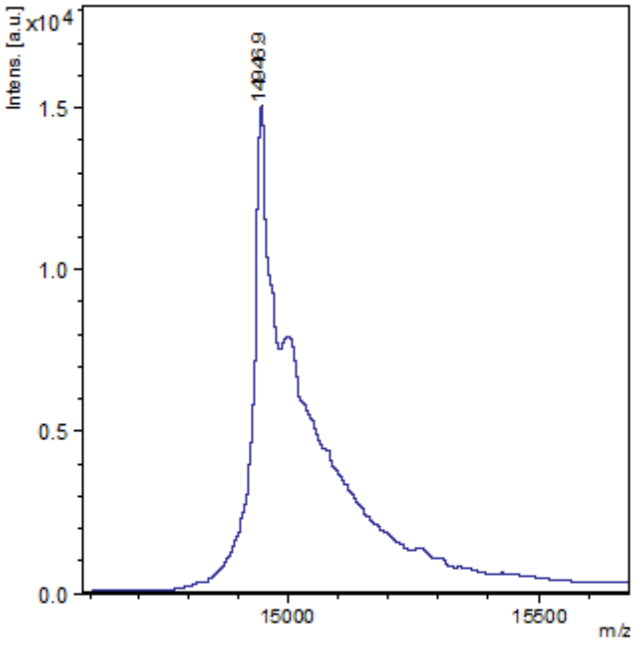


Figure S4: MALDI spectrum of gal-7, showing a single major species corresponding to a molar mass of 14946.9 Da, in good agreement with the predicted molar mass of the native sequence without the initial methionine (14943.8 Da).

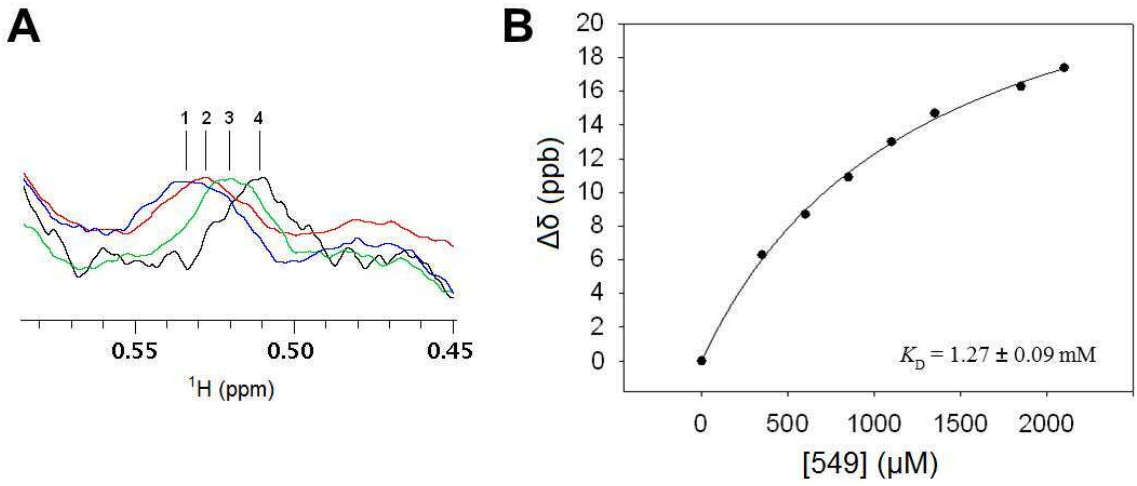


Figure S5: ^{1}H-NMR titration of phenoxyphenyl-based fragments with gal-7. Strong shielding of the gal-7 L24 δ-protons was observed upon addition of these compounds. Being this signal reasonably isolated at around 0.53 ppm in the gal-7 ^{1}H-NMR spectrum [3], a K_D for their binding to gal-7 could be determined by monitoring its chemical shift displacement at increasing concentrations of the titrated compounds (A), being compound 549 the most potent of the series, with a K_D of 1.27 ± 0.09 mM (B).

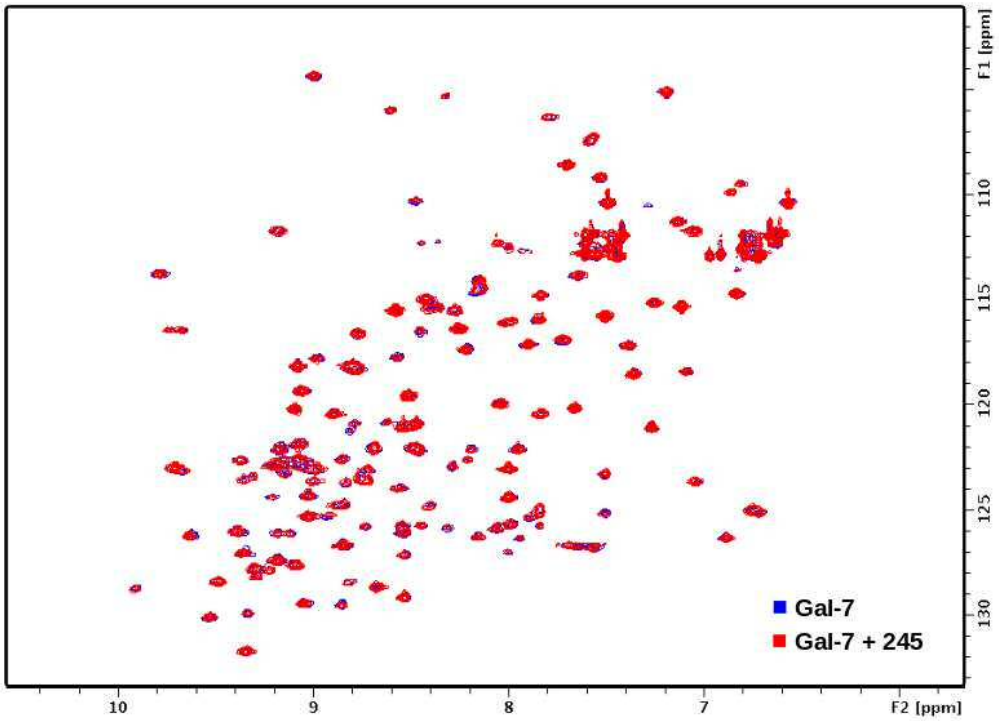


Figure S6: Superimposition of ^{1}H-^{15}N HSQC spectra of 250 µM Gal-7, in the absence (blue) and in the presence of fragment 245 (red).

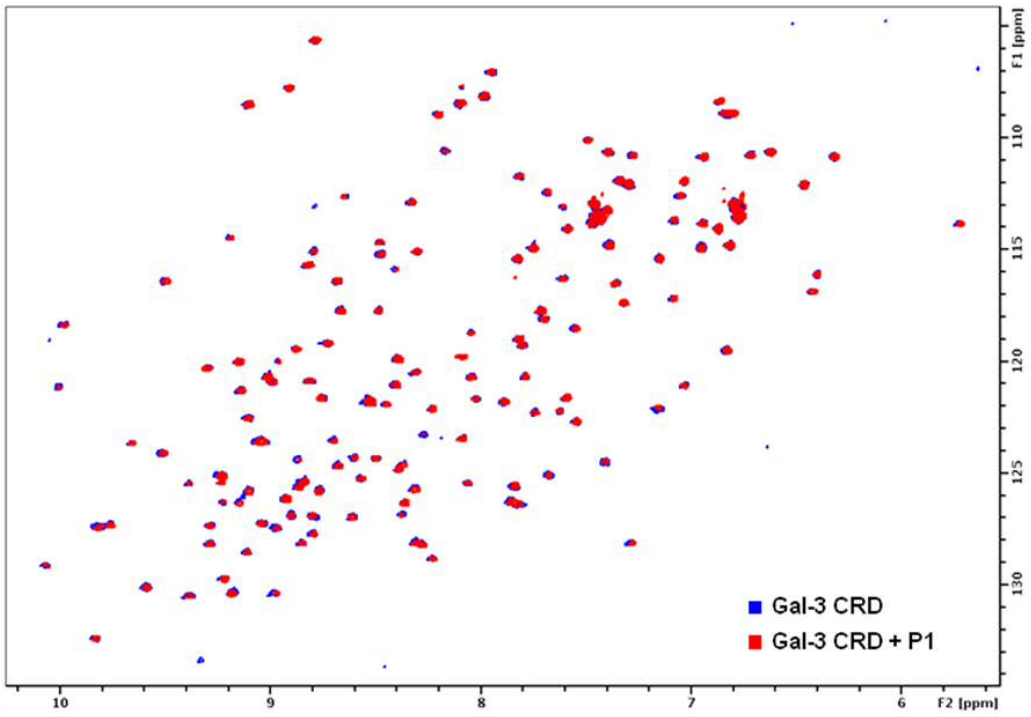


Figure S7: Superimposition of ^{1}H-^{15}N HSQC spectra of the Gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of P1 (red).

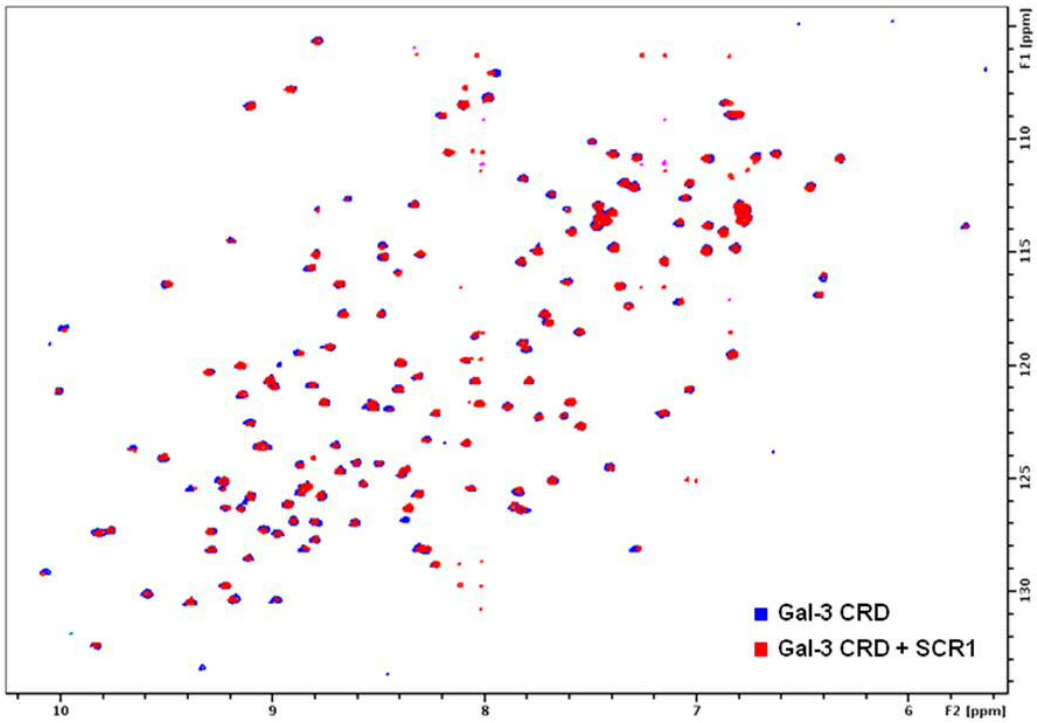


Figure S8: Superimposition of ^{1}H-^{15}N HSQC spectra of the Gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of SCR1 (red).

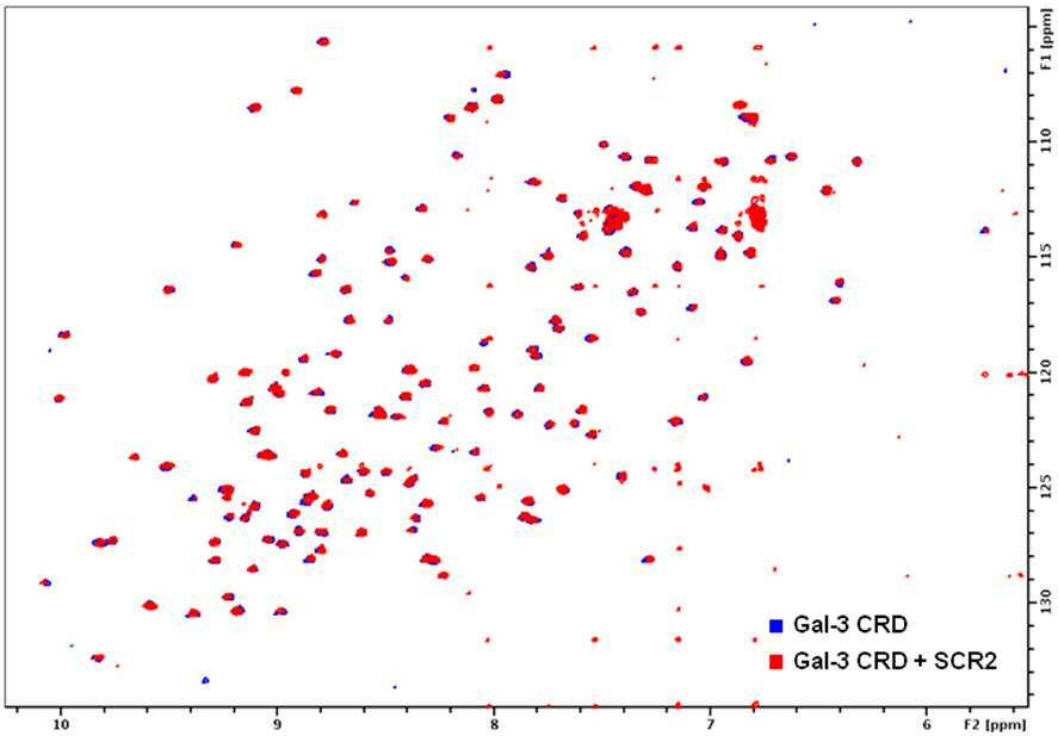


Figure S9: Superimposition of ^{1}H-^{15}N HSQC spectra of the Gal-3 CRD, in the absence (blue) and in the presence of ten equivalents of SCR2 (red).

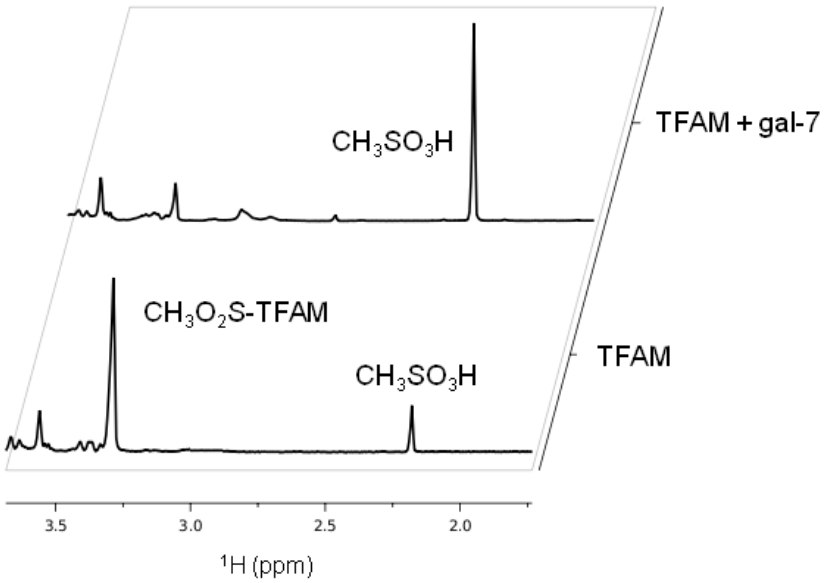


Figure S10: ^{1}H-NMR spectra of TFAM, in the absence (lower spectrum) and in the presence of 1 eq gal-7 (upper spectrum), showing the release of CH_3SO_3H from the tag upon disulfide-bond formation with the gal-7 Cys38.

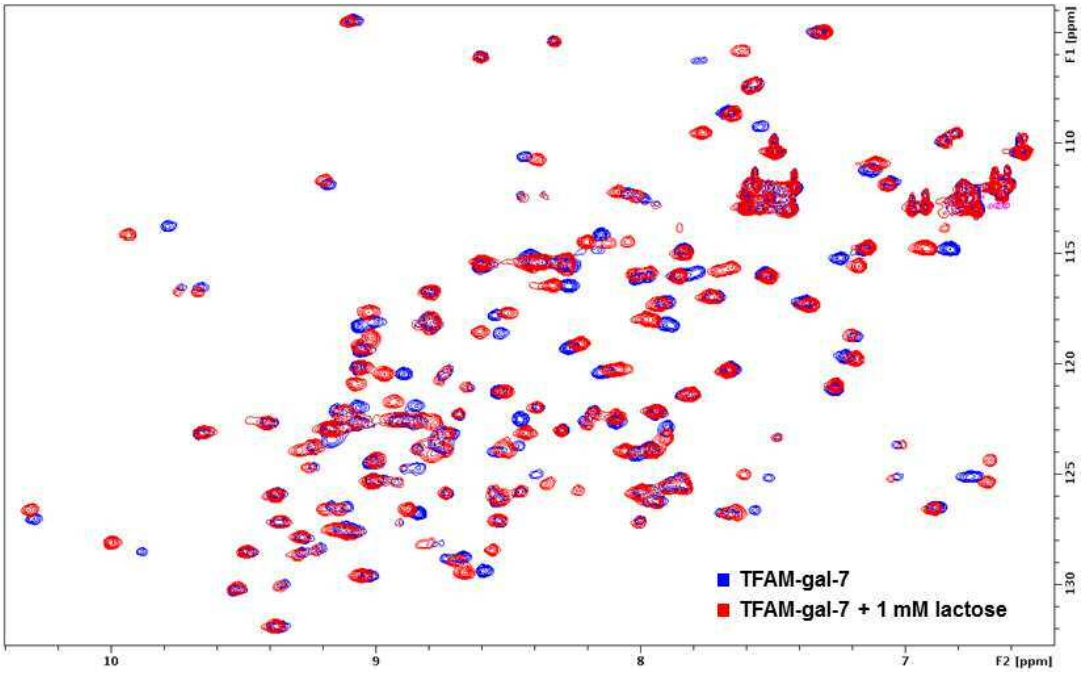


Figure S11: ^{1}H-^{15}N HSQC of TFAM-labeled gal-7, in the absence (blue) and in the presence of 1 mM lactose, showing selective cross-peak perturbation upon addition of lactose to the protein after TFAM labeling.

References

[1] Groves, P., Palczewska, M., Molero, M. D., Batta, G., Canada, F. J., and Jimenez-Barbero, J. (2004), *Anal Biochem* **331**, 395-397.

[2] Jansson, P. E., Stenutz, R., and Widmalm, G. (2006), *Carbohydr Res* **341**, 1003-1010.

[3] Nesmelova, I. V., Berbis, M. A., Miller, M. C., Canada, F. J., Andre, S., Jimenez-Barbero, J., Gabius, H. J., and Mayo, K. H. (2012), *Biomol NMR Assign* **6**, 127-129.

PUBLICATIONS

Journal papers

1. Madariaga, D., Martínez. N., Somovilla, V., García, L., Berbís M.A., Martín-Santamaría, S., Hurtado, R., Asensio, J.L., Jiménez-Barbero, J., Avenoza, A., Busto, J., Corzana, F., Peregrina, J. Serine versus threonine glycosylation with α-*O*-GalNAc: unexpected selectivity in their molecular recognition with lectins. *Chem. Eur. J.* (2014) To article ⇨

2. Canales A., Mallagaray, A., Berbís M.A., Navarro-Vázquez A., Domínguez G., Cañada F.J., André S., Gabius H.J., Pérez-Castells J., Jiménez-Barbero J. Lanthanide-chelating carbohydrate conjugates are useful tools to characterize carbohydrate conformation in solution and sensitive sensors to detect carbohydrate-protein interactions. *J. Am. Chem. Soc*. (2014) 136, 8011-7. To article ⇨

3. Ippel H., Miller M.C., Berbís M.A., Suylen D., André S., Hackeng T.M., Cañada F.J., Weber C., Gabius H.J., Jiménez-Barbero J., Mayo K.H. 1H, 13C, and 15N backbone and side-chain chemical shift assignments for the 36 proline-containing, full length 29 kDa human chimera-type galectin-3. *Biomol. NMR Assign.* (2014). To article ⇨

4. Berbís M.A., André S., Cañada F.J., Pipkorn R., Ippel H., Mayo K.H., Kübler D., Gabius H.J., Jiménez-Barbero J. Peptides derived from human galectin-3 N-terminal tail interact with its carbohydrate recognition domain in a phosphorylation-dependent manner. *Biochem. Biophys. Res. Commun.* (2014) 443, 126-31. To article ⇨

5. Prinsen P., Rencoret J., Gutiérrez A., Liitia T., Tamminen T., Colodette J., Berbís M.A., Jiménez-Barbero J., Martínez A.T., Del Río J. Modification of the lignin structure during alkaline delignification of Eucalyptus wood by kraft, soda-AQ and soda-O2 cooking. *Ind. Eng. Chem. Res.* (2013) 52, 15702-12. To article ⇨

6. Ermakova E., Miller M.C., Nesmelova I.V., López-Merino, L., Berbís M.A., Lagartera L., Daragan V.A., Andre S., Cañada F.J., Jimenez-Barbero J., Solís D., Gabius H.J., Mayo K.H. Lactose binding to human galectin-7 (p53-induced gene 1) induces long-range effects through the protein resulting in increased dimer stability and evidence for positive cooperativity. *Glycobiology* (2013) 23, 508-23. To article ⇨

7. Fernández, M.C., Díaz D., Berbís M.A., Marcelo F., Cañada J., Jimenez-Barbero J. Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr. Protein Pept. Sci* (2012), 13, 816-30. To article ⇨

8. Nesmelova I.V., Berbís M.A., Miller M.C., Andre S., Jimenez-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, 127-9. To article ⇨

9. Berbís M.A., Sánchez-Puelles J.M., Cañada F.J., Jiménez-Barbero, J. Structure and function of bacterial UDP-glucose pyrophosphorylase, a drug target candidate. *Curr. Med. Chem.* (*2014*) (accepted).

Book chapters

10. Ardá A., Berbís M.A., Blasco P., Canales A., Cañada F.J., Fernández-Alonso M.C., Marcelo F., Jiménez-Barbero J. Recent advances on the application of NMR methods to study the conformation and recognition properties of carbohydrates. In: *Carbohydrate Chemistry, Vol. 38.* Royal Society of Chemistry (2012). To article ⇨

11. Fernández-Alonso M.C., Berbís M.A., Canales A., Ardá A., Cañada F.J., Jiménez-Barbero J. New applications of high-resolution NMR in drug discovery and development. In: *New applications of NMR in drug discovery and development.* Royal Society of Chemistry (2013). To article ⇨

12. Berbís M.A., Canales A., Sastre J., Unione L., Fernández-Alonso M.C., Blasco P., Cañada F.J., Jiménez-Barbero J. Advanced NMR techniques: defining carbohydrate structures and ligand-receptor interactions. In: *Carbohydrates Chemistry: State-of-the-art and challenges for drug development*. Imperial College Press (*2014*) (in press).

Manuscripts in preparation

13. Berbís M.A., Mallagaray, A., Canales A., Cañada F.J., André S., Gabius H.J., Pérez-Castells J., Jiménez-Barbero J. Disulfide-based paramagnetic tagging of carbohydrates for the study of protein-sugar interactions. (in preparation)

14. Berbís M.A., Mallagaray A., Cañada F.J., André S., Gabius H.J., Pérez-Castells J., Jiménez-Barbero J. Fluorine-19 spin labeling of proteins: application to diffusion spectroscopy and paramagnetism-enhanced detection of protein-ligand interactions. (in preparation)

15. Berbís M.A., André S., Cañada F.J., Gabius H.J., Jiménez-Barbero J. Fragment-based targeting of galectin-7 (p53-induced gene 1): small molecule inhibition of its dimerization and binding at a novel site on the gal-7 surface. (in preparation)

16. Valderrama J.A., Berbís M.A., Durante-Rodríguez G., Cañada F.J., Jiménez-Barbero J., Díaz E. The AccS histidine kinase domain of Azoarcus sp. CIB is regulated through an oxidation-dependent inhibition of its dimerization. (in preparation)

17. Van de Weerd R., Berbís M.A., Maaskant J.J., Sparrius M., Boot M., Paaw N.J., de Vries N., Boon L., Baba O., Geurtsen J.J., Canada F.J., Jiménez-Barbero J., Appelmelk B.J. A murine monoclonal antibody to glycogen: characterization of epitope-fine specificity by NMR and its use in mycobacterial capsular alpha-glucan cell-surface research. (in preparation)

18. Singh A.K., Berbís M.A., Jiménez-Barbero J., Menendez M., Kilcoyne M., Joshi L., Ballmann M.Z., Benko M., Harrach B., van Raaij M.J. Structure and sialyllactose binding of the carboxy-terminal head domain of the fiber from a Siadenovirus, turkey adenovirus 3. (in preparation)