UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

Novel ligand-based ^{19}F-NMR and computational approaches for studying sugar-lectin molecular recognition events.

Metodologías innovadoras basadas en ^{19}F-RMN de ligando y técnicas computacionales para el estudio de procesos de reconocimiento molecular azúcar-lectina.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

JOSE DANIEL MARTÍNEZ ORDÓÑEZ

DIRECTORES

JESÚS JIMÉNEZ BARBERO
FRANCISCO JAVIER CAÑADA VICINAY

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

Novel ligand-based ^{19}F-NMR and computational approaches for studying sugar-lectin molecular recognition events.

Metodologías innovadoras basadas en ^{19}F-RMN de ligando y técnicas computacionales para el estudio de procesos de reconocimiento molecular azúcar-lectina.

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

JOSE DANIEL MARTÍNEZ ORDÓÑEZ

DIRECTORES

JESÚS JIMÉNEZ BARBERO
FRANCISCO JAVIER CAÑADA VICINAY

CIC bioGUNE
MEMBER OF BASQUE RESEARCH & TECHNOLOGY ALLIANCE

CIB CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

A mi familia, por su amor y apoyo infinito en todos los sentidos.

Agradecimientos

Me gustaría comenzar agradeciendo a mis directores de Tesis, Jesús y Javier, la oportunidad que me brindaron hace ya algunos años de poder unirme a su grupo de investigación, primero en el CIB en Madrid, y más tarde en CIC bioGUNE en Vizcaya.

Cuando Jesús me propuso trasladarme a Bilbao para realizar mi tesis doctoral allí bajo su dirección, no era consciente aún de la inmensidad de lo que se me venía por delante. Sin esa primera oportunidad, no habría podido disfrutar de la etapa más estimulante intelectualmente (y personalmente) de mi vida. Siempre le estaré muy agradecido por ello. También por su ayuda en todo cuanto he necesitado hasta ahora, por sus enseñanzas sobre azúcares y RMN o sobre el funcionamiento del mundo científico, su eficacia para resolver cualquier tipo de problema, y sobre todo por la confianza depositada al haberme permitido siempre seguir mi camino con la flexibilidad para dejarme 'hacer y deshacer' a mi criterio (sé que soy un poco *especialito* a veces), sugiriendo lo que creía mejor para mí en cada momento.

Javier Cañada fue mi primer mentor en el estudio por RMN de la estructura de biomoléculas y sus interacciones (mis conocimientos previos, como los de cualquier estudiante de Química, se limitaban a la elucidación estructural de compuestos orgánicos). Junto a él me sumergí, neófito, en todo este universo fascinante que es la teoría sobre la que se erigen los experimentos de RMN que aplicamos a diario (*relajación, bendita locura*), y comencé a utilizar un espectrómetro de RMN en la planta baja del CIB. Muy pronto comprendí que Javier, siempre tan humilde, era un científico superlativo que sabía sobre casi todo, mucho. Para mí ha sido todo un ejemplo de humildad, tesón, paciencia e impecables valores científicos desde la primera conversación que tuvimos a mediados de mi curso de Máster cuando le conocí hasta el día de hoy.

Asimismo, me gustaría agradecer a mis compañeros del grupo *Chemical Glycobiology Lab* la ayuda prestada de una u otra forma: Ana Poveda, Ardá y Gimeno (*la terna de las Anas*), Marta, Luca, Javier, June e Íker. A Sandra muy especialmente por su ayuda en todo lo que se refiere a las bio-cosas, la expresión y purificación de DC-SIGN, tan 'intensita' sobre todo durante el primer año de Doctorado. A María Moure por su ayuda con la síntesis estos últimos meses, en los que estaba a otras cien cosas a la vez. Al *presi* Antonio, por las fructíferas discusiones científicas compartidas, por ser un ejemplo de perseverancia y dedicación a la ciencia, pero especialmente por su voluntad desinteresada de ayudar a cualquiera cuando es necesario y la buena energía que desprende siempre. ¡Viva Morón! También quería transmitir mi gratitud a los doctorandos (algunos de ellos ya doctores). A Pablis, mi hermano científico, que es una de las personas más humildes y bondadosas que he conocido. Nuestra amistad comenzó en aquel Máster sobre cómo descubrir fármacos, y ha ido *in crescendo* hasta el día de hoy. Gracias por escuchar todas mis preocupaciones científicas y vitales, discutir y rebatir cada uno de mis argumentos objetivamente, y por el

apoyo y la ayuda prestada siempre que te he necesitado, sin excepciones. Más allá del universo del análisis molecular, he tenido la gran suerte de estar a tu lado compartiendo aventuras de todos los sabores, incluyendo viajes, congresos veraniegos en la playa, charlas interminables de cine, ciencia (y ficción), videojuegos o de nuestros *Yo* pasados y presentes… ¡Espero haber estado a la altura, amigo! Gracias de corazón a Sarita, con la que conecté muy rápidamente desde el día en que aterrizó en el laboratorio (¡*vacaciones pagadas!*), y que se convirtió poco después en una pieza fundamental para completar el engranaje del *ICS group* junto a Pablis. Has sido un gran apoyo para mí dentro y, sobre todo, fuera del trabajo, donde siempre nos hemos entendido genial (mirada, guiño, risita y agur). Desde Valencia a Val di Mello, pasando por muchas experiencias *Mystic*as, paseos al Eroski o aprovisionamiento de Vitamina D, ha sido un placer. Nada de esto habría sido la mitad de divertido, por supuesto, sin Jon(i), nuestro *Alexa* personal y guía espiritual en las vascongadas, reportero fotográfico complementario en nuestros viajes, y siempre dispuesto a echar una mano cuando hace falta; y sin María Pía, nuestra querida 'madre' en Bilbao, que muy a menudo ha puesto algo de cordura a todas nuestras tonterías diarias. A las nuevas generaciones también les debo haber tenido un final de Tesis tan divertido. A Marquitos, nuestro *bobeco*, cuyo rendimiento purificando Gal-8 va inevitablemente ligado al número de veces que ha escuchado *Dákiti* ese día. Y a Unai Atxabal, un tipo *mamadísimo* que, a pesar de todos los fagos que le acompañan, es capaz de hacer un *clean* de 150 kg en botas y sin despeinarse. El futuro del grupo está en "buenas manos". Por último, a Ilaria y Helena, con las que compartí laboratorio y cervezas una buena temporada en tierras vascas, y que se marcharon hace ya un tiempo, gracias también por haber formado parte de esta historia.

Agradezco enormemente al grupo *Computational Chemistry Lab* su colaboración en todos los ámbitos, su buen rollo y su ejemplo de trabajo metódico y de calidad. En particular a Gonzalo por estar siempre disponible para echar una mano, escuchar mis inquietudes y resolver mis profundas dudas científico-existenciales o proponer constantemente nuevas ideas súper interesantes, y por las estimulantes charlas científicas que surgen a menudo con él. Y por supuesto a Reyes, *ReyNunFra*, cuya disposición, ayuda y afecto dentro y fuera del laboratorio, especialmente en este último tramo del trayecto, han sido fundamentales para poder terminar de dar forma a todo este trabajo (interior y exteriormente). Reyes no es solo una científica perspicaz, metódica y perseverante con la que es un placer trabajar y aprender. Más importante aún, es una persona muy especial llena de luz que irradia energía positiva a todo aquel que tiene la suerte de poder estar cerca de ella. Gracias por toda esta electricidad, *Southern*. ¡No dejes de resplandecer nunca!

Me gustaría dar las gracias también, por la compañía y buenos ratos disfrutados, a todos los compañeros con los que he tenido el placer de coincidir en el laboratorio de Madrid tanto en mi época de estudiante de Máster, como en el Doctorado: Bea y Javierín (*menuda parejita*), Ángeles, Carmen, Silvia (*si estás ahí, manifiéstate!*), Eva, Paola, Amaia, Ioanna, Mónica… En especial, Bea, a ti por

tu ayuda decenas (¿centenas?) de veces en temas burocráticos que se me escapan, y por los viajes, congresos e incontables cotilleos divertidos que hemos vivido estos años. A mis amigos y amigas del Máster, con los que he compartido cervezas, viajes, festivales, desamores y charlas científico-musicales a partes iguales: Carlos, Israel, Jorge, Julia, Alejandro y Ainhoa. También al bonito grupo de químicos orgánicos, computacionales y de análisis que formamos aquel año loco en Toledo cuando trabajamos en Janssen… y que ha trascendido mucho más allá de la capital manchega (*lo que pasa en Moscú, se queda en Moscú*) ¡*JD* siempre llevará a los *Sintéticos* muy dentro!

Finalmente, a mi *crew* de Bilbao. A Eli, una de las personas más humanas y justas que he conocido jamás, y que ha sido también la mejor compañera de viaje durante estos años: nunca podré agradecerte del todo el apoyo, el cariño y el amor que me has brindado siempre, tu compañía y cálida acogida en estas tierras, el haberme dejado formar parte de tu fantástica cuadrilla; nada habría sido ni remotamente igual de bueno sin ti este tiempo. Gracias por haber sido tan buena conmigo, haber estado ahí cerca para todo, y haber sido tan flexible -a pesar de tu genética norteña clásica (*joking*!)- con mis volteretas vitales. A mis primeros y más puros amores en el Norte, Diego (*Diegus*) y Lucía (*Luciérnaga, Lucifer*), mejores entre los mejores amigos (*soulmates*) que uno pudiera soñar tener; nunca jamás he reído y pasado unos años tan divertidos en casa, de viaje mochilero, tirados en una playa o bailando reguetón en mi vida, como los que he tenido la suerte de compartir con vosotros, chicos. Os quiero, y este amor es eterno :) A Itsaso, una de esas personitas con un brillo especial tan raras de encontrar, que conocí nada más trasladarme a Bilbo y que con su alegría y buen rollo constante se convirtió casi instantáneamente en una amiga de las que se cuentan con los dedos. Y finalmente, a las últimas incorporaciones de la pandilla, Ainhoa, y Beatriz. La primera es nuestra bióloga heavy-vasca favorita, que apareció un día por casa y se quedó a vivir en el cuarto de al lado; y que mola tanto que juega más a la PS4 que yo. La segunda es nuestra enfermera particular preferida (no conocemos a otra), siempre preparada con una historia desagradable sobre entrañas para compartir, que un día se plantó en La Salve bajo el brazo de Lucía y ya nunca quisimos que se fuera de nuestro lado. ¡Gracias por todo amig@s!

No podría terminar esta sección sin agradecer, a su vez, la enorme ayuda que diversos grandes científicos me han brindado en diferentes etapas de este periodo predoctoral. A Tammo, porque junto a él he podido profundizar enormemente en el funcionamiento más fundamental de los espectrómetros de RMN que utilizamos, y porque siempre ha estado dispuesto a echarme una mano y resolver mis dudas más enrevesadas. A Paco Corzana, porque dedicó un día entero (¡8 horas!) a introducirme en el universo de la dinámica molecular de una forma tan entretenida y estimulante, que despertó en mí unas ganas tremendas de aprender a fondo a utilizar la metodología desde ese encuentro breve en Logroño. A David Baker, que me acogió durante 3 maravillosos meses en su magnífico laboratorio en Seattle (*Institute for Protein Design*), y del que aprendí tantísimas cosas sobre el diseño de proteínas *de novo* y sobre técnicas

punteras de producción y caracterización de proteínas, que es imposible resumirlas en unas pocas líneas. Finalmente, quiero extender este agradecimiento a muchos otros investigadores que de una forma u otra han contribuido también a mi formación como científico, y al desarrollo de este trabajo en última instancia.

Por último, y muy especialmente, a mi familia. A mis padres, Galo y Mercedes, por su amor y apoyo ilimitado en todos los sentidos imaginables. A mis hermanos, Gloria y Alberto, por su empatía, cariño y ayuda, tanto en el fondo como en la forma, en el desarrollo de mi carrera científica y personal. Sois la mejor familia que uno pudiera soñar jamás :)

Puedo afirmar sin duda alguna que estos cuatro años de doctorado en el Norte han sido los más bonitos e intensos de mi vida. Gracias, en definitiva, a todos los que habéis formado parte, de una forma u otra, de este maravilloso camino.

It is a capital mistake to theorize before you have all the evidence. It biases the judgment.

Sherlock Holmes

It is better to be roughly right than precisely wrong.

Alan Greenspan

CONTENTS

Abbreviations

ABC	ATP-binding cassette
ATP	Adenosine TriPhosphate
AMBER	Assisted Model Building with Energy Refinement
CLR	C-type Lectin Receptor
CORCEMA	Complete Relaxation and Conformational Exchange Matrix
COSY	COrrelation SpectroscopY
CPMG	Carr-Purcell-Meibom-Gill
CRD	Carbohydrate Recognition Domain
CSA	Chemical Shift Anisotropy
CSP	Chemical Shift Anisotropy
DC-SIGN	Dendritic Cell-Specific Intercellular adhesion molecule -3-Grabbing Non-integrin
DD	Dipole-Dipole
DNA	DeoxyriboNucleic Acid
DQ	Double-Quantum
ECD	ExtraCellular Domain
EDTA	Ethylene Diamino Tetraacetic Acid
EM	Electron Microscopy
ER	Endoplasmatic Reticulum
FID	Free Induction Decay
FT	Fourier Transform
GAFF	General AMBER Force Field
GGBP	Glucose and Galactose Binding Protein
GPI	GlycosylPhosphatidylinositol
GPU	Graphical Processing Unit
HIV	Human Immunodeficiency Virus
HML	Human Macrophage Lectin
HSQC	Heteronuclear Single Quantum Correlation
HTS	High-Throughput Screening
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	Lysogeny Broth
LNFP III	Lacto-N-FucoPentaose type III
MC	Monte Carlo
MD	Molecular Dynamics
MGL	Macrophage Galactose-type Lectin
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OD	Optical Density
PAMP	Pathogen-Associated Molecular Pattern
PBC	Periodic Boundary Conditions

PBP	Periplasmic Binding Proteins
PCS	PseudoContact Shifts
PET	Positron Emission Tomography
PME	Particle Mesh Ewald
PSA	Pisum Sativum Agglutinin
QM	Quantum Mechanics
RF	RadioFrequency
RDC	Residual Dipolar Coupling
SAR	Structure-Activity Relationship
SPR	Surface Plasmon Resonance
SQ	Single-Quantum
STD	Saturation Transfer Difference
TOCSY	TOtal Correlation SpectroscopY
UV	UltraViolet spectroscopy
ZQ	Zero-Quantum

RESUMEN

Metodologías innovadoras basadas en ^{19}F-RMN de ligando y técnicas computacionales para el estudio de procesos de reconocimiento molecular azúcar-lectina

Los carbohidratos juegan un papel fundamental en una enorme variedad de procesos biológicos. Se encuentran en todos los organismos vivos en la naturaleza, donde intervienen en funciones que abarcan desde su uso como fuente de energía o como fragmentos estructurales, hasta procesos de infección en organismos superiores. En vertebrados, se localizan tanto en la superficie celular como en el espacio extracelular, formando estructuras muy diversas y complejas, pero también están presentes en el núcleo y citoplasma de células eucariotas unidos a proteínas (glicoproteínas). Su localización casi universal en el organismo les confiere la capacidad de intervenir en un gran número de procesos de 'comunicación' con otras entidades, por ejemplo, interacciones intercelulares, célula-molécula y célula-matriz extracelular. Además, los carbohidratos median procesos de reconocimiento molecular entre distintos organismos, como el reconocimiento de patógenos y parásitos por la célula de un huésped.

De entre todas las biomoléculas que interaccionan con carbohidratos, en esta Tesis nos hemos centrado en su reconocimiento molecular por parte de lectinas. Las lectinas son proteínas que carecen de actividad catalítica y no son producidas por una respuesta inmune (no son anticuerpos). Sin embargo, juegan un papel fundamental en la comunicación entre células, median en procesos de infección e inmunidad, o participan en la limpieza de glicoproteínas extracelulares entre otros.

Para abordar el estudio a nivel atómico/molecular de los procesos de reconocimiento entre azúcares y lectinas, hemos empleado la espectroscopía de Resonancia Magnética Nuclear (RMN) como técnica analítica principal. En particular hemos utilizado ampliamente RMN de flúor, ya que las posibilidades que ofrece este heteronúcleo desde el punto de vista del análisis de interacciones ligando-receptor son enormes. Hemos abarcado estos estudios fundamentalmente desde el punto de vista del carbohidrato, para lo que hemos utilizado distintos glicomiméticos fluorados (es decir, moléculas químicamente modificadas que imitan a carbohidratos naturales). Nuestro objetivo ha sido doble: implementar nuevas metodologías basadas en RMN de flúor para la identificación y caracterización del estado unido de azúcares que interaccionan con lectinas; y profundizar en el entendimiento de cómo estas interacciones suceden.

De forma paralela a los experimentos de RMN, hemos utilizado métodos computacionales para dar soporte a los datos experimentales obtenidos. En concreto, hemos hecho uso principalmente de simulaciones de Dinámica Molecular (MD) que han sido esenciales en la interpretación de los resultados de RMN, así como para guiar los pasos experimentales a seguir en algunos casos. Hemos empleado a su vez otras herramientas computacionales en menor medida, como es el caso de la *matriz de relajación completa e intercambio conformacional* (CORCEMA-ST) que permite calcular intensidades de STD teóricas para un modelo ligando-receptor dado, pero que sin duda han resultado de suma utilidad.

A lo largo de los próximos capítulos, expondré las contribuciones que hemos realizado en los últimos años. Éstas incluyen:

- La descripción de una metodología basada en filtros de relajación T_2 para, de forma simultánea, identificar en una librería de azúcares aquellos que se unen a lectinas, así como realizar el mapeo químico de los grupos OH implicados en la interacción.
- La detección y caracterización -aplicando la metodología anterior- de un modo de unión desconocido de la interacción entre la lectina DC-SIGN y la manosa.
- La aplicación del nuevo experimento 2D-STD-TOCSYreF, con detección en flúor, para estudiar la interacción de DC-SIGN con un glicomimético fluorado, derivado del núcleo trimanósido que se encuentra en el centro de las cadenas glicánicas de las N-glicoproteínas de mamíferos. Este experimento resulta especialmente útil en casos en los que el solapamiento de señales en el espectro de ^{1}H-RMN hace imposible la estimación adecuada de los efectos STD en cada protón.
- El desarrollo de una metodología para estudiar múltiples modos de unión de ligandos a un mismo receptor en disolución, utilizando datos de STD-RMN, simulaciones de MD, cálculos de CORCEMA y un nuevo programa desarrollado para el análisis combinado de todos estos datos, *BM-Mixer*.

ABSTRACT

Novel ligand-based ^{19}F-NMR and computational approaches for studying sugar-lectin molecular recognition events

Carbohydrates play a central role in a large myriad of biological processes. They are found in all living organisms in nature, participating in different functions ranging from their use as energy source or as structural fragments, to infection-related processes in complex organisms. In vertebrates, they are located both in the cell surface and in the extracellular space, forming very diverse and intricate structures, but they are also present in the nucleus and cytoplasm of eukaryotic cells bound to proteins (glycoproteins). Their location almost ubiquitous in the organism confers them the capacity of mediate in a large number of 'communication' processes with other entities, for instance, in cell-cell, cell-molecule and cell-matrix interactions. In addition, carbohydrates intervene in molecular recognition processes between different organisms, such as the pathogen and parasite recognition by host cells.

Among all the biomolecules that interact with carbohydrates, we have focused on their molecular recognition by lectins in this Thesis. Lectins are proteins lacking catalytic activity and they are not produced by an immune response (they are not antibodies). Nevertheless, they play a fundamental role in intercellular communication, mediate in infections and immune processes, and participate in extracellular glycoprotein cleaning among others.

To address the study at the atomic/molecular level of the recognition processes between sugars and lectins, we have applied Nuclear Magnetic Resonance (NMR) spectroscopy as the main analytical technique. In particular, we have largely used fluorine NMR, since the possibilities offered by this heteronucleus are enormous. We have fundamentally covered studies from the carbohydrate point of view, for which different fluorinated glycomimetics (i.e., chemically modified molecules that mimics the natural carbohydrates) have been employed. Our objective has been twofold: the implementation of new methodologies based on fluorine NMR to identify and characterize the bound states of sugars that interact with lectins; and to deepen in the understanding of how these interactions occur.

In parallel to the NMR experiments, we have used computational methods to support the obtained experimental data. In particular, we have mainly employed Molecular Dynamics (MD) simulations that resulted to be indispensable in the interpretation of NMR data, as well as a guide for future

experimental directions to follow in some cases. Furthermore, we have made use of other computational tools to a lesser extent, such as the *COmplete Relaxation and Conformational Exchange MAtrix* (CORCEMA-ST) that allows calculating theoretical STD intensities for a given ligand-receptor model, which undoubtedly have turned to be very helpful too.

Throughout the following chapters, I will illustrate the contributions to the field that we have achieved over the last years. These include:

- The description of a methodology based on T_2 relaxation filters to, simultaneously, identify from a library of sugars those that bind lectins, and chemically map the OH groups involved in the interaction.
- The detection and characterization –by applying the previous methodology- of an unidentified binding mode in the interaction between DC-SIGN lectin and mannose.
- The application of the novel 2D-STD-TOCSYreF experiment, with fluorine detection, to study the interaction of DC-SIGN with a fluorinated glycomimetic derived from the trimannose core ubiquitous in mammal N-glycoproteins. This experiment turns out to be particularly useful in cases in which the signal overlapping in the ^{1}H-NMR spectrum prevents the proper estimation of individual STD effects.
- The development of a new methodology to study multiple binding modes of ligands to the same receptor in solution, using STD-NMR data, MD simulations, CORCEMA-ST calculations and a new program specifically developed to perform the combined analysis of these data, *BM-Mixer*.

CHAPTER 1

Introduction and theoretical framework

The chemistry of carbohydrates and Lectins

1. CARBOHYDRATES

Carbohydrates (also referred to as *glycans*, *saccharides* or simply *sugars*) are polyhydroxylated and chiral organic molecules present in all living organisms. They are involved in a myriad of functions, among which the production of energy, their role as structural constituents of cell walls and extracellular matrix, or their participation in molecular recognition events phenomena and cell signalling are some of major importance (Figure 1). Although they exist as individual entities, they are usually found conjugated to proteins or lipids, forming *glycoconjugates*. In particular, glycoconjugates are among the main constituents of the external surface of mammalian cells and are expressed at different phases of differentiation. Modifications in oligosaccharide composition of the cell surface are related with many different pathological conditions [1,2].

Carbohydrates, along with proteins, nucleic acids and lipids, are one of the four essential classes of molecules found in all living systems. Unlike proteins or DNA, glycans are typically highly branched molecules with a much wider conformational variability. Therefore, the study of these molecules is particularly challenging and still relatively underexplored in comparison with the development in other fields of life sciences (such as genomics or proteomics).

It has been our aim within the work performed along these years to contribute in extending the knowledge about these fundamental chemical entities and their role in the biological context.

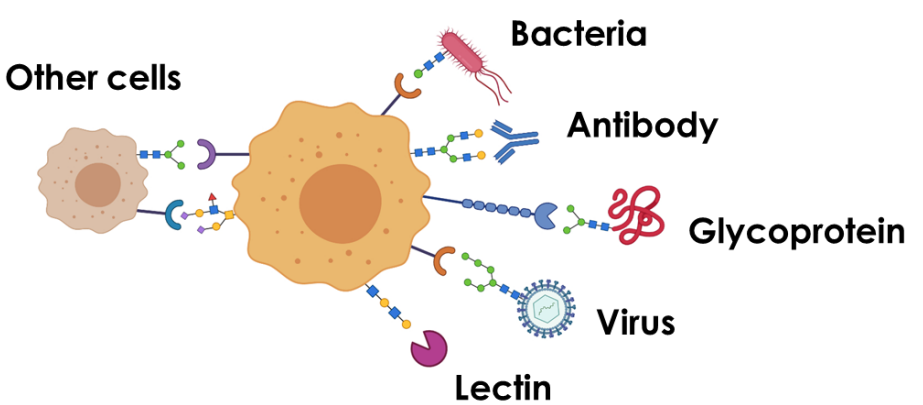


Figure 1. Carbohydrates participate in many molecular recognition processes in living organisms.

1.1. CHEMICAL FEATURES OF CARBOHYDRATES

Among the possible classification patterns in which sugars can be sorted, it is especially convenient the one referring to the number of the simplest 'building blocks' or units that conform a *carbohydrate*. In that sense, the minimum unit that possesses the fundamental features of carbohydrate is named *monosaccharide*. More precisely, monosaccharides are polyhydroxy aldehydes (*aldoses*) or ketones (*ketoses*) with at least three carbon atoms (Figure 2). Other molecules derived from these basic units, for instance, by oxidation of hydroxyl terminal groups to carboxylic acids, or by substitution of hydroxyl groups by hydrogen or other heteroatoms, are considered carbohydrates as well. Higher order carbohydrates arise by connecting monosaccharide units through *glycosidic linkages*: *oligosaccharides* are formed by 2 to 10 monosaccharide units and *polysaccharides*, by more than 10 units [3,4].

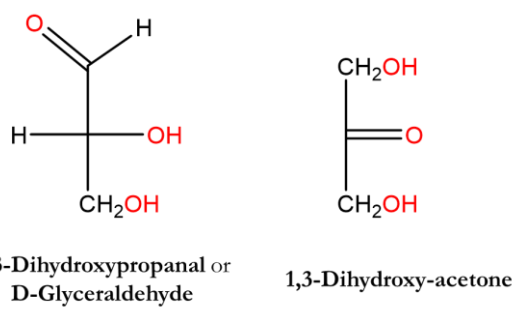


Figure 2. The simplest aldose and ketose molecules in Fischer projection.

To set the scene, some basic description of carbohydrate nomenclature is required. For aldoses, the aldehyde function has always the highest priority for numbering the carbon atoms, thus it is assigned as C1. In the case of ketoses, the carbon of the ketone function receives the lowest possible number. Additionally, to distinguish between enantiomers, the *D*/*L* nomenclature introduced by Fischer is employed:

If the hydroxyl group attached to the asymmetric carbon with the highest numbered position projects on the right when is represented in the Fischer projection, the compound has the D-configuration by definition. The reverse situation corresponds to the L-configuration.

D-sugars are the most commonly found in nature. Although several theories have been proposed to explain this homochirality[1], its exact origin is still unknown.

[1] Homochirality is also observed in the preference for L- over D-amino acids in nature.

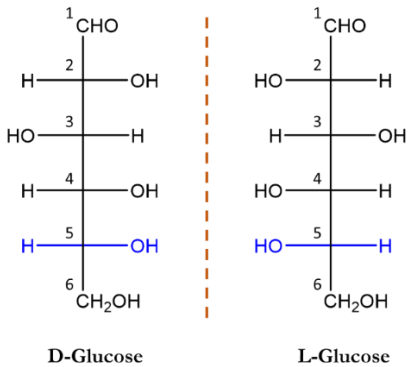


Figure 3. Representation of the D and L enantiomers of glucose. The hydroxyl group attached to the asymmetric carbon furthest from the aldehyde function (C5) is represented in blue.

In general, trivial names are employed when referring to aldoses and ketoses with up to six carbon atoms, as shown in Figure 4.

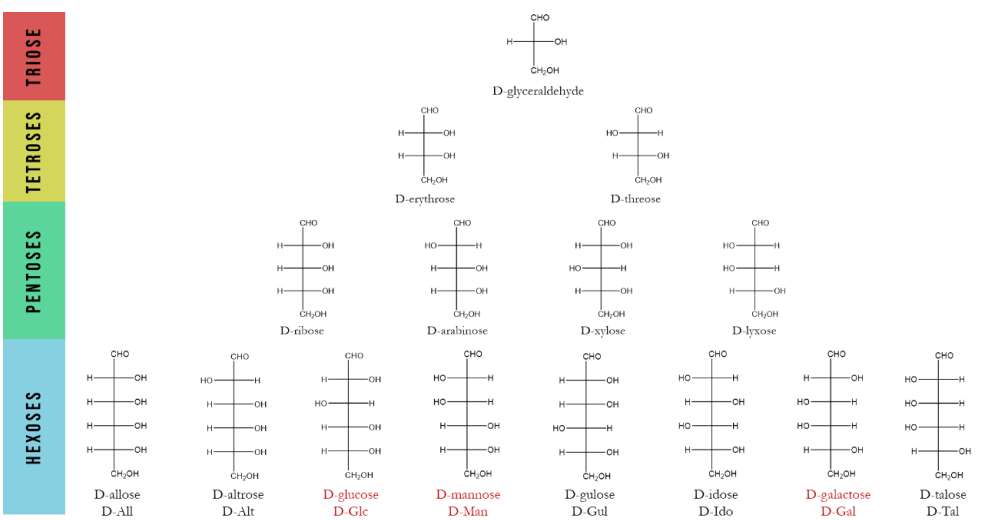


Figure 4. Structures of the D-aldose series, with up to six carbon atoms. The sugars highlighted in red (or derivatives) are the ones most commonly employed in this work.

In aqueous solution, monosaccharides of at least four carbon atoms can undergo an intramolecular hemiacetal formation: the hydroxyl group at C4 or C5 attacks the aldehyde carbon, to yield stable five or six membered rings. In fact, the equilibrium is highly displaced towards the cyclic forms, and the acyclic (linear chain) form exists only in trace (< 0.1%) amounts in the case of pentoses and hexoses. Rings of six and five members are named *pyranose*

and *furanose* respectively. This equilibrium between acyclic and cyclic forms in solution is known as *mutarotation*[2] [5] (Figure 5).

The intramolecular hemiacetal formation implies the creation of a new stereocenter, since C1 becomes now asymmetric. The hemiacetal formation can take place from either of the two faces:

The asymmetric center arising at C1 from the intramolecular hemiacetal formation is named *anomeric* and the two possible isomers are referred to as alpha (α) and beta (β) *anomers*. Using the Haworth projection, the α anomer is the one in which the exocyclic OH (or -OR) group at the anomeric center and the -CH_2R group (i.e., the substituent at C5) are on opposite faces. If the two groups are on the same face, the isomer has the β configuration.

Haworth projections (Figure 5, bottom) are simplifications of the real 3D carbohydrate structures, useful to illustrate stereochemical relationships. However, the 3D structure of carbohydrates crucially influences their physico-chemical properties and reactivity. Thus, it is appropriate at this point to turn into a more accurate sugar representation, closer to a realistic view of the disposition of the ring carbon atoms and their substituents in space. For pyranose rings, it turns out that the *chair* conformation is the most favoured energetically (see the pyranose conformation of D-Man in Figure 5). The conformational preferences of pyranose rings will be addressed in the following section.

[2] In the presence of an alcohol or amine in acidic conditions, the reaction continues to the acetal formation, and the resulting product is named as *glycoside*. In that case, since C1 is an acetal function in glycosides (i.e., the substituent at C1 is a –OR group), mutarotation does not take place in water for such species.

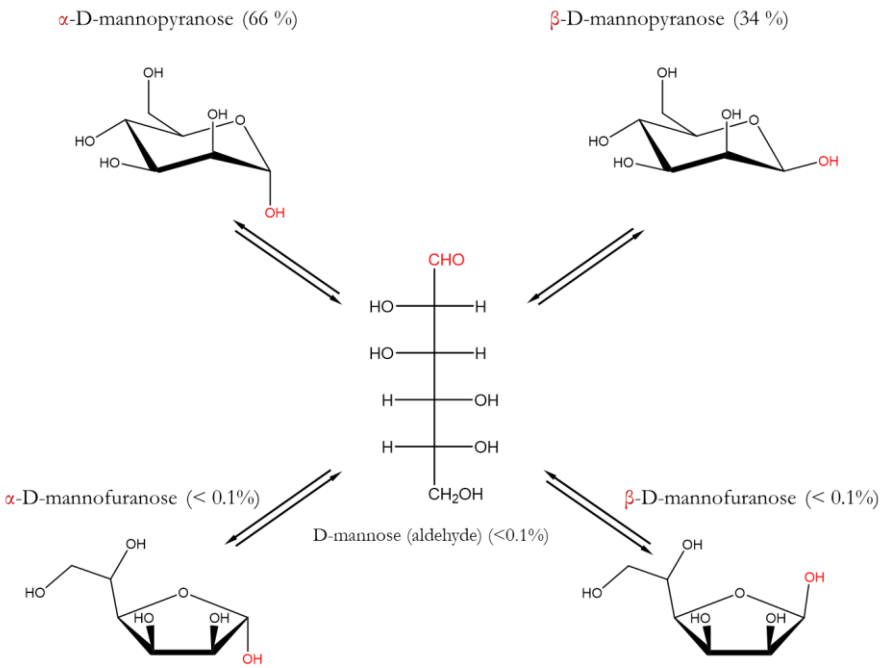


Figure 5. Mutarotation equilibria for D-Man, using the Haworth projection for the furanose rings and the chair representation for pyranoses. The anomeric hydroxyl (aldehyde function in the linear form) is shown in red. Notice that for the α anomer, the anomeric hydroxyl is in axial disposition in the pyranose ring, while for the β anomer this hydroxyl is equatorial.

The scientific projects addressed throughout this dissertation mainly involve the study of mono- and oligosaccharides of *aldohexoses* (aldose with six carbon atoms), forming virtually only pyranose rings in solution. Therefore, the following discussion will be particularly focused in these six-membered ring sugars.

1.2. CARBOHYDRATE CONFORMATIONS

The conformations adopted by carbohydrates are key to understand their physico-chemical properties and reactivity. As it will be introduced in the following lines, monosaccharides fundamentally exist in pyranose and furanose forms, since these shapes minimize eclipsing interactions and favour certain stable angular torsions. Once the study of the monosaccharide conformation is addressed, the conformational preferences of oligosaccharides, which are determined by the torsional angles across the glycosidic linkage, will be covered.

1.2.1. Monosaccharides

Pyranose (and furanose) rings can in theory adopt several conformations that interconvert one into another just by rotating bonds. The IUPAC classifies pyranose ring shapes into 38 different conformations that in turn are distributed in 5 groups [6,7]: 2 *chairs*, 6 *boats*, 6 *skew-boats*, 12 *half-chairs*, and 12 *envelopes* (Figure 6). For energetic reasons, in general only the chair conformations, abbreviated C, are dominant for pyranose rings. Nevertheless, although other conformations can be expected in some cases due to the influence of ring substituents or for reaction intermediates [8]. In pyranoses, the chair conformation is defined by displaying the four atoms C2-C3-C5-O in a reference plane (being O the endocyclic oxygen). The remaining carbon atoms are then above or below this plane. The positions of these out-of-plane atoms are given as superscripts and subscripts for the atoms above and below the plane, respectively.

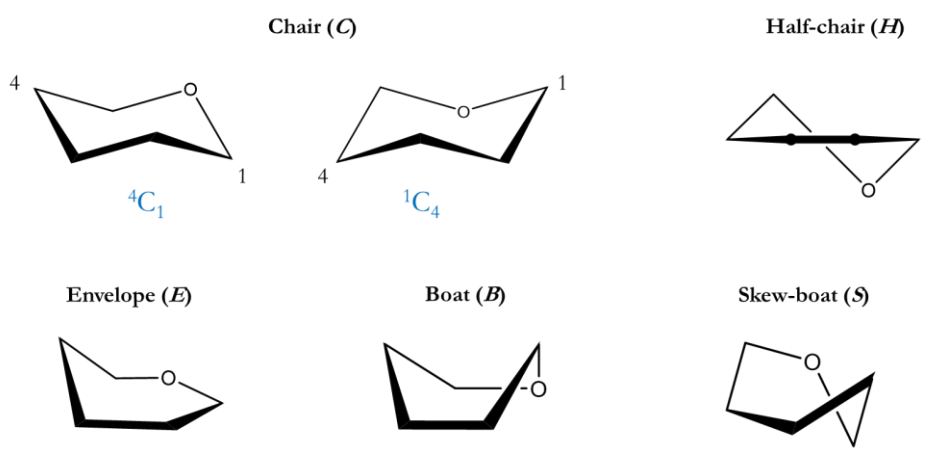


Figure 6. Pyranose conformations from the 5 different groups. In the case of the chair representations, the only two possible conformers are depicted.

Depending on the particular sugar entity, the $^4C_1$ or $^1C_4$ chairs are preferred (Figure 7, see the explanation below). Ultimately, the preference for one or another of the chair conformations depends on the equilibrium between torsional and angular energies, steric strain, as well as electrostatic effects, modulated by the presence of different substituents that might interact between each other. Moreover, when the substituents are polar groups, the solvent plays also an important role in determining the preferred conformations.

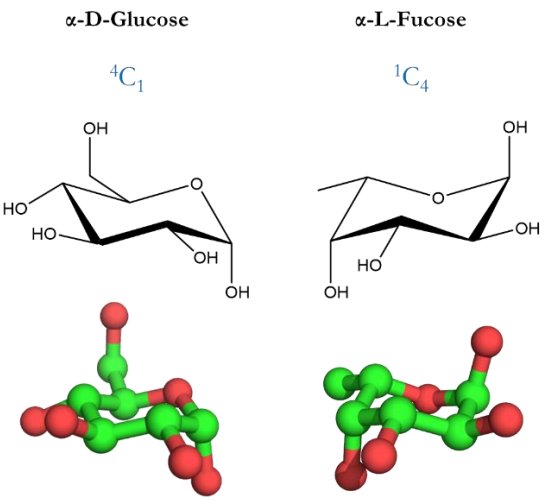


Figure 7. Chair representations (top) and 3D models (bottom) of α-D-Glc and α-L-Fuc. Due to the different substitution patterns, different chairs are preferred in each case. In the 3D models, hydrogens are omitted for clarity.

An interesting feature of the chair conformations is that substituents may adopt either axial or equatorial arrangements.

For substituted cyclohexanes, the conformation that places more substituents in an equatorial disposition is typically the most stable.

This observation has to do with a number of reasons. For instance, unfavourable *1,3-diaxial interactions*, which arise by electronic and steric repulsion between two axially distributed electronegative substituents separated by two C-C bonds, are avoided in the di-equatorial alternative. In monosubstitued cyclohexanes, an additional driving force towards the equatorial conformer is that the C–X bond (where X is the substituent) is antiperiplanar (or *trans*[3]) to two C–C bonds, whereas in the case of the axial conformer, the C–X bond is synclinal (*gauche*). When more substituents come into play, the specific situations must be analyzed similarly and the preferred conformations will ultimately depend on the nature of those substituents [5].

For carbohydrates in the pyranose form, however, remarkably a predilection of substituents at C1 position for an axial over an equatorial orientation is observed (in general, for tetrahydropyran derivatives).

The preference of tetrahydropyrans bearing an electronegative substituent at position C2 (equivalent to C1 in sugars) for an equatorial conformation is known as the *anomeric effect*.

[3] *Gauche* and *trans* torsion angles correspond to ±60º and ±180º respectively, within an interval of ±30º (goldbook.iupac.org/terms/view/C01088).

The anomeric effect was originally observed by Edwards and named by Lemieux [9,10]. The effect becomes more pronounced with the increase in the electronegativity of the substituent and decrease with the dielectric constant of the solvent. Two possible explanations to this phenomenon were initially proposed: a) the stabilization of the axial configuration at C1 arise from the favourable interaction of the lone electron pair of a molecular orbital on the endocyclic oxygen and the σ^* orbital of the exocyclic C-X bond (this is known as *hyperconjugation*); and, b) the dipole-dipole repulsion involving the endocyclic oxygen and X is minimized in the axial configuration at C1. Although it is likely that the anomeric effect is the result of both of these different contributions, the former explanation is more broadly accepted (Figure 8).

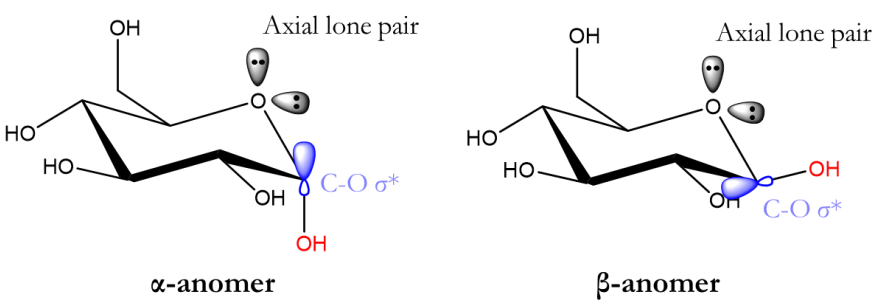


Figure 8. Simplified representation of the anomeric effect for D-Glucose, based on favourable/unfavourable orbital overlapping of the empty antibonding C-O σ^* (in blue) and one occupied orbital on the endocyclic oxygen. The electron density is more favorable transferred for the α anomer.

Finally, the conformations adopted by the exocyclic hydroxymethyl (or in general, -OR) group in pyranoses must also be noted. The potential energy surface for rotation about the C5-C6 bond has three minima, giving rise to three distinct rotamers. To classify these conformers a two-letter nomenclature is employed, so they are categorized as GG (*gauche-gauche*), GT (*gauche-trans*) and TG (*trans-gauche*). The first letter indicates the torsion angle between the endocyclic oxygen O5 and the exocyclic oxygen O6, and the second letter refers to the dihedral angle between C4 and O6 (Figure 9). It has been observed that the orientation of the substituent at C4 is the fundamental factor determining the relative rotameric populations, which are modulated by 1,3-diaxial interactions, *gauche effects*[4], solvation and hydrogen

[4] The gauche effect refers to the stabilization of the gauche conformation (torsion angle of about ± 60º) of two vicinal electronegative substituents, over the anti (± 180º) that would reduce steric repulsions. It can be explained by *hyperconjugation*, similarly to the previously discussed anomeric effect.

bonding involving O4 and O6. Additionally, solvent effects, the configuration at C1 and the particular substituents of the pyranose ring are also important to a lesser extent [11,12]. It has been observed that D-Glc and D-Man, whose OH at C4 is in an equatorial conformation, have rotamer populations of approximately gg:gt:tg ≈ 60:40:0 [11–14]. On the contrary, for D-Gal, with axial OH at C4, the relative populations found are roughly gg:gt:tg ≈ 10-20:50-60:20-30 [11,15,16].

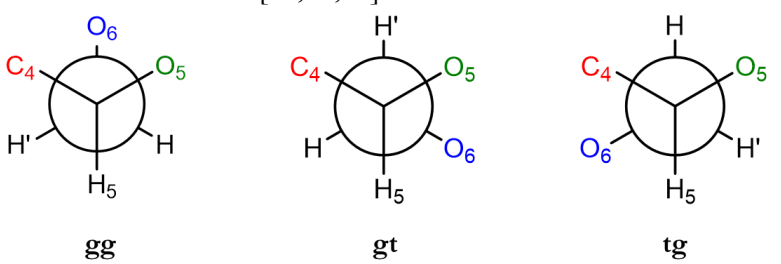


Figure 9. Newman projections around the C5-C6 bond showing the GG, GT and TG conformations of the exocyclic hydroxymethyl group.

1.2.2. Oligosaccharides

Oligosaccharides are formed when a monosaccharide reacts through its anomeric carbon (C1) with any of the OH groups of another monosaccharide, formally giving rise to an acetal and a water molecule. The new bond created between monosaccharide residues is called *glycosidic linkage*. In the case of two monosaccharides reacting through their anomeric OH's, the obtained disaccharide is classified as a *non-reducing* sugar. On the other hand, if one of the units still conserve the hemiacetal function at C1 (i.e, if the glycosidic linkage is not formed between the two anomeric carbons of the different monosaccharide units), the resulting sugar is a *reducing* one[5]. In such case, the sugar may continue reacting with other monosaccharides to form higher order oligosaccharides (trisaccharides, tetrasaccharides, etc.). Whatever the case, since the glycosidic linkage always implies the anomeric carbon of at least one of the monosaccharides, it can exist in either an α or β configuration [4,5].

Whereas the previous conformational portray presented for monosaccharides in the previous section applies likewise for oligosaccharides, in order to fully characterize their three-dimensional

[5] These terms, which are used for monosaccharides substituted at the anomeric carbon as well, arise from the capacity of sugars to act or not as reducing agents. They are classified as *reducing* sugars if the hemiacetal group is present (or an aldehyde or ketone group in the open-chain form). If the anomeric carbon is 'blocked' as an acetal function, they cannot behave as reducing agents and are named *non-reducing* sugars instead.

features it is necessary to introduce following torsion angle definitions [17,18]:

$$\phi\text{: O5}(i) - \text{C1}(i) - \text{O}n(i-1) - \text{C}n(i-1)$$

$$\psi\text{: C1}(i) - \text{O}n(i-1) - \text{C}n(i-1) - \text{C}(n-1)(i-1)$$

$$\omega\text{: O6}(i) - \text{C6}(i) - \text{C5}(i) - \text{C4}(i)$$

where n is a position in the ring and i indicates a given residue.

ϕ and ψ torsion angles describe the spatial disposition around the glycosidic linkage and are always required to describe the 3D structure of oligosaccharides. The angle ω is also important when the glycosidic linkage involves an exocyclic CH_2OH group. In fact, ω is the torsion angle around the C5-C6 bond discussed above for monosaccharides in terms of GG, GT and TG rotamers. Figure 10 shows schematically these dihedral angles in a disaccharide.

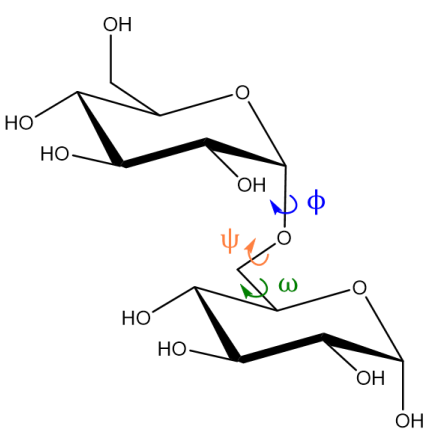


Figure 10. Torsion angles ϕ, ψ and ω in a disaccharide linked through carbons 1→6. Notice than ω is defined between atoms of the same residue.

The conformations about the anomeric C-O bond are of exceptional importance in carbohydrate chemistry. As it was firstly observed by Lemieux [19], substituents at the anomeric OH adopt preferentially a conformation in which $\phi \approx 60°$ for both α and β anomers. This phenomenon was called *exo-anomeric effect*, and, similarly to the anomeric effect, it can be explained in terms of the favorable interaction between an electron rich molecular orbital on the exocyclic oxygen when is parallel to the σ^* orbital of the C1-O5 bond, and by steric repulsion effects of substituents at C2 (Figure 11).

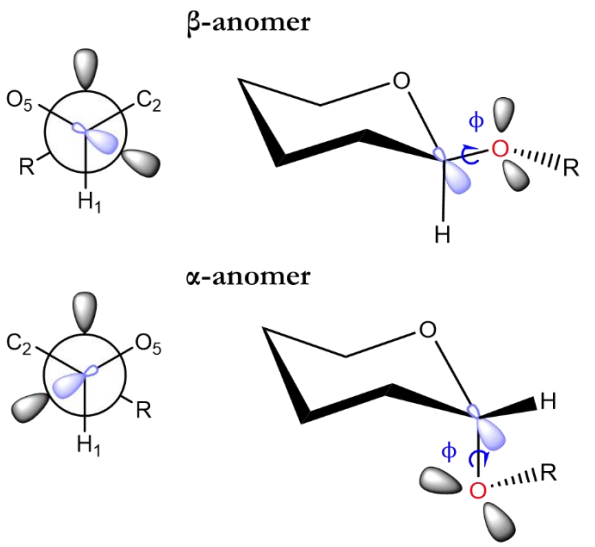


Figure 11. The exo-anomeric effect is a consequence of favorable orbital overlapping and steric repulsion. These are the preferred conformations for α and β anomers.

1.3. GLYCANS IN THE BIOLOGICAL CONTEXT

Carbohydrates are broadly spread in nature. In mammals, they can be found in virtually all intracellular compartments and in the extracellular region, both as free entities or attached to proteins and lipids forming an enormous variety of glycoconjugates. In fact, all animal cells are covered by a dense coat of complex glycans, referred to as the *glycocalyx*.

The assembly of glycoconjugates takes place by addition of sugars to proteins or lipids in a non-templated manner. This enzymatic process, mediated by *glycosyltransferases* and *glycosidases*, is known as *glycosylation* and take place in the Golgi apparatus and endoplasmic reticulum (ER). Briefly, proteins can be linked to glycans through the oxygen atom of the hydroxyl groups of Ser or Thr residues, or through the nitrogen atom of the amino group of Asp residues, giving rise to *O-glycans* and *N-glycans* respectively. Moreover, they can also be linked to *glycosaminoglycans*[6], *phosphorylated* glycans, or *glycosylphosphatidylinositol* (GPI) anchors[7]. Among glycolipids, *glycosphingolipids* are one the most commonly found in cellular membranes, and are formed by the linking of a glycan to a ceramide [20–22].

Deviations from normal glycosylation patterns in humans is related to different diseases. For instance, altered glycosylation strongly is associated

[6] Linear polysaccharides formed by repeating disaccharide units
[7] A short glycolipid embedded in the cell membrane.

with tumor development and spreading, autoimmune and chronic inflammatory diseases, as well as congenital disorders of glycosylation.

Glycans are also found abundantly in the surface of many pathogens, together with sugar binding proteins. In fact, most infectious diseases are related to protein-carbohydrate interactions between pathogen and host cell surfaces. For instance, human innate immune cells are able to identify the so-called pathogen-associated molecular patterns (PAMPs), which are glycans and glycoconjugates in many cases. The recognition process of these PAMPs initiates a cascade of intracellular signaling mechanisms leading to the adaptive immunity [2,23].

Hence, these pathophysiologically significant interactions have attracted much attention both to understand the molecular recognition processes and to develop sugar-based drugs capable of interfere on them. However, carbohydrate-like drugs still represent only a very limited region of the therapeutic world, probably due to the intrinsic pharmacokinetic limitations they display [24].

1.3.1. Glycomimetics: F-saccharides

From a physiological point of view, carbohydrates possess relatively poor pharmacokinetic properties. This is related with their physico-chemical features: they are highly polar molecules that suffer from passive permeation problems and low metabolic stability, which translate into low drug bioavailability, potency and fast renal excretion [24]. To overcome these flaws, synthetically modified carbohydrate analogues have been widely developed in recent years. These chemically altered molecules are known as *glycomimetics* [24,25]. Glycomimetics may not only be superior to natural carbohydrates in terms of general bioavailability and stability against enzymatic degradation, but they also can be devised with the goal of improving the affinity and/or selectivity towards a target receptor compared to the original sugar [26].

Nevertheless, the growth of glycomimetic design and synthesis does not only serve to the purpose of improving the natural sugar properties to increase its value as potential therapeutic agents. In fact, glycomimetics are commonly used in the biochemical field in a variety of scenarios, such as in the study of sugar-receptor interactions or as probes for *in vivo* imaging to assess drug distribution, by employing Nuclear Magnetic Resonance (NMR) or Positron Emission Tomography (PET). In this scenario, fluorinated carbohydrates are of central importance [27–30]. Typically, and particularly in the context of this dissertation, fluoro-sugars are glycomimetics in which exocyclic OH groups are substituted by F atoms, in a *deoxy-fluorination* process. Other F-substitutions have been described as well, such as those

involving the endocyclic oxygen -giving rise to fluoro-carbasugars-, or di and trifluorination reactions in which C-H groups are also replaced by F [31].

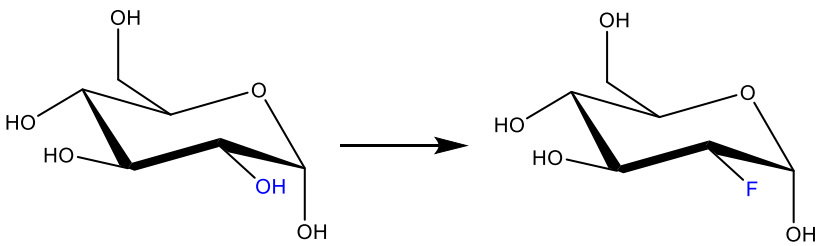


Figure 12. The substitution of the exocyclic OH by F atoms give rise to deoxy-fluoro sugars. In this example, 2-deoxy-2-fluoro-α-D-glucose (right) is the fluorinated derivative at position C2 of α-D-glucose.

Even though O and F are relatively analogous in terms of size and electronegativity, the C-F bond is stronger, shorter, and more polarized than the C-O bond. Importantly, the replacement of a hydroxyl group by fluorine leads to the loss of the hydrogen bond donating capacity and to an evident weakening of the hydrogen bond acceptor ability, due to the higher electronegativity of the F atom. Consequently, OH by F substitutions modulate important physico-chemical properties such as conformation, anomeric equilibria or lipophilicity, which are related with cell permeability of sugars and the way they interact with receptors [32–34].

In general, the shape of the pyranose ring does not seem to suffer important distortions with respect to that of the natural sugars, especially at the chair level [35,36]. As mentioned in the previous section, the substituent at C4 may importantly influence the rotamer distribution around the C5-C6 bond, but in the case of the F substitution, similar populations are in fact observed with respect to the natural monosaccharides [11,31,37]. Regarding the anomeric equilibria, it has been observed that deoxyfluorination cause a slight preference for axial anomers in glucose and galactose derivatives, which might be explained by reduced 1,3-diaxial repulsion as a consequence of the stronger electron withdrawing capability of F in comparison with that of OH [38]. Lipophilicity of carbohydrates is also increased with fluorination, to an extent that depends on the number of fluorine substituents introduced, their position and the stereochemistry of the particular sugar [35,39]. Lipophilicity is essential in the context of medicinal chemistry, since it is related with membrane permeability –higher lipophilicity correlates with better permeability-, as well as affinity of the interaction with carbohydrate receptors, since it influences the desolvation energy contribution to the total binding energy.

Along the following chapters, the work that has been developed using different fluorinated-sugars, mainly as probes for NMR studies, will be presented. The enormous advantages of employing ^{19}F-NMR spectroscopy in certain scenarios rather than ^{1}H-NMR will be discussed as well.

2. LECTINS

Lectins are a subclass of proteins that reversibly interact with sugars in a non-covalent but high-specific way. They do not possess catalytic activity and, unlike antibodies, they are not products of an immune response. Lectins are mainly located at the cell surface or as soluble proteins, both at the intra- and extracellular region. Structurally, they are frequently oligomeric proteins containing two or more carbohydrate binding sites that belong to a more general carbohydrate recognition domain (CRD), which may exist as an individual entity as well. This polyvalence enables lectins to participate in multivalent interactions with carbohydrates of the cell surface, leading to cross-linking of cells and precipitation, a process known as *cell agglutination*. Moreover, lectins can also induce cross-linking between glycoproteins or other glycoconjugates, causing their precipitation as well [1,40,41].

Lectins are ubiquitous in nature. They belong to different protein families that broadly diverge in size, structure, and in their binding site composition, although they show amino acid sequence similarities and common structural features. They were first discovered in plants at the end of the 19th century[8] and in animals during the early 20th, although it was not until 1960s that they attracted the attention of the research community when they were shown to be invaluable tools in the study of complex carbohydrates in solution, especially in physiological and pathological cellular processes involving carbohydrates. In fact, lectins are mediators in numerous molecular recognition processes, including cell-cell adhesion, host-pathogen interactions, serum-glycoprotein control, malignancy and metastasis, and innate immune responses [2,41,42].

Apart from the trivial classification of lectins by the organism on which they are found, lectins can be further grouped according to other convenient criteria, for example, their structural or sequence similarities. Examples of such are the *C-type lectins*, in which the sugar binding is mediated by Ca^{2+} ions, or *galectins*, whose constituent CRDs are folded beta-sheets of about 140 amino acids. One classification that results particularly useful is that made on the basis of their carbohydrate specificity, in particular, on the monosaccharide (or monosaccharides) to which they display the highest affinity. Nevertheless, some lectins show a rather similar specificity for different sugars, so that grouping is merely qualitative in some cases.

[8] It is believed that the first description of a lectin corresponds to Peter Hermann Stillmark, who isolated *ricin* from seeds of *Ricinus communis*, as shown in his doctoral thesis in 1888 (University of Dorpat).

In the following sections, a brief overview of the particularities of the sugar-lectin interactions, as well as a short description of lectins employed in this work, will be exposed.

2.1. SUGAR-LECTIN INTERACTIONS AND MOLECULAR RECOGNITION PROCESSES

In general, lectins bind their preferred carbohydrates with relatively high specificity, but with low affinity (with K_D typically in the range of $10^{-3} - 10^{-6}$ M). Although binding affinity may increase with the size of the sugar, the extent of this improvement still lies in the micromolar range. To overcome this hurdle, multivalency is one of the main strategies followed by nature. Thus, the ligand may display many copies of the same binding-epitope, or the interaction could take place through multiple and simultaneous protein-ligand contacts [2,43].

The overall binding affinity (K_a) is determined by the Gibbs free energy (G) change of the binding process, and depends on the subtle balance between *enthalpic* (H) and *entropic* (S) contributions from the molecules in the free and bound states:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

$$\Delta G = -RT \ln K_a \quad (2)$$

where R is the universal gas constant and T is the temperature.

On the enthalpic side, the interaction between proteins and sugars at the atomic level results from hydrogen bonding, pure electrostatic forces, and hydrophobic contacts, in which the solvent molecules –that solvate the sugar and lectin- play a crucial role as well (Figure 13, Figure **14**). It is common to find polar charged amino acids (such as Arg, Asp or Glu) and cations (for instance, Ca^{2+} in C-type lectins) in the carbohydrate binding sites of lectins. Although there are some glycans that contain charged residues or substituents, most carbohydrates are neutral molecules that mostly interacts with lectins through complex networks of hydrogen bonds (and hydrophobic contacts, including CH/π interactions, to a lesser extent). These interactions are the cornerstones of the stereospecificity of a lectin towards a particular sugar over others [44–46].

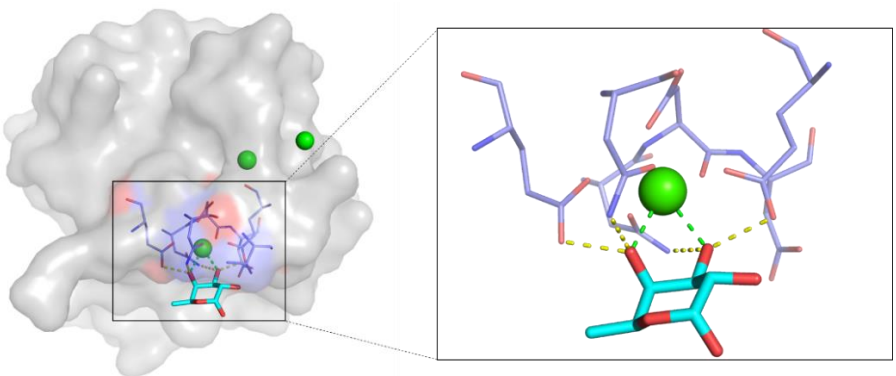


Figure 13. DC-SIGN lectin bound to α-L-Fuc (from the crystal structure deposited with PDB ID 1SL5; only the Ca^{2+} interacting residue of the ligand is shown). The hydrogen bond (yellow dashed lines) network at the binding site determines specificity of lectins towards particular carbohydrates. The polar interaction between Ca^{2+} and OH groups is shown as green dashed lines.

On the other hand, entropic factors include the (most commonly) unfavorable configurational and/or conformational entropy lost by the ligand/protein upon binding, and the desolvation/reorganization of solvent molecules both solvating the ligand and receptor, a process that may or may not be favourable (Figure 14) [47,48]. Therefore, when evaluating the binding affinity it is crucial to consider the role of the solvent. In fact, X-ray crystallographic structures and molecular simulations of proteins in the free state have shown that water molecules generally occupy equivalent positions to those taken by the hydroxyl groups of the bound ligand. Hence, qualitatively it is expected that the displacement of water molecules upon ligand binding does not imply a remarkable gain in terms of enthalpy. On the contrary, displacing waters from the binding site to the bulk might have associated important entropic changes [49].

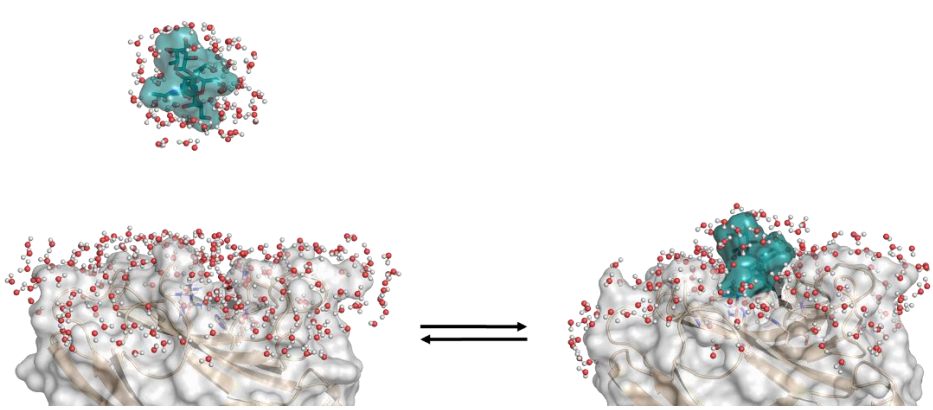


Figure 14. Solvation plays a central role in both enthalpy and entropy of binding.

Another important factor to take into account when considering binding affinity is the surface of the receptor accessible to the solvent. Indeed, solvent exposed binding sites makes easier for solvent molecules to displace the ligand in comparison with buried active sites, which implies faster off-rates (i.e, short residence time of the ligand at the active site) and lower affinities. Moreover, due to the differences in the dielectric constant between the inner and outer areas of the protein, hydrogen bonds are enthalpically more valuable in protein cavities. Altogether, these features translate into higher affinities for lectins displaying buried CRDs [50] (Figure 15).

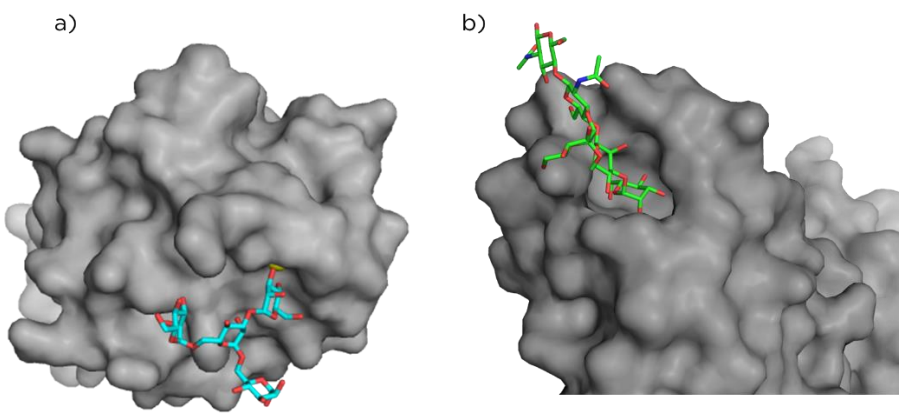


Figure 15. The human lectin DC-SIGN (*a)*) and the bacterial lectin FimH (*b)*) bound to different high mannose derivatives. FimH has a buried binding pocket, which translates into slower off-rates and much higher affinities for mannose oligosaccharides than most mammal lectins, such as DC-SIGN, that display a shallow and solvent exposed binding site.

2.2. SELECTED LECTINS STUDIED THROUGHOUT THIS THESIS

Chapters 4 and *5* of this Thesis deal with the human lectin Dendritic Cell-Specific Intercellular adhesion molecule 3-Grabbing Non-integrin (DC-SIGN) and its interaction with Man derivatives. *Chapter 3* describes the development and evaluation of a screening and chemical-mapping strategy that allows deducing the binding profile of lectins in solution. In that work, several lectins from different sources have been employed: Pisum Sativum Aglutinin from pea seeds, Glucose and Galactose Binding Protein (GGBP) found in bacteria, and the human Macrophage Galactose-type Lectin (MGL, HML). Although they will be described more extensively in the corresponding chapters, a short description of each is presented herein.

2.2.1. Pisum Sativum lectin

Pisum Sativum Lectin (also known as Pisum Sativum Agglutinin (PSA) is a member of the legume lectin family, found mostly in pea seeds. Structurally, PSA is a dimeric protein with a molecular weight of approximately 50 kDa. The monomeric units are identical and each contains one carbohydrate-binding site. In addition, two different metal-binding sites are present in each of them, one occupied by a calcium ion and the other by a transition-metal, usually manganese. The Ca^{2+} is located near the sugar recognition spot and helps into pulling the involved amino acids in the proper positions for sugar binding, whereas the second metal seems to contribute to the organization of the previous one. Importantly, neither of these cations participate directly in sugar binding [51,52].

PSA shows specificity for mannose and glucose entities, and even though it is able to agglutinate human erythrocytes, it is not specific for any blood group antigen. It also presents mitogenic activity, similarly to Concanavalin A lectin [53–55].

2.2.2. Human Macrophage Lectin

The human Macrophage Galactose-type Lectin (MGL), also dubbed Human Macrophage Lectin (HML), belongs to the type II transmembrane C-type Lectin Receptors (CLR) family, which recognizes carbohydrates in a Ca^{2+}-dependent manner. These CLR are a type of pattern recognition receptors expressed by antigen presenting cells, such as macrophages or dendritic cells, capable of identifying sugar type structures of pathogens and host tissues. MGL contains a single extracellular CRD and is galactose-specific. In particular, interacts with terminal Gal/GalNAc moieties such as the Tn-Antigen (α-GalNAc-Ser/Thr), a known epitope for being expressed in aberrantly glycosylated Mucin 1 as well as in adenocarcinoma cells. Therefore, it may also play an important role in tumor immunity [56–60].

2.2.3. Glucose and Galactose Binding Protein

The Glucose/Galactose-Binding Protein (GGBP) is a member of the sugar-binding protein subclass named Periplasmic Binding Proteins (PBP). These proteins, present in the periplasmic space between the inner and outer membrane of gram-negative bacteria, recognize a number of small molecules and ions. They act as intermediary receptors in the ATP-binding cassette (ABC) transport system and chemotaxis, crucial for the bacterial survival [61,62].

Structurally, GGBP is an α/β protein[9] composed by two-fold domains joined by three peptide segments. It has been shown that it exists in open and closed forms, being the latter associated with the bound state. Indeed, ligand binding provoke a classical hinge-bending motion that finally leads to a global change in the orientation of the domains, which in turn causes the equilibrium to move from the open to the closed form through a conformational selection process [63–65]. Unlike other lectins, it shows specificity towards both glucose and galactose with rather higher affinity (K_D around 0.2-0.4 μM).

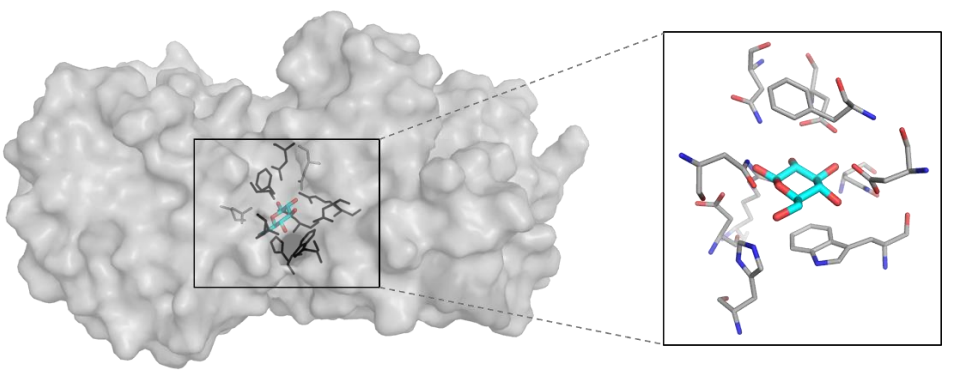


Figure 16. X-Ray crystallographic structure of GGBP/β-D-Glc complex (PDB: 2FVY). The binding affinity of Glc and Gal towards GGBP is rather high compared to other sugar-lectin systems. This has to do with the somewhat buried location of the binding pocket, and the high number of interactions taking place with the binding site residues.

2.3. DC-SIGN

The Dendritic Cell-Specific Intercellular adhesion molecule 3-Grabbing Non-integrin (DC-SIGN), is a transmembrane receptor belonging to the same class of proteins than MGL and expressed mainly in dendritic cells [66,67]. Its involvement in a wide range of pathologies related to the immune system, such as infections caused by virus or bacteria, cancer and autoimmune diseases, is now well established. Actually, it has attracted the attention of the research community for the last twenty years as it is a key piece in the infection by the Human Immunodeficiency Virus (HIV) [68]. DC-SIGN binds HIV with high affinity through a glycoprotein present on the envelope of the virus, gp120, which displays abundant N-linked high-mannose oligosaccharides. Furthermore, it has been shown its participation

[9] Classification of a protein by their folding pattern, composed of alternating α-helices and β-strands.

in other pathogen-related diseases such as Ebola, dengue, or hepatitis C viruses.

Each DC-SIGN monomer displays an extracellular domain (ECD) distributed into a CRD, a neck region, and a small cytosolic tail. The neck region is involved in the assembly of other three lectin monomers, giving rise to a characteristic tetrameric structure [69]. DC-SIGN is less specific than most lectins, although it preferentially recognizes fucose and mannose derivatives with low affinities [67,70,71]. Indeed, DC-SIGN is a prototypical example of a 'promiscuous' lectin.

NMR in the study of ligand-receptor interactions

1. INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy [72] is a central technique in the investigation of the structure, dynamics, kinetics and molecular interactions of a broad variety of biochemical systems. Since it was applied for the first time to address problems related with structural biology, NMR has experienced remarkable developments in many technological aspects (instrumental electronics, hardware, magnets and probe capacities) and in the available methodologies, importantly extending its applicability.

NMR is well established nowadays as one of the most powerful tools for understanding biological processes at the atomic level. Depending on the time scale of the particular process under study, different NMR experiments may be applicable. The intrinsic characteristics of the technique makes it particularly useful to analyze the structure and dynamical properties of complex molecules, such as carbohydrates, in solution. Among all the different NMR parameters that can be exploited for such purpose, relaxation represents a fundamental part. Furthermore, a plethora of experiments have been developed to study small-molecule interactions with receptors, from both the point of view of the ligand and the macromolecule.

Unlike other biophysical techniques, such as Surface Plasmon Resonance (SPR) or Isothermal Titration Calorimetry (ITC) [73–75], NMR spectroscopy is extremely versatile not only due to the numerous approaches that provides for studying many different aspects of biomolecular interactions, but also because of its applicability over a wide range of binding affinities and the complexity of the systems that can be analyzed [76–79]. Although X-ray crystallography is particularly powerful for the determination of protein structure and can typically yield a direct, very accurate picture of the binding mode [80], it lacks the flexibility to study interactions in physiological conditions, test different ligands/receptors simultaneously, provide rapid feedback about interaction features, not to mention the analysis of dynamical properties such as conformational changes or the dissociation constant of the ligand-receptor complex [76].

In the end, the combination of the flexibility and dynamic outlook provided by NMR spectroscopy with the accurate atomic-disposition information obtained from X-ray crystallographic structures, constitute an extremely powerful team to tackle the study of very difficult problems.

2. FUNDAMENTALS OF NMR

From a quantum mechanical perspective, an atomic nucleus with nuclear *spin angular momentum* quantum number I (which can be an integer or half-integer), in an external magnetic field $\boldsymbol{B_0}$ (which, in general, it is arbitrarily assigned to be applied in the $+z$ direction), has associated $2I+1$ non-degenerate energy levels, characterized by the *magnetic quantum number* m (which can take values in the interval $(-I, I)$ in steps of 1). If the nucleus is isolated (i.e., neglecting chemical shift and J-coupling effects, which will be discussed shortly after), the energy associated to each level is:

$$E = -\hbar\gamma B_0 m = \hbar\omega_0 m$$
$$\omega_0 = -\gamma B_0 \qquad (3)$$

where $\hbar$ is the reduced plank constant[10], γ is an intrinsic property of each nucleus known as the *gyromagnetic ratio* (which may have either sign), B_0 is the magnetic field strength, and ω_0 is the so called *Larmor frequency,* arguably one of the central concepts in NMR spectroscopy [81,82] (Figure 17). Equation 3 states that:

When a nucleus with $I \neq 0$ is placed in an external magnetic field, an interaction between its spin angular moment and the field occurs. The separation of the nuclear spin levels is named *nuclear Zeeman splitting*.

From here on, the discussion will be focused on nuclei with $I = \frac{1}{2}$ (abbreviated *spin one-half*) since these are the most exploited in NMR and employed in the work described herein (^{1}H, ^{19}F, ^{13}C or ^{15}N). For spin one-half nuclei, 2 energy levels are predicted from Equation 5:

$$E_\alpha = \frac{1}{2}\hbar\omega_0$$
$$E_\beta = -\frac{1}{2}\hbar\omega_0 \qquad (4)$$

since m can be $-\frac{1}{2}$ or $\frac{1}{2}$. Typically, these energy states are labelled as α ($m = \frac{1}{2}$) and β ($m = -\frac{1}{2}$). The transition between α and β has an associated energy given by:

$$\Delta E_{\alpha\to\beta} = E_\beta - E_\alpha = -\hbar\omega_0 \qquad (5)$$

This transition can be associated to a peak in the NMR spectrum at the Larmor frequency ω_0 (Figure 17, *b)*).

[10] $\hbar = \frac{h}{2\pi}$

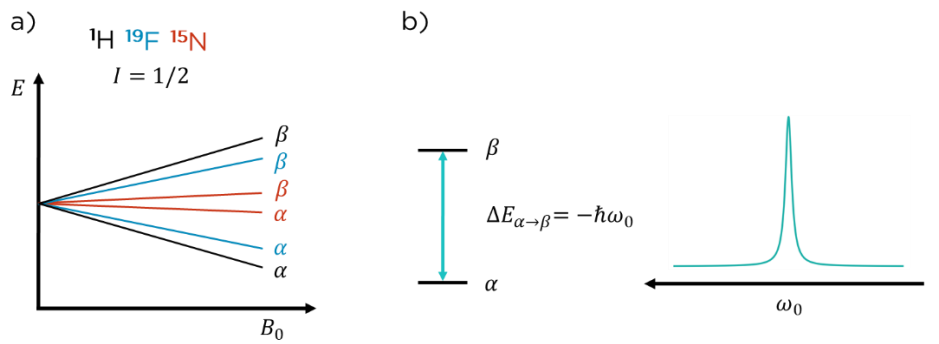


Figure 17. a) Energy levels arising from the interaction of the nuclear spin angular momentum of different one-half nuclei (^{1}H in black, ^{19}F in blue, and ^{15}N in red) with an external magnetic field. The difference in energy between the α and β states is proportional to γ for each nucleus. b) A transition between states can be associated with a line in the NMR spectrum.

The presence of an external magnetic field also affects the electronic density surrounding the nucleus, since electrons possess spin angular momentum as well. Electrons respond to this perturbation producing small induced local fields that subtract from or add to the external magnetic field. Consequently, the nucleus experiences a slightly different magnetic field, and hence a different Larmor frequency. This contribution from the electronic chemical environment of the nucleus is known as *chemical shift*, and provides important local molecular information about the observed nucleus.

When two or more nuclei are connected through chemical bonds, they might interact between each other. This phenomenon is known as *scalar* or *J coupling*, and is characterized by the *coupling constant*, J, which is independent of the strength of the magnetic field. The scalar coupling entails an increment on the available energy states for this multiple-spin system in comparison with those shown on Figure 17 for one spin. In turn, other new possible transitions between states arise, which leads to the appearance of splitting patters in the NMR spectrum. Since these patters depend on the specific features of the interaction between the spins (for instance, the chemical bonds that separates the nuclei or their relative orientation in space), the scalar coupling is a key NMR parameter to deduce bond connectivity in molecules.

Although the complete understanding of NMR spectroscopy relies on the laws of quantum mechanics, it is worthwhile to make use of the classical, more intuitive picture provided by the *vector model*[11]. Therefore, the

[11] The vector model is a very useful and simple description of NRM phenomena. It relies on classical physics and is only fully applicable to non-interacting one-half spins, but for such systems the results are identical to those derived from quantum mechanics.

perspective from both approaches will be employed herein to discuss the NMR phenomena presented when possible. According to the vector model, the interaction energy between a nuclear magnetic moment $\boldsymbol{\mu}$ and a uniform magnetic field $\boldsymbol{B_0}$ is determined by their scalar product:

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B_0} \qquad 6$$

As it can be observed in Equation 6, the minimum energy situation occurs when the magnetic moment of the nucleus is perfectly aligned with -and in the same direction than- the field, whereas the exact opposite case corresponds to the highest energy state (Figure 18).

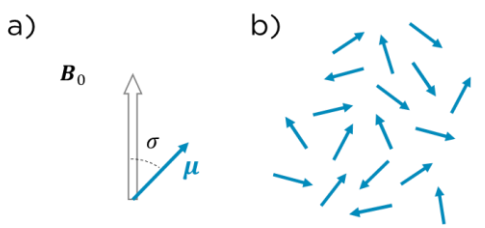


Figure 18. a) The interaction of a uniform magnetic field and the magnetic moment of the nuclear spin. b) The nuclear magnetic moments are randomly distributed in a sample at thermal equilibrium in the absence of an external magnetic field.

However, in a typical macroscopic sample, such isolated nuclear spin situation does never occur. Rather, a collection of millions of magnetic moments at *thermal equilibrium* is found. In the absence of an external magnetic field, the magnetic moments of all nuclei are *isotropically* distributed (since there is not any preferential direction for them to adopt, Figure 18, *b)*).

When the sample is placed in a magnetic field, an interaction energy of the field with the nuclear magnetic moments arises.

This energy, however, is much lower than the thermal energy, and it is precisely the random thermal motion of the molecules which prevents all spins to adopt the magnetically preferred configuration. Still, in time, and averaged over the whole sample, a slight preference of the magnetic moments to align with the field exists, which give rise to a net magnetization or *bulk magnetization* or *equilibrium magnetization*. This magnetization of the sample as a whole can be described by a *bulk magnetization vector* $\boldsymbol{M}$, which is stationary and aligned parallel to $\boldsymbol{B_0}$ (Figure 19).

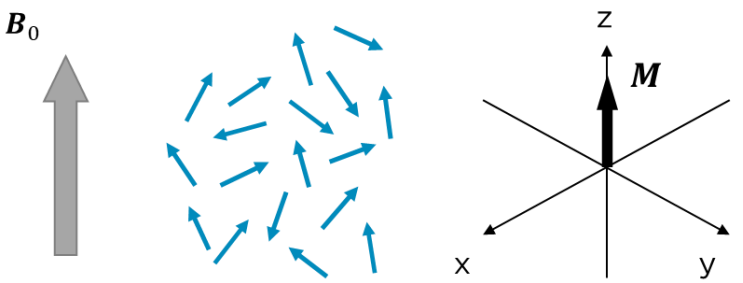


Figure 19. In the presence of $\boldsymbol{B_0}$, a small preference of the nuclear spin magnetic moments (shown in blue) to align with the field arises. This leads, on average, to a net magnetization of the sample, represented by the bulk magnetization vector $\boldsymbol{M}$. The vector model deals with the behavior of $\boldsymbol{M}$ to describe the NMR phenomena.

It can be proved that the magnitude[12] of $\boldsymbol{M}$ is proportional to the difference in population between the two energy levels predicted in Equation 4. At thermal equilibrium it is given by

$$M_{eq} = \frac{1}{2}\hbar\gamma\Delta n_{eq} = \frac{\hbar^2\gamma^2 N B_0}{4k_B T} \qquad 7$$

where k_B is the Boltzmann's constant, N the number of individual nuclei, T the temperature, and Δn_{eq} represents the difference in the α and β state populations, which can be computed using the Boltzmann distribution.

M_{eq} represents the maximum magnetization achievable at equilibrium (also referred to as *polarization*) for a particular nucleus, and thus it is related with the *sensitivity* of that nucleus to the NMR experiment.

In practice, the ratio $\frac{\Delta n_{eq}}{N}$ is of the order of 10^{-5} for ^{1}H at 298 K, which means that NMR spectroscopy is intrinsically a low-sensitivity technique. Importantly, since the net magnetization M_{eq} shows a quadratic correlation with γ, nuclei with higher gyromagnetic constants are more sensitive to the effects of external magnetic fields, and therefore, their use is appropriate for NMR spectroscopy. This is, for instance, the case of ^{1}H or ^{19}F.

This slight net magnetization along the z-axis, however, is still not measurable. In order to generate an *observable* NMR signal, $\boldsymbol{M}$ must be perturbed somehow and brought into the transverse xy-plane, producing *transversal magnetization* since the spectrometer detector (a coil wound) is placed there (Figure 20). A brief description of a usual NMR experiment is given in the following lines.

[12] I.e., the norm of $\boldsymbol{M}$.

By applying an oscillating transverse magnetic field at (or close to) the Larmor frequency, called *radiofrequency (RF) field* or *RF pulse* and typically labeled $\boldsymbol{B_1}$, it is possible to move $\boldsymbol{M}$ away from the z-axis. Such field is said to be *on-resonance* with the Larmor frequency of the nuclei. Depending on the duration the $\boldsymbol{B_1}$ is applied, the magnetization vector can be rotated an angle θ (**Figure 20**). The transversal magnetization is maximum at $\theta = 90º$, and in such case its magnitude is equal to $\boldsymbol{M}$.

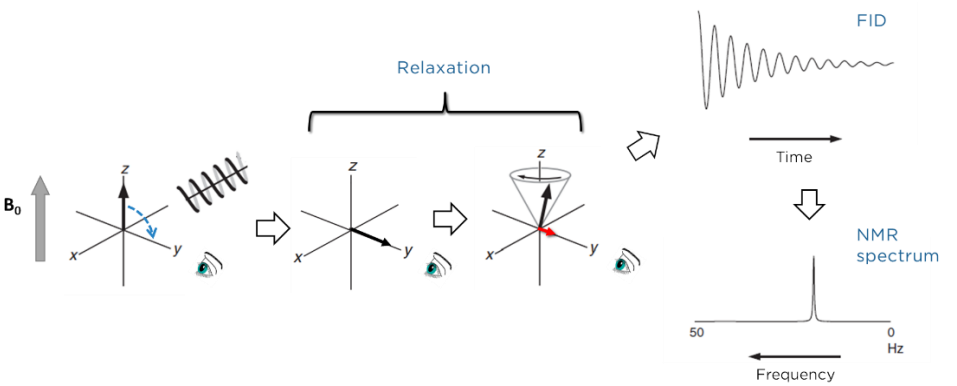


Figure 20. Schematic view of the NMR experiment. First, the bulk magnetization vector is displaced to the transverse plane by the effect of a RF-pulse oscillating at the Larmor frequency of the nucleus. The bulk magnetization comes back to its equilibrium state by precessing about the external field at the Larmor frequency. The free induction decay signal is recorded by the detector (the 'eye' in the figure) and fourier transformed to yield a line in the NMR spectrum at the Larmor frequency

Once the bulk magnetization vector has been tilted from its equilibrium position, it starts rotating about the direction of the magnetic field $\boldsymbol{B_0}$, in a motion called *precession*. It turns out that the precession frequency of $\boldsymbol{M}$ is precisely the Larmor frequency ω_0. The oscillating magnetization vector induces a current in the detection coil, which is subsequently amplified and recorded. This signal is known as the *free induction decay* (FID). Afterwards, the FID, which is a time-domain signal, is submitted to a mathematical procedure named Fourier Transform (FT), to yield the common frequency-domain NMR spectrum (Figure 20)[13]. This whole process -from the application of the RF-pulse to the signal acquisition- constitutes a *scan*, and it is repeated a *number-of-scans* times so that the signal-to-noise ratio increases [81,83].

[13] Notice that the complete picture for applications in chemistry, such as chemical shifts and coupling constants, can only be fully described using a quantum mechaniccal treatment, outside the scope of this Thesis.

As it will be shown in the following section, from the moment it is perturbed, the bulk magnetization progressively returns to its equilibrium state in a process named *relaxation* (this is the reason why the free induction signal actually decay), which is central in NMR spectroscopy.

2.1. Relaxation in NMR

In the context of NMR, relaxation is described as the phenomena through which the bulk magnetization returns to its equilibrium value after the perturbation caused by a RF-pulse. Following with the simplified view provided by the vector model, at equilibrium $\boldsymbol{M}$ is aligned with the external field, so that only z-magnetization exists. Mathematically, the components of $\boldsymbol{M}$ in this scenario can be expressed as:

$$\begin{aligned} M_z &= M_{eq} \\ M_x &= 0 \\ M_y &= 0 \end{aligned} \qquad 8$$

When a RF-pulse close to the Larmor frequency of the observed nuclei is applied, $\boldsymbol{M}$ is tilted away from its equilibrium position. In that case, the situation is the following:

$$\begin{aligned} M_z &< M_{eq} \\ M_{xy} &\neq 0 \end{aligned} \qquad 9$$

Obviously, the components of the magnetization vector changes with time as a consequence of the precession of $\boldsymbol{M}$ about the $\boldsymbol{B_0}$. From the moment it is perturbed, the net magnetization tries to recover its equilibrium value, i.e., to return to the situation of Equation 8. The relaxation phenomenon can be defined in terms of two different relaxation processes:

The *longitudinal* relaxation depict the return of z-magnetization ($\boldsymbol{M_z}$) to its equilibrium value, a process that can be described as an exponential decay characterized by a constant named *longitudinal relaxation time* $\boldsymbol{T_1}$. On the other hand, the *transverse* relaxation refers to the decay of transverse magnetization (M_x, M_y) to zero, which is approximately exponential and can be described by the *transverse relaxation time* $\boldsymbol{T_2}$.

The equations describing the time evolution of the components of the magnetization vector are

$$M_z(t) = [M_z(0) - M_{eq}]e^{-t/T_1} + M_{eq} \qquad 10$$

$$M_{xy}(t) = M_{xy}(0)e^{-t/T_2} \qquad 11$$

where M_{xy} represents the projection of $\boldsymbol{M}$ on the *xy*-plane, i.e., the size of transverse magnetization (Figure 21) [84,85].

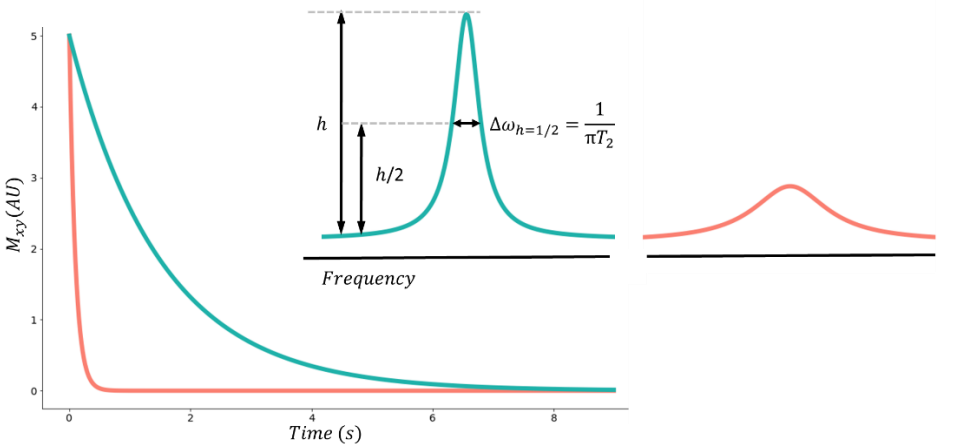


Figure 21. T_2 determines the rate of decay of the M_{xy} magnetization after a RF-pulse. In the simulations on the left, the signal colored in teal is characteristic of a small molecule ($T_2 = 1.5$ s, in this simulation), whereas the one in warm pink is characteristic of a large macromolecule ($T_2 = 0.1$ s). The corresponding FTs of these decaying signals are represented on the right. T_2 is also related with the linewidth at half height ($\Delta\omega_{h=1/2}$) of the observed NMR peak.

Microscopically, for relaxation to occur it is necessary the presence of oscillating *local magnetic fields* that interact with individual spins, producing changes in the orientation and magnitude of their magnetic moments.

In particular, the components of their magnetic moments that are perpendicular to the transverse components of the local field will be changed. Importantly, the local fields must be oscillating at or near the Larmor frequency. From a quantum-mechanical perspective, this situation corresponds to transitions between the allowed spin states of the individual nuclei.

Local fields may arise from different sources, but the two most important mechanisms contributing to relaxation for spin-half nuclei are the through-space *dipole-dipole* (DD) or *dipolar* mechanism, and the *chemical shift anisotropy* (CSA) mechanism [81,82,85]. In the case of transverse relaxation, an additional contribution to relaxation exists, and is brought about local fields not necessarily oscillating at the Larmor frequency. This distinctive characteristic of transverse relaxation turns out to be key in the application of T_2-based methods to the detection of interactions between small ligands and macromolecules, as it will be presented later on.

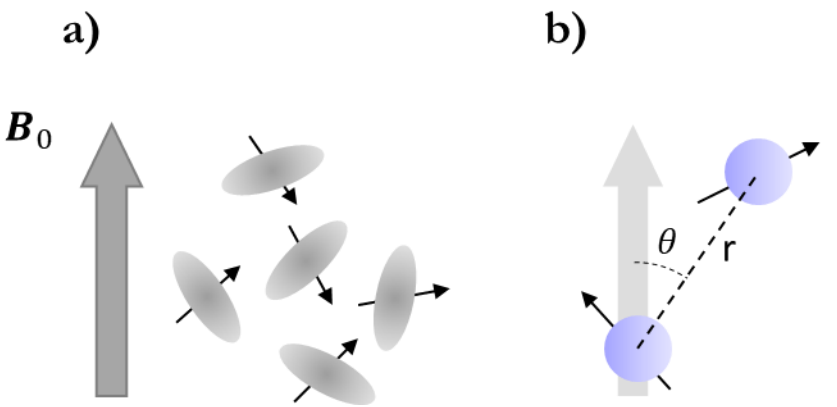


Figure 22. Local magnetic fields lead to relaxation. For one-half nuclear spins, these are mainly generated by: a) CSA of the electron clouds and b) through space dipole-dipole interactions. The DD interaction depends on the distance between the involved nuclei, their gyromagnetic constant and the angle between the direction of the external field and the vector connecting the two spins

In the dipolar mechanism, the local field is produced by one nuclear spin, propagated through space, and experienced by another nucleus nearby (Figure 22, b)). The local field depends linearly on the gyromagnetic ratio of each spin, the orientation between the vector joining the nuclei and the external magnetic field, and is strongly distance dependent ($1/r^3$, where r is the distance between the two nuclei). In practice it is only effective at, say, up to distances of 5 Å for two ^{1}H nuclei. Importantly, this mechanism does not implicate the electron clouds of the atoms.

The chemical shift anisotropy refers to the different chemical shift experienced by a nucleus depending on its orientation with respect to the external magnetic field[14] (Figure 22, a)). Since the chemical shift is a consequence of minor local fields produced by electrons around the nucleus, which varies in direction and size as the molecule tumbles in solution, CSA is also a source of relaxation. Therefore, the more anisotropic is the electron density around a nucleus, the more effective is the CSA mechanism. This is the case of ^{19}F, and indeed it is one of the main reasons that makes it such an interesting nucleus to study ligand-receptor interactions, as it will be presented in the following sections. CSA relaxation is proportional to the square of the applied magnetic field.

As before mentioned, the local fields required for relaxation arising from DD and CSA mechanisms must be varying continuously in time, which is brought about by thermal motion. A molecule in solution experiences

[14] In solution, the rapid tumbling of molecules produce an effective (macroscopic) averaged chemical shift, called isotropic shift.

collisions with others, and gradually is rotated in small steps, in a characteristic motion named *rotational diffusion*. This motion, in turn, is described by the time it takes on average for a molecule to modify its orientation about 1 radian from its starting position, known as the *correlation time* τ_c. Ultimately, relaxation occurs because rotational diffusion cause motion of averaged frequency ω, in the proper range to cause relaxation. It can be proved that:

The rapid motion of small molecules in solution is less effective in causing transverse relaxation than the motion of large macromolecules. This means that bigger molecules (such as proteins) exhibit smaller transverse relaxation times.

The reason has to do with the fraction of total motion that is present at the proper frequency to cause relaxation, and is provided by the *spectral density function*[15]:

$$J(\omega) = \overline{B_{loc}^2} \cdot \frac{2\tau_c}{1 + \omega^2 \tau_c^2} \qquad 12$$

where $\overline{B_{loc}^2}$ is the average of the square of the local fields.

It is possible now to describe the relaxation rates R_i (the inverse of the corresponding relaxation times T_i, where $i = 1,2$) as the sum of the contributions from DD ($R_{i,DD}$) and CSA ($R_{i,CSA}$) mechanisms[16], and as functions of $J(\omega)$:

$$\begin{aligned} R_1 &= R_{1,DD} + R_{1,CSA} = a \cdot J(\omega_0) \\ R_2 &= R_{2,DD} + R_{2,CSA} = b \cdot J(\omega_0) + c \cdot J(0) \end{aligned} \qquad 13$$

being ω_0 the Larmor frequency, and *a*, *b* and *c* constants that include the gyromagnetic ratio of the particular nucleus. Figure 23 shows the evolution of R_1 and R_2 with τ_c. As the correlation time increases, the longitudinal relaxation rate reaches a maximum when $\tau_c = 1/\omega_0$. On the contrary, the transverse relaxation rate increases permanently, which is a consequence of its dependence on $J(0)$ (see Appendix A.I).

[15] This is the simplest example of a spectral density function, assuming spherical molecules in viscous medium that undergo isotropic tumbling.

[16] Assuming that DD and CSA mechanisms are uncorrelated.

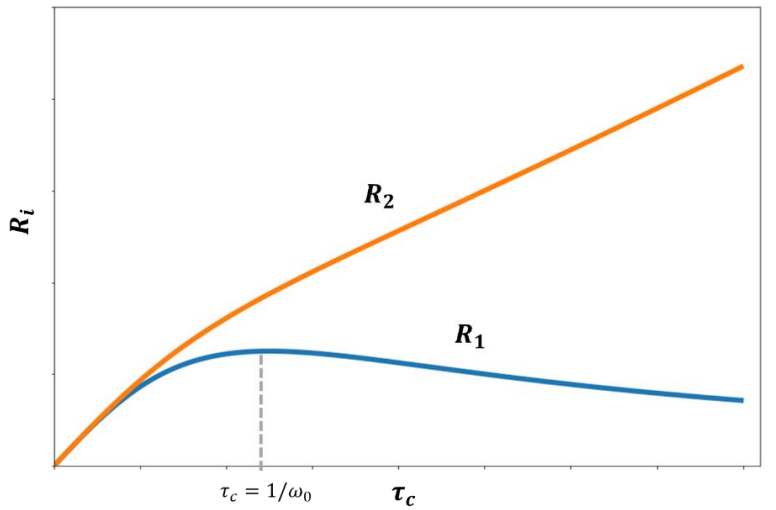


Figure 23. Dependence of R_1 and R_2 relaxation rates with the molecular correlation time. R_1 reaches a maximum at $\tau_c = 1/\omega_0$, while R_2 is a monotonically increasing function.

It is common in NMR to distinguish between two different motional regimes. Motion is considered *fast* on the relaxation NMR time scale when $\omega_0\tau_c \ll 1$, and *slow* when $\omega_0\tau_c \gg 1$. The former is the case observed for small molecules, because these usually exhibit fast molecular motion with low τ_c; the latter is associated with large molecules, with slower tumbling and higher τ_c values. As it will be shown in the following sections, these differences can be exploited to study the interactions of small molecules with larger biomolecules by a variety of NMR methods [86].

2.2. Longitudinal relaxation of two nearby nuclei. The Nuclear Overhauser Effect (NOE)

In this section, the previous discussion of relaxation via the dipole-dipole mechanism is extended to include a key concept involving the longitudinal relaxation of two close spins: *cross-relaxation*. Cross-relaxation refers to the through-space transfer of z-magnetization from one spin to another nearby, by means of different relaxation pathways [81,84,87].

The cross-relaxation is responsible of the so called *Nuclear Overhauser Effect (NOE)*[17], which can be defined as the change in the intensity of resonances from one spin, when the transitions of a sufficiently close spin are perturbed from equilibrium.

[17] This effect was firstly observed in metals as the polarization of the nuclei when irradiating their electron spin resonances [198].

As shown in previous sections, the z-magnetization at equilibrium is proportional to the population difference between the α and β energy levels, according to the Boltzmann distribution. Thus, it is convenient at this point to discuss the longitudinal relaxation in terms of the possible occurring spin transitions in a simple isolated two-spin system coupled through space.

Quantum mechanics shows that for a system comprised by two spin one-half nuclei, sufficiently close between each other so that they interact through dipolar coupling, four possible energy states arise. In the diagram shown in Figure 24, L and S are the two nuclei, and the states α and β are the spin states corresponding to $m = +1/2$ and $-1/2$ respectively, as shown before. Let us consider now an ensemble of these nuclei L and S. According to the Boltzmann distribution, the $\alpha\alpha$ state will be slightly more populated than the $\beta\beta$. For simplicity, let us assume as well that the $\alpha\beta$ and $\beta\alpha$ states are isoenergetic, so that their populations are the same at thermal equilibrium (Figure 24).

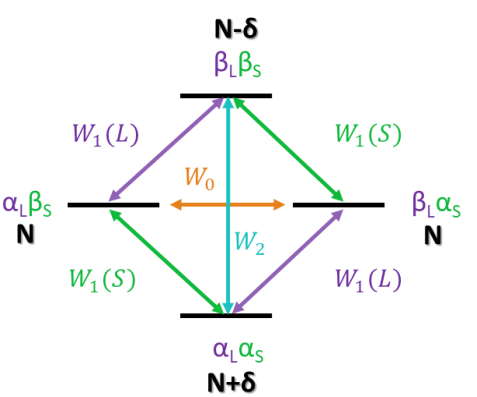


Figure 24. Schematic representation of the energy levels arising from the coupling of a homonuclear two-spin system. The available relaxation pathways are shown in colored arrows, and the population of each energy state at equilibrium is shown in bold.

If the equilibrium population of the system is perturbed somehow, longitudinal relaxation comes into play in an attempt to re-establish the original situation. In this scenario, six possible transitions or relaxation pathways between those states arise for this two-spin system. Each transition has associated a *transition rate* constant or *transition probability*, which describe the rates at which transitions come about. These have been labelled as $W_{\Delta m}$ in Figure 24, where Δm indicates the variation in the magnetic quantum number associated with the particular transition.

W_1 characterizes the already described longitudinal relaxation pathway by which spins either of L or S can flip between their allowed states (while those

of the other partner remains unchanged), known as *single-quantum* (SQ) transition. W_0 and W_2, however, are the transition probabilities associated to the simultaneous flipping of both spin states of L and S, which are referred to as *zero-quantum* (ZQ) (between states $\alpha\beta \leftrightarrow \beta\alpha$) and *double-quantum* (DQ) transitions (between states $\alpha\alpha \leftrightarrow \beta\beta$). Together, these transitions are generally named cross-relaxation pathways, and are responsible of the Nuclear Overhauser Effect (NOE). It is important to notice here that, according to quantum mechanical selection rules, only states with $\Delta m = \pm 1$ are 'allowed', so that ZQ and DQ transitions are 'forbidden'. In practice, this implies that these transitions cannot be directly observed (i.e., they are not associated with a signal in the NMR spectrum), although they can be manipulated indirectly. On the contrary, SQ transitions are 'allowed' and in fact these are the ones that produce peaks, at the associated frequency of the transition, in the NMR spectrum.

Deviations from the equilibrium population difference involving the SQ transitions, as a consequence of ZQ and/or DQ processes, will translate into changes in the height of the peak in the NMR spectrum, which is the experimental manifestation of the NOE.

The system can be perturbed by selectively equilibrating the populations of, say, S, using a sufficiently long interval of low-power on-resonance RF irradiation -a process known as *saturation*-. The consequence of this saturation is that the net z-magnetization for S will vanish, since the populations of its α and β energy levels will be equal. As a consequence, the SQ transitions of S do not participate in the relaxation process (Figure 25).

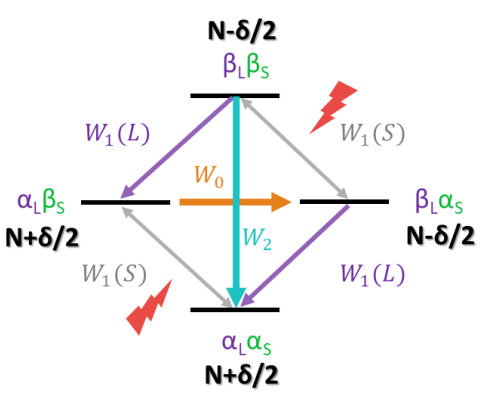


Figure 25. When the SQ transitions of S are saturated, the recovery of the equilibrium magnetization can be driven by W_0 and W_2 processes, giving rise to the NOE.

The same would in principle apply for those SQ transitions of L in the shown diagram, since the difference in population of the involved energy

levels remains constant (Figure 25). However, from the moment the system is perturbed, longitudinal relaxation by dipolar (and other) mechanisms modifies the initial populations and hence relaxation via W_1 occurs.

On the other hand, spins from the $\alpha\alpha$ state towards the $\beta\beta$ state will flow by the W_2 proccess, to re-establish the equilibrium population differences on the S energy states. This, in turn, will increase the imbalance in the population difference across the L transitions. In practice, the W_2 process results in an overall increase of the L spin signal intensities in the spectrum; i.e., a *positive NOE*. Similarly, via the W_0 process spins will be transferred from the $\beta\alpha$ to the $\alpha\beta$ state, trying to equilibrate the equilibrium populations of the S energy levels. This time, a decrease in the population difference over the two L transitions occurs, and thus the W_0 pathway leads to a net drop of the L spin resonance intensities in the spectrum; i.e., a *negative* NOE.

In the end, the subtle balance between the three processes competing between each other (the zero-, single- and double-quantum pathways) will determined the sign of the observed NOE, in the event it is developed (since, for instance, if $W_1(L)$ is very efficient, then the NOE may not be developed at all). When the system is perturbed as described herein[18], the competition of relaxation pathways can be succinctly gathered using the *Solomon equations* to yield the NOE enhancement:

$$\eta_L\{S\} = \frac{\gamma_S}{\gamma_L}\left[\frac{W_2 - W_0}{W_0 + 2W_1(I) + W_2}\right] = \left[\frac{\sigma_{LS}}{\rho_{LS}}\right] \qquad 14$$

where γ_S and γ_L are the gyromagnetic ratios of the S and L nuclei respectively. σ_{LS} is the cross-relaxation constant for the two spins, and its sign determines the overall NOE effect observed.

As it was shown before, the longitudinal relaxation rate depends on the spectral density at the Larmor frequency $J(\omega_0)$. Thus, it is not surprising that the transition probabilities W depends on ω_0 as well. Specifically, it can be proved that

$$\begin{aligned} W_1 &\propto J(\omega_L) \\ W_0 &\propto J(\omega_L - \omega_S) \\ W_2 &\propto J(\omega_L + \omega_S) \end{aligned} \qquad 15$$

where ω_L and ω_S are the Larmor frequencies of the L and S nuclei respectively.

The implications of Equation 15 are very important to understand the NOE effects observed for molecules of different sizes.

[18] This is known as *steady-state* NOE experiment.

Small molecules, which tumble rapidly in solution and has low correlation times τ_c, typically exhibit a wide distribution of tumbling frequencies $J(\omega)$. This means that small molecules may use DQ transitions for cross-relaxation more efficiently, and hence $\sigma_{LS} > 0$ giving rise to a positive NOE. In contrast, high molecular weight molecules tumble more slowly, displaying longer τ_c, and so the lower frequencies are especially populated. Therefore, large molecules will benefit from ZQ transitions for cross-relaxation more efficiently. Thus, in general for large macromolecules, $\sigma_{LS} < 0$ and the resulting NOE will be negative.

Cross relaxation and NOE effects for the different tumbling regimes of small and macromolecules represent the foundation behind the popular Saturation Transfer Difference (STD) NMR experiment, which will be discussed extensively afterwards.

2.3. ^{19}F-NMR

In the last decades, ^{19}F-NMR has emerged as an invaluable tool in a wide range of scenarios involving the study of biomolecular systems. From the NMR point of view, the fluorine nucleus presents very attractive magnetic properties with inherently appealing characteristics. Its active isotope, ^{19}F, has a natural abundance of 100% and is a spin-half nucleus as ^{1}H. This means that the analysis of ^{19}F-NMR spectra is straightforward, as the same rules than for ^{1}H apply. Importantly, its gyromagnetic ratio is relatively high (83% that of γ_H), meaning that it is almost as sensible as ^{1}H for NMR purposes[19]. Moreover, the absence of the strong NMR signal from water media in ^{19}F-NMR spectra, avoid problems associated with the dynamic range of the experiments and possible overlapping of ligand signals. Finally, ^{19}F exhibits a very large chemical shift range in comparison with ^{1}H (and even compared to ^{13}C), thus being a very sensitive reporter of even minute changes in its electronic local environment.

^{19}F-NMR has been extensively employed to address the study of biological systems. Although it has been applied to investigate structural features, conformation and function of proteins [88–90] and nucleic acids [91–93], even in living cells [94] (*in-cell* NMR) from the point of view of the receptor, it is among the NMR ligand-based approaches where it has resulted to be more fruitful [95–100], with a especial relevance in the drug-discovery field. Remarkably, ^{19}F-NMR is arguably one of the most potent tools for screening mixtures of compounds against a target receptor [29,101–103].

[19] The gyromagnetic ratio has also important consequences for dipolar relaxation, as discussed before.

The reasons why the ^{19}F nucleus is an excellent probe for ligand-observed NMR experiments are numerous. To begin with, ^{19}F-NMR spectra do not suffer from the presence of background signals neither coming from receptors nor ligand molecules that otherwise difficult spectra interpretation. In this sense, the use of ^{1}H-NMR limits to a few the number of molecules that can be employed at the same time for the binding experiments, which is a critical disadvantage for screening purposes. Furthermore, ^{19}F detection also overcome the need of using deuterated solvents and buffers that could affect protein stability and solubility, let alone the cost of the highly deuterated reagents. Finally, the large CSA and exchange contribution exhibited by fluorine, which will be discussed in detail later on, makes it an extremely powerful reporter of binding events using different NMR strategies [104,105].

3. BINDING EQUILIBRIA FROM THE NMR POINT OF VIEW

At this point, it is important to dedicate a few lines to address the equations and features that rules binding equilibria. In the simplest situation, a one-site binding process between a ligand and a receptor can be described as:

$$L + P \rightleftharpoons LP \qquad 16$$

in which L and P are respectively the ligand and receptor species in the free state, and LP is the their bound form. The rate constant of the complex formation, k_{on}, is related with the probability of having a productive encounter between ligand and receptor, i.e., an encounter that leads to formation of LP. The dissociation of LP to yield L and P is accounted by the off rate of the process, k_{off}. These rate constants are related through the temperature-dependent equilibrium dissociation constant (or its inverse equilibrium association constant, K_a):

$$K_D = \frac{1}{K_a} = \frac{[L][P]}{[LP]} = \frac{k_{off}}{k_{on}} \qquad 17$$

where the brackets indicate concentration of the species. It might be useful to define as well the total ligand concentration ($[L]_T$) in solution (either free or bound), and the fractions of free (p_f) and bound (p_b) ligand in equilibrium:

$$[L]_T = [L] + [LP]$$
$$p_f = \frac{[L]}{[L]_T}$$
$$p_b = \frac{[LP]}{[L]_T} \qquad 18$$

The interpretation of binding experiments by NMR requires a closer view to the *chemical exchange* process. Chemical exchange refers to the change in the chemical environment of a nucleus when it is found in different states, and can be described by an *exchange rate* constant k_{ex} (Appendix A.II) [86]. Indeed, this is the case for L and P nuclei since they exchange between the bound and free state in solution. When the ligand binds the receptor, it adopts the macromolecule NMR characteristics for the time the complex is formed. Likewise, the receptor's binding site is perturbed by the presence of the ligand in the bound state.

For a given NMR parameter Q (either from the ligand or the receptor perspective), it is possible to distinguish between three general cases: if $k_{ex} \gg \Delta Q$, the exchange is considered *fast* on the NMR timescale; if $k_{ex} \ll \Delta Q$, the exchange is considered *slow*; and when $k_{ex} \approx \Delta Q$ the exchange is referred to as *intermediate* ($\Delta Q = |Q_B - Q_F|$; Q_B and Q_F are the parameters of the bound and free state, respectively).

Let us illustrate this with a practical example. Let Q be the Larmor frequency, and consider an ensemble of uncoupled spins of the same type giving rise to a NMR signal characteristic of the free state, and a different signal characteristic of the bound state. For simplicity, let us consider that the population of the two states are equal. The fast exchange regime translate into a new signal appearing at the frequency corresponding to the weighted-average frequencies of the characteristic signals of the free and bound states. When the molecule is in slow exchange, two different signals appear, each corresponding to a different state (free and bound). In the case of intermediate exchange, a situation in between the other two takes place, and the signals are typically broader with lower intensity (Figure 26) [86,106,107].

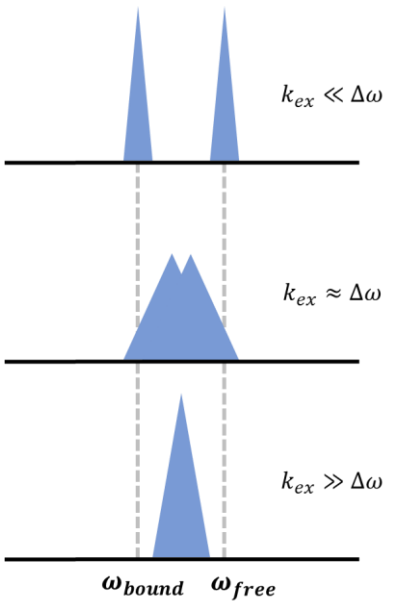


Figure 26. Schematic representation of the exchange effect on an ensemble of similar spins between two equally populated states.

4. NMR TECHNIQUES TO STUDY LIGAND-RECEPTOR INTERACTIONS

NMR spectroscopy is extraordinarily valuable for studying interactions of biomolecules with ligands in physiological conditions [108–119]. It allows monitoring ligand-receptor[20] interactions by tracking changes in NMR observables (chemical shifts, relaxation, etc.) of any of the involved partners (ligand or receptor), induced by the binding event (Figure 27). Many NMR experiments and methodologies have been proposed to exploit those differences [86,107,120–132], and can be broadly classified into receptor-based or ligand-based approaches depending on the observed species (Table 1). In fact, typically a combination of different sets of experiments is employed to get a complete picture of the protein-ligand interaction.

[20] In the context of the work presented in this dissertation, the *receptor* or *target* molecules are lectins, and *ligands* are carbohydrates.

Table 1. Summary of common NMR strategies to study ligand-receptor interactions

Observation	Disturbed NMR parameter upon binding	NMR measure	What to look for	Main use(s)
RECEPTOR	Chemical shifts	Chemical-Shift Perturbation (CSP)	Variations in chemical-shifts of the receptor resonances upon binding. There might be variations in signals shape or intensity too.	Detection of binders. Mapping receptor binding sites. SAR by NMR. Affinity measurements.
	Dipolar couplings	Residual Dipolar Coupling (RDC)	Differences in RDCs in the free and bound state	Bound state orientation of ligands
LIGAND/RECEPTOR	Relaxation/NOE	Paramagnetic Relaxation Enhancement (PRE)	Broadening/reduced-intensity of receptor/binder signals as a consequence of paramagnetic relaxation effects	Detection of binders
LIGAND	Relaxation/NOE	Saturation Transfer Difference (STD)	Changes in signal intensities of the ligand, due to receptor-ligand NOE effects when the receptor is irradiated.	Detection of weak binders. Epitope mapping. Affinity measurements by competition.
		T_2 relaxation	Broadening/reduced-intensity of ligand signals as a consequence of increase in R_2 relaxation upon binding	Detection of binders. Screening. Affinity measurments by competition.
		WaterLOGSY	Changes in signal intensities of the ligand by NOE effects from water molecules	Detection of weak binders
		Transferred NOE/ROE	Differences in NOEs signs of the ligand in the free and bound state. Intermolecular receptor-ligand NOE effects.	Detection of weak binders. Conformation of flexible ligands in the bound state.
	Response to magnetic field gradient pulses	Diffusion Ordered SpectroscopY (DOSY)	Differences in diffusion rates of ligands upon binding	Detection of binders

The most widespread receptor-based method, Chemical Shift Perturbation (CSP), exploits differences in chemical shifts observed in $^{15}N/^{1}H$ (or $^{13}C/^{1}H$) two-dimensional correlation spectra of the target in the presence and absence of a ligand [125]. If previous knowledge of the protein resonance assignment exists (usually this is possible for proteins up to 40 kDa), it permits to locate specific binding sites of the receptor [133]. Moreover, it allows studying both higher and lower affinity ligands. However, the need of over-express and purify high quantities of protein (in the order of mg), finding expression hosts capable of facing isotope enrichment for assignment (^{13}C, ^{15}N, ^{2}H), as well as the stability of the protein samples over the time required to perform the potentially time-consuming NMR experiments, are important factors that may limit the applicability of the approach.

In general, ligand-based methods are devoted to detect ligand-protein interactions even if the binding is (very) weak and transient, by scrutiny of changes in different NMR observables of the ligand upon binding (especially those involving relaxation, as shown in Table 1) [76]. These methods overcome the problem of preparing mg quantities of isotope-labeled protein, requires lower protein concentration (some of them, only in the range of low μM) and can be performed typically on a more reasonable time scale than receptor-based strategies. Therefore, although in principle they do not provide information about receptor binding site(s), they turn out to be very useful as screening methods allowing the study of large mixtures of compounds simultaneously, or to validate binding of other molecules.

Furthermore, they can be applied to unravel ligand binding epitopes and their bound state conformation, which is an especially powerful approach when combined with structural information of the receptor if it is available (for instance, obtained by X-Ray crystallography, NMR or molecular modelling).

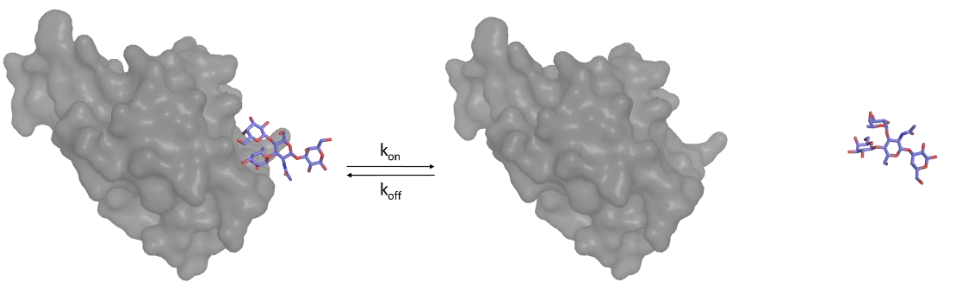


Figure 27. Ligands display different NMR properties in the free and bound states. The ^{1}H, ^{15}N and ^{13}C resonances of aminoacids at (or close to) the binding site are also perturbed by the presence of the ligand in the bound state.

4.1. Ligand-observed ^{19}F-NMR-R$_2$ method to detect ligand-receptor interactions

As discussed before, ^{19}F-NMR is particularly useful for studying mixtures of compounds, among other things, due to the large CSA exhibited by ^{19}F which hugely reduces the likelihood of signal overlapping. But as seen before, CSA can also be a relaxation source. It turns out that, in the case of the ^{19}F nucleus, the CSA mechanism contributes very importantly to the relaxation process. As it will be shown, this can be exploited to detect binding by monitoring changes in the T_2 (or R_2) relaxation time of the ligands when a protein is present in solution.

In previous sections, it was shown that both R_1 and R_2 depends ultimately on the correlation time τ_c of the molecules (Equation 13). However, the dependence of R_2 on the spectral density function at $J(0)$ entails a linear relationship with the correlation time of the molecule (and hence roughly, with its molecular weight). When a ligand binds to the receptor, it mimics its correlation time during the time the interaction last, what provokes an important decrease in its overall transversal relaxation rate. In a NMR spectrum, this entails a broadening of the observed peak for the bound ligand (Figure 21). On the contrary, R_1 does not depends on $J(0)$, and reaches a maximum value when plotted against the correlation time (Figure 23). Hence, it is an insensitive reporter of binding.

Compared to ^{1}H, these effects are much more intense for ^{19}F because ^{1}H nuclei mainly relax through the DD mechanism. However, due to the high anisotropy of the ^{19}F nuclei, besides the dipolar relaxation there exists an

additional contribution from the CSA mechanism (see Appendix A.I) [86,106]. In fact, at the strong magnetic fields used routinely nowadays, the CSA contribution to the ^{19}F nucleus relaxation is the dominant mechanism. As a consequence, the large CSA of ^{19}F results in a very significant difference in linewidth of the fluorine resonance of a molecule between its free and bound states. Hence, the particular features of ^{19}F relaxation make it possible to detect even a minuscule fraction of bound molecules [134].

Although it is possible to monitor the differences in linewidths of fluorine resonances of ligands in the presence and absence of a receptor, in practice it is more convenient to employ an alternative approach: measuring signal intensity evolution via the *Carr-Purcell-Meibom-Gill* (CPMG) scheme [135,136].

CPMG is classically used to measure the transversal relaxation times T_2 of molecules. A graphical representation of how the experiment works is shown in Figure 28. The pulse sequence is as follows: [D-90_x-(τ-180_y-τ)$_n$-acquire], where 90_x and 180_y are RF-pulses applied about the x and y axis[21], D is the *recovery delay* (i.e., the time passed before a new scan starts), τ is the *evolution time* and n indicates the repetitions of the *spin-echo* before signal acquisition. Briefly, the 90º pulse creates transversal magnetization by tilting $\boldsymbol{M}$ to the xy-plane. Then, the magnetization is allowed to evolve during a time τ and then *refocused* by applying an on-resonance 180º degree pulse, after which the magnetization evolves during τ again. Finally, the FID is acquired. The sequence (τ-180_y-τ) is referred to as *spin-echo*, and has the effect of rotating the magnetization vector in the xy-plane[22]. Since T_2 relaxation drives naturally xy-magnetization to its equilibrium value of zero, after each repetition of the spin-echo block the measured transversal magnetization will be lower and so it will the intensity of the acquired NMR signal. By recording the FID at increasing time intervals t (i.e., at increasing spin-echo repetitions, n), a set of intensity-time data can be obtained and then fitted to the exponential decay curve:

$$I(t) = I_0 e^{-t/T_2} = I_0 e^{-n2\tau/T_2} \qquad 19$$

where $I(t)$ is the intensity at time t and I_0 the intensity at $t = 0$.

[21] The phase shift between the 90º and 180º pulses aims at compensating imperfections in the pulses that otherwise may accumulate during the NMR experiment.

[22] The objective of the spin-echo sequence is to eliminate the effects caused by field inhomogeneities in the estimation of T_2. Microscopically, after the 90º pulse each individual spin evolves at its larmor frequency (ω_0) plus a contribution caused by field inhomogeneities (ω_{inhom}), which tend to disperse the transverse magnetization. The ω_{inhom} term can be reversed by applying a spin-echo sequence.

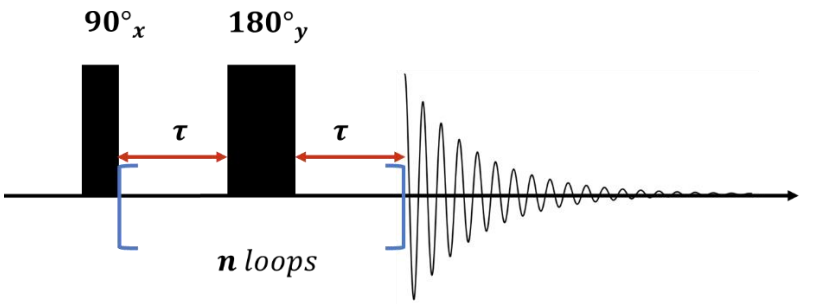


Figure 28. Schematic representation of the CPMG pulse sequence.

In practice, ^{19}F-NMR CPMG experiments (also dubbed ^{19}F-NMR-R_2 or *relaxation filtered*) for monitoring ligand-receptor interactions are performed first on a sample of the ligand alone, and then on a sample of both protein and ligand . Differences in T_2 as a consequence of binding translate into a reduction of the signals for a particular number of spin-echo repetitions n [101]. Thus, it is possible to detect binding simply by comparing peak heights in the presence and absence of receptors at different n. Alternatively, the full decay of transversal magnetization can be tracked for both samples, which allows the quantitative calculation of the transversal relaxation times of the ligand nuclei in the free state, $T_{2,f}$ and their observed relaxation times in the presence of the receptor, $T_{2,obs}$ [137,138].

In the case of weak affinity binders, the observed transversal relaxation rate in the presence of the protein is given by [86]:

$$R_{2.obs} = p_f R_{2,f} + p_b R_{2,b} + R_{ex} \qquad 20$$

where $R_{2,f}$ and $R_{2,b}$ are the transversal relaxation rates of the ligand in the free and bound states respectively. p_f and p_b are the fraction of free and bound ligand, as defined in Equation 18. The last term in Equation 20 is usually named *exchange contribution*, and arises from the non-equivalence between free and bound ligand chemical shifts [104,139]:

$$R_{ex} = p_b {p_f}^2 \tau_{res} 4\pi^2 (\delta_f - \delta_b)^2 \left(1 - \frac{p_f \tanh(\frac{\tau}{p_f \tau_{res}})}{\tau/\tau_{res}} \right) \qquad 21$$

where τ_{res} is the ligand residence time, δ_f and δ_b are the chemical shifts of the ligand in the free and bound states, respectively, and τ is the evolution time of the CPMG pulse sequence.

Equation 20 shows that binding can be detected if the averaged relaxation rate is dominated by the term $p_b R_{2,b}$, which implies a broadening of the observed signal as a consequence of the decrease in τ_c. Due to the fact that

typically $R_{2,b} \gg R_{2,f}$, it is feasible to detect binders even under the typical experimental conditions of $p_b \ll p_f$ for weak binders. Interestingly, R_{ex} depends on the strength of the applied magnetic field, meaning that the effects of the binding will be even more pronounced at higher fields. Moreover, R_{ex}, and hence $R_{2.obs}$, can be modulated by controlling the pulsing rate $1/(2\tau)$ of the CPMG experiment. In the limit of fast exchange ($\tau_{res} \to 0$), R_{ex} becomes zero. In any case, it can be manipulated to control the broadening effect.

The sensitivity and simplicity of this methodology explains why it is regarded as one of the most powerful tools for the screening of small molecules against a larger receptors, broadly used for High-Throughput Screening (HTS) [126].

4.2. STD-NMR

When low or medium molecular-weight ligand molecules bind to macromolecules, such as proteins, the observed ligand NOEs may undergo strong changes (for the reasons exposed in previous sections). This transferred NOE effect is the origin on which the STD-NMR experiment is based [115,140]. STD-NMR is one of the most employed ligand-based techniques for the detection and characterization of ligand-receptor interactions of weak binders. It has been extensively applied to study a variety of biomolecular systems, including interactions between sugars and lectins [141–146], ligands with diverse proteins [114,147–150], nucleic acids [151–153] and viruses [154–156].

The following discussion will focus on ^{1}H-STD experiments, since ^{1}H is by far the most employed nucleus in this kind of experiment. Subsequently, the particularities of STD experiments based on ^{19}F will be discussed as well.

In general, the samples studied by STD-NMR experiments comprises a solution containing a ligand (or small library of compounds) and a target receptor, typically in a 1:20 to 1:100 receptor/ligand molecular ratio. In its simplest form, the STD experiment consists on observing differences in the intensity of ligand resonances in the NMR spectrum recorded when the receptor is selectively irradiated *(on-resonance* spectrum), and in another spectrum recorded without protein saturation (*off-resonance* spectrum). Then, subtraction of these two spectra is performed, so that, in principle, only the signals of those nuclei from the ligand that are close enough to the protein (≤ 5 A) will remain in the *difference* spectrum. On the contrary, resonances of other ligand ^{1}H's far from the protein or other species that do not bind the receptor, if present, will vanish (Figure 29) [140,157].

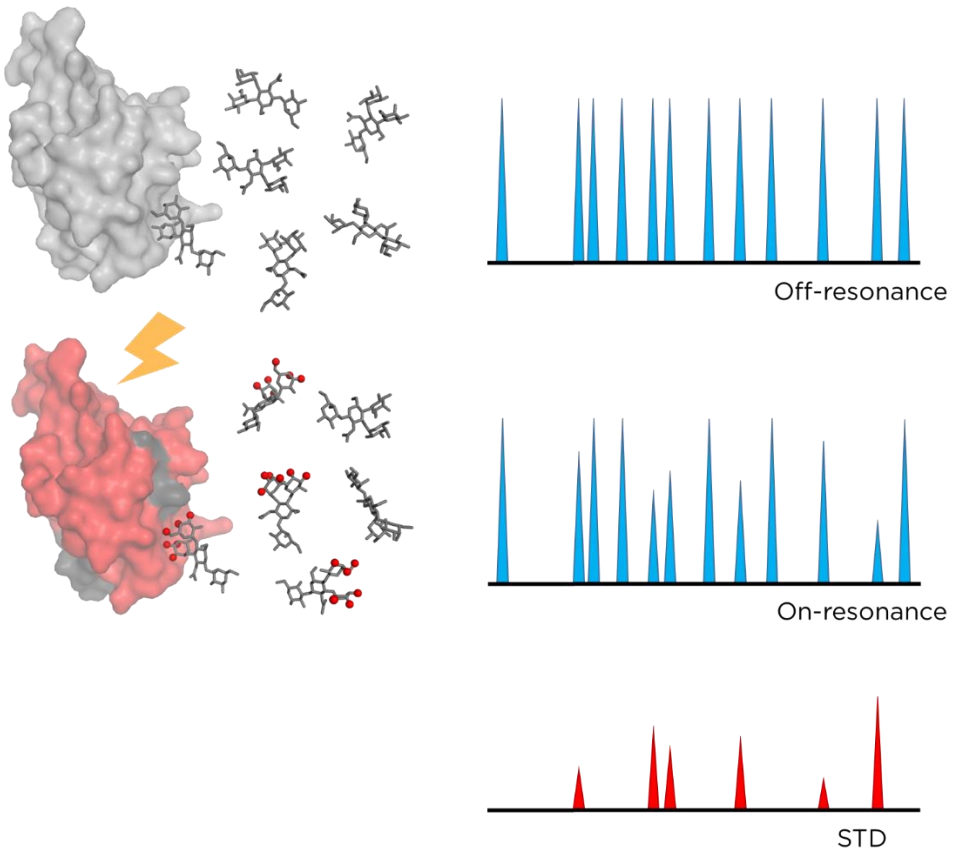


Figure 29. Representation of the STD NMR experiment. The off-resonance experiment, without protein saturation (upper panel), gives rise to a spectrum with unperturbed signal intensities, I_{off}. When the protein is selectively saturated in the on-resonance experiment (middle panel), the magnetization is transferred to the protons of the ligand closer to the receptor (represented as red spheres) by dipolar relaxation. The saturated ligands are accumulated in solution, and the on-resonance spectrum shows a decrease in the intensity of those proton signals, I_{on}. The difference spectrum (bottom) reveals only the signals corresponding to the protons that received saturation from the protein, with intensity I_{STD}.

STD values are usually specified as the fractional intensity change between the on- and off-resonance spectra for each proton of the ligand:

$$\%\,STD = \frac{I_{off} - I_{on}}{I_{off}} \cdot 100 = \frac{I_{STD}}{I_{off}} \cdot 100 \qquad 22$$

where I_{off} and I_{on} are the intensity of a particular ligand signal in the off- and on-resonance spectra respectively, and I_{STD} is the absolute intensity change.

In the on-resonance experiment, only protein resonances must be selectively saturated, so it is particularly important to select an area free from

ligand frequencies. Thus, typically the selective irradiation is performed on the aliphatic (< 1 ppm) or aromatic (> 6 ppm) region, depending on the density of ligand signals in those regions. Once some protein protons have been saturated, the magnetization spread rapidly to other protons of the protein via cross-relaxation very efficiently, a process known as *spin diffusion* (as the polarization diffuses over many spins, Figure 30).

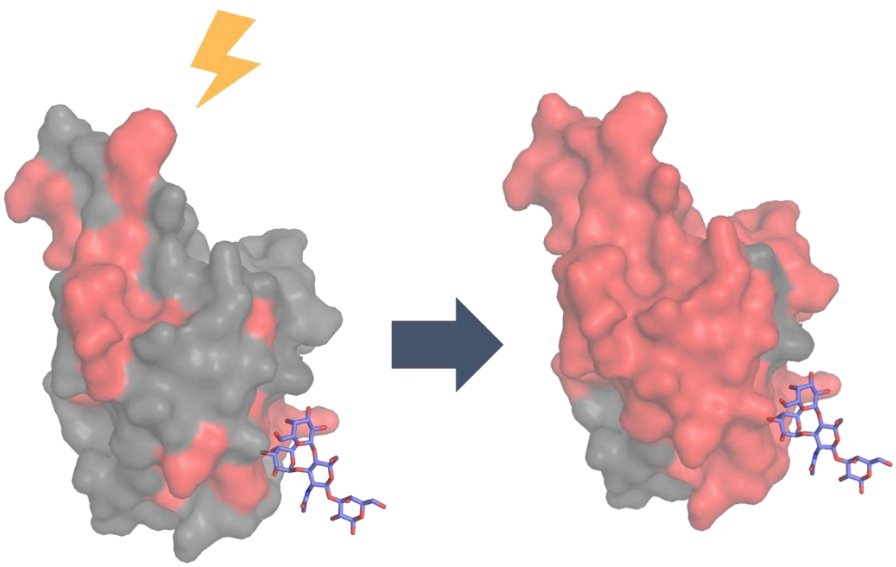


Figure 30. Schematic representation of the spin-diffusion. After selective saturation of some protein resonances (left), the magnetization is quickly spread all over the protein residues (right). The effectiveness of the spin diffusion is roughly proportional to the molecular weight of the receptor.

When a ligand binds to the protein, the saturation will be transferred preferentially to those protons closest to the receptor. Globally, the degree of saturation of the ligand nuclei depends on the residence time ($1/k_{off}$) of the ligand at the binding site. Upon dissociation, the ligand moves back to the bulk solution carrying the saturation information from the protein. These saturated ligand molecules free in solution give rise to resonance signals with thin linewidths (as these are characteristic of the free state), but likely different intensities, in the NMR spectrum.

The underlying idea behind the STD-NMR experiment is that the saturation of the protein and bound ligand occurs much faster than the time it takes for the saturated ligand nuclei to re-establish their equilibrium magnetization via longitudinal relaxation. Consequently, the off-rates must be fast enough so that saturated ligands can accumulate in solution before they lose the transferred magnetization from the protein. This is typically the case for medium/strong binders, with K_D in the low μM or nM range.

However, if the off-rate is extremely fast, the STD effects will be very weak (or negligible) as the probability of the molecules being in the bound state is minimal. Thus, STD-NMR is specially suitable for dissociation constants K_D approximately in the range of 10^{-3}-10^{-8} [140].

Importantly, from the discussion above it is inferred that the intensity of the signals in the difference spectrum are not directly proportional to the binding strength in general. In contrast, STD intensities increase with the saturation time and the ligand concentration, up to a maximum value. Actually, an increment in the ligand/protein ratio implies an increment in the absolute STD signal I_{STD} although the proportion of saturated ligands is reduced [115,124]. In addition, the use of high ligand/protein ratios is convenient since it decreases the likelihood of saturated ligands re-entering the protein binding site, which would preclude otherwise obtaining trustable information about the conformation of the ligand with respect to the protein residues at the binding site, the ligand *binding epitope*.

It is worth mentioning that binding affinity can be measured from STD-NMR spectra by competition titration experiments, as long as the binding constant of the competing molecule to the receptor is known. By monitoring the decrease in the peak height of a STD signal of the ligand upon increasing amounts of the competing molecule, it is possible to obtain its dissociation constant employing a one-site competition model [115,124].

4.3. STD-NMR for determining the binding epitope

As it has been shown, the STD-NMR experiment can be used not only as a method to detect binding, but also to measure ligand binding constants and to determine the binding epitope of a ligand to a target protein. The latter typically requires more thorough experimental designs, such as irradiation at different saturation times or the use of different protein/ligand ratios, to avoid unambiguous interpretations of the ligand STD profile. It is worth noting here that the definition of 'binding epitope' is somehow diffuse. It may refer to the specific parts of a polymeric molecule (for example, a polysaccharide) that are closer to the protein residues upon binding, without specifying exactly their orientation in space. Alternatively, 'binding epitope' may denote in some cases the detailed location of the protons from the ligand with respect to those of the protein at the binding site, and hence, the precise ligand conformation at the bound state that can be deduced in combination with theoretical saturation transfer calculations [115,145,158].

In general, some difficulties inherently associated with the experiment exists, which become more evident as the information intended to be obtained is more precise. For instance, intra-ligand relaxation rates may give rise to an 'apparent' epitope map, which could not be fully representative of

the bound state epitope (this might be particularly pronounced for geminal protons, which can display smaller STDs even if they are close to the protein)[158]. Moreover, the use of longer saturation times, which yield stronger STD signals, has associated other potential issues such as the somewhat complex influence of the free ligand correlation time and conformation on STD intensities [158].

Another issue that can affect the interpretation of STDs for epitope mapping is the residence time of the ligand. If it is too long, spin diffusion may effectively 'dilute' the transferred magnetization over the ligand protons, making the distinction between closer and farther protons from protein residues within the binding site impossible. Similarly, the size of the receptor can influence the observed STD profiles. In general, the bigger the target protein is, the more intense STD signals will be observed (due to the larger difference between ligand and receptor correlation times). Although this can be advisable if the only purpose of the experiment is the detection of binding, it might be an obstacle when trying to deduce subtle details about the ligand conformation at the binding site, since the spreading of the magnetization throughout the ligand would difficult the STD interpretation (this is analogous to the effect observed for too long residence times of the ligand). In this sense, using a smaller portion of the receptor (for example, only the CRD in the case of lectins) can serve both to achieve a somewhat more selective saturation spot in the protein (due to the reduce effectiveness of spin diffusion as a consequence of the lower molecular weight) and a well-defined epitope mapping of the ligand, with substantial differences in the STD intensities of each proton. Additionally, irradiation of the receptor at different frequencies (for instance, at aromatic and aliphatic protons independently) can be exploited as well to obtain further information about the mode of binding [123,149,159].

Fortunately, a meticulous experimental design and analysis can help reducing the uncertainties of the results. For instance, STD with short saturation times do not depend on the free ligand correlation time and conformation, and indeed reflects more precisely the ligand-protein interproton distances at the bound state. In contrast, short saturation times translate into poor STD intensities with low signal to noise (S/N) ratios. To circumvent this dichotomy, it is a recommendable practice to perform a set of STD experiments varying the saturation time of the receptor, and fitting the STD intensities of each proton to the following equation [151]:

$$I_{STD}(t) = {I_{STD}}^{max}(1 - e^{(-k_{sat}*t)}) \qquad 23$$

where ${I_{STD}}^{max}$ is the maximum value of the I_{STD} curve at long saturation times, and k_{sat} represent the experimental saturation rate constant. This

allows extrapolating the STD values to zero saturation time, which represent better the 'true' saturation profile in terms of ligand-protein interproton distances and reduce the likelihood of introducing noise in the observed STD intensities [151]:

$$\left(\frac{d\left(I_{STD}(t)\right)}{dt}\right)_{t=0} = {I_{STD}}^{max} k_{sat} \qquad 24$$

To conclude, the use of computational models of the complex and other tools to support the experimental results and help with their interpretation is advisable, especially in the ligand-based approaches such as STD in which no explicit information of the receptor is obtained. If the structure of the complex or the protein alone is available, the COmplete Relaxation and Conformational Exchange MAtrix (CORCEMA) theory might be a very useful resource to complement the experimental results [158]. For a given model (from X-ray crystallography, NMR, or computational prediction) CORCEMA is able to estimate the absolute or fractional intensity changes of the STD signals for the reversible binding of a ligand and a protein forming a binary complex, thus providing a quantitative measure of the theoretical STD profile of the system under study. Then, this profile can be compared with the experimental STD using an error function, dubbed *R-NOE factor*, to extract conclusions about the suitability of the proposed model to explain the STD results.

4.4. ^{19}F-STD-NMR based experiments

The basic 1D STD-NMR experiment can be combined with other pulse sequences (for example, 2D-TOCSY [140]) to extend the information they provide and circumvent the limitations associated to 1D NMR spectroscopy.

Interestingly, STD experiments has also been adapted in the last decades for ^{19}F detection. In particular, two versions of the experiment have been developed: one with direct heteronuclear magnetization transfer (via dipolar coupling) $^{1}H_{receptor} \rightarrow {}^{19}F_{ligand}$ [160]; and the more sensitive $^{1}H \rightarrow {}^{1}H$-STDreF [160,161], in which the magnetization transfer in the first step is homonuclear $^{1}H_{receptor} \rightarrow {}^{1}H_{ligand}$, followed by an intra-ligand relay to ^{19}F (reF) through scalar $J_{H,F}$ coupling for observation. The explanation behind this gain in sensitivity in the reF experiment is the less effective heteronuclear cross-relaxation $^{1}H \rightarrow {}^{19}F$ with respect to the homonuclear case.

More recently, the evolution of the $^{1}H \rightarrow {}^{1}H$-STDreF lead to the development of the 2D STD-TOCSYreF experiment [37]. This new NMR strategy overcomes the lack of spatial coverage of the previous mono dimensional experiment, limited to the magnetization transfer coming from geminal and trans-vicinal protons to fluorine, where $J_{H,F}$ coupling is large.

Clearly, that limitation was insufficient for epitope mapping. However, in the 2D STD-TOCSYreF version, the introduction of a $^{1}H,^{1}H$-TOCSY building block allows efficiently extend the STD polarization to all protons of the spin system, followed by the reF transfer from geminal and trans-vicinal protons of ^{19}F nuclei.

Molecular Dynamics simulations

1. INTRODUCTION

Computational simulations are an invaluable tool for interpreting experimental data regarding the structure, dynamics and molecular interactions of complex systems. Ideally, these techniques would be able to work the other way around: predict the outcome of experiments beforehand. Although the field has undergone a huge breakthrough within the last two decades, along with the development of computing power and Graphics Processing Units (GPUs) as potent platforms for scientific calculations, universally applicable models are still far from being a reality. In fact, this is not a (big) problem, since different approaches might be employed for distinct situations [162,163].

In general, the size of the system under study and the scientific question that one wishes to answer, determines the methodology that can be applied. For instance, Quantum Mechanical (QM) methods explicitly consider electronic effects on the molecule and make use of different approximations to solve the Schrödinger equation. Accordingly, these methods can be used to study reaction mechanisms, bond formation and cleavage, predict spectroscopic data, or to study molecular structures of excited states, to name a few. Although they provide the most rigorous approaches to investigate molecular systems, they require an enormous computational cost and are in practice only applicable to subsets of a few hundred atoms at most. On the other hand, Molecular Dynamics (MD) simulations rely on molecular mechanics to describe systems. MD simulations neglects the quantum nature of atoms and use classical physic laws to examine the interactions between them. This approximation hugely simplifies the problem and consequently MD methods can be used to study the structural and dynamical evolution of systems comprised by thousands of atoms, explicitly, even at the millisecond time scale [163,164].

Given the inherent structural complexity of carbohydrates, and aiming at better understanding the way they interact with lectins from a dynamic point of view, MD simulations have been chosen as the main computational technique employed herein to support the experimental NMR data, as it will be presented in the following chapters.

2. STATISTICAL MECHANICS

The link between computer simulations –which provide detailed information of a system at atomic level-, and experimental observations, is the field of statistical mechanics. Statistical mechanics establishes relationships between experimentally observed thermodynamic properties of a system, such as pressure or volume, and the positions and momenta of each particle comprising that system.

Let us consider a system containing N atoms. The position $\boldsymbol{r}$ and momenta $\boldsymbol{p}$ of all the atoms of the system define a point in a 6N-dimensional space referred as *phase space*. If we denote $\Gamma(t) = (\boldsymbol{r}(t), \boldsymbol{p}(t))$ as a particular point in the phase space at time t, then the instantaneous value of a property A can be written as $A(\Gamma(t))$. Since the system evolves in time, it is reasonable to assume that the experimental value observed for the property, A_{obs}, should be equal to the time average of A over an (infinite) time interval [165,166]:

$$A_{obs} = \langle A(\Gamma(t)\rangle_{time} = \lim_{t\to\infty} 1/t \int_{t=0}^{t} A(\boldsymbol{r}(t), \boldsymbol{p}(t))\, dt \qquad 25$$

There are several difficulties with this approach. In principle, it would be possible to determine the evolution of $\Gamma(t)$ by applying the Newton's equations of motion (if the forces acting between particles are known) for a simple classical system. However, for the typical number of particles of real systems ($N \approx 10^{23}$), it is not practicable to calculate even a small fraction of the whole trajectory since that would imply the integration of the equations of motion for a huge system during an enormous (infinite) period of time. To circumvent this problem, Maxwell, Boltzmann and Gibbs postulated the *ergodic hypothesis*, which is central in statistical mechanics:

The *time-average* is equal to an *ensemble average* of a large number of replications of the system.

In this regard, an *ensemble* is a (virtual) large collection of systems, say m, each one being a replica on a macroscopic (thermodynamic) level of the particular system of interest. For instance, for an isolated system characterized by having N atoms in a volume V, with total energy E, the ensemble would be formed by mN atoms, in a volume mV with total energy mE. Each replica corresponds to a (potentially) different system configuration (*microstate*) compatible with the fixed (at constant N, V and E, -or NVE in short-) thermodynamic state (*macrostate*). This is known as the *Postulate of Equal a priori Probabilities*:

Any system has an equal probability of being in any microstate consistent with its fixed macrostate.

In other words, the systems forming an ensemble are thermodynamically identical, but not necessarily identical at the atomic level. As before, each system configuration conforming the ensemble can be described as a point $\Gamma(\boldsymbol{r},\boldsymbol{p})$ in phase space. Then, Equation 25 can be replaced by [165,167]:

$$A_{obs} = \langle A(\Gamma)\rangle_{ens} = \iint A(\boldsymbol{r},\boldsymbol{p})\rho(\boldsymbol{r},\boldsymbol{p})_{ens}d\boldsymbol{r}d\boldsymbol{p} \qquad 26$$

where $\langle A(\Gamma)\rangle_{ens}$ represents the average value (or expectation value) of A over all the replicas of the ensemble[23]. The term $\rho(\boldsymbol{r},\boldsymbol{p})$ is the *probability density function* of the ensemble, which is defined by the chosen macroscopic parameters (NVE, NVT, etc.). $\rho(\boldsymbol{r},\boldsymbol{p})$ indicates the probability of finding a member of the ensemble in the interval $(\boldsymbol{r},\boldsymbol{p}) + (\boldsymbol{r},\boldsymbol{p})d\boldsymbol{r}d\boldsymbol{p}$.

One of the most commonly used ensembles in statistical mechanics is the NVT or *canonical* ensemble, in which N, V and T are fixed. It turns out that for this ensemble, the probability density function is the Boltzmann distribution:

$$\rho(\boldsymbol{r},\boldsymbol{p})_{NVT} = \frac{e^{-E(\boldsymbol{r},\boldsymbol{p})/(kT)}}{Q_{NVT}} \qquad 27$$

where k is the Boltzmann constant, T is the temperature, E is the total energy[24] and the normalization factor Q is the *partition function* of the ensemble. The previous equation is a central result in physical-chemistry: the partition function is the bridge between the individual molecular configurations of the system and its thermodynamic properties. In other words, by knowing Q one can basically access to all the information about the macroscopic properties of the system. In the case of the canonical ensemble, Q has the following particular form [167]:

$$Q_{NVT} = \iint e^{-E(\boldsymbol{r},\boldsymbol{p})/(kT)}d\boldsymbol{r}d\boldsymbol{p} \qquad 28$$

Q has different expressions depending on the ensemble of interest (Table 2). However, for systems of the size of Avogadro's number, these differences vanish[25]. This behavior is known as the *thermodynamic limit* and it is universal for any thermodynamic property in any ensemble. As it will become clear soon, this property is important for molecular simulations of large systems,

[23] A double integral sign is used for clarity, but rigorously there are 6N integral signs corresponding to the 3N positions and 3N momenta of the N particle system.

[24] The total energy is the sum of the kinetic and potential energies: $E = K(\boldsymbol{p}) + V(\boldsymbol{r})$

[25] It can be proved that the fractional fluctuation of measurable properties in any ensemble scale roughly as $1/\sqrt{N}$.

since it allows us to choose the ensemble that is mathematically most appropriate to work with, or the one that better mimics the laboratory experimental conditions.

Table 2. Summary of common ensembles

Ensemble	Partition function
Microcanonical (NVE)	$\omega(E)$
Canonical (NVT)	$\iint e^{-E(\boldsymbol{r},\boldsymbol{p})/(kT)}$
Isobaric-Isothermal (NPT)	$\iint e^{-(E(\boldsymbol{r},\boldsymbol{p})+PV(\boldsymbol{r},\boldsymbol{p}))/(kT)}$
Grand Canonical (μVT)	$\iint e^{-(E(\boldsymbol{r},\boldsymbol{p})+\mu N(\boldsymbol{r},\boldsymbol{p}))/(kT)}$

$\omega(E)$ is the degeneracy (number of microstates or configurations) for a particular energy E.

From Equations 27 and 28, it is now possible to estimate the average value for any given observable. The only remaining question is how one can generate replicas of the system to conform an ensemble. As we will see below, this is the goal of molecular simulation methods.

3. A GENERAL VIEW OF MOLECULAR DYNAMICS SIMULATIONS

Molecular simulation methods aim to produce a set of relevant microstates for a particle-based system. There are two main techniques devised to achieve that goal, referred to as *Molecular Dynamics* (MD) and *Monte Carlo* (MC) simulations, and both can be used to calculate relevant properties of the system by averaging over the output microstates. In this dissertation, only MD simulations will be discussed.

The molecular dynamics method produces an ensemble of states by integrating the Newton's equations of motion. Unlike MC, which is a stochastic method, MD is deterministic in the sense that it provides information about the 'time'[26] evolution of the properties of a system. In fact,

[26] Notice that this is not a 'real time' in any physical sense. As stated before, in MD simulations properties are computed from *ensemble averages* instead of *time averages*. This is just an index used to account for the steps of the simulation.

knowing the position and velocities of the particles forming the system at an initial (simulation) time τ_0, it is possible to predict its configuration at any time. The trajectory followed by every particle can be obtained by solving the following equations:

$$-\nabla V(\boldsymbol{r}_i) = \boldsymbol{F}_i \tag{29}$$

$$\frac{\boldsymbol{F}_i}{m_i} = \frac{d^2\boldsymbol{r}_i}{d\tau^2} \tag{30}$$

where $V(\boldsymbol{r}_i)$ and $\boldsymbol{F}_i$ are respectively the potential energy and the force acting over the particle i, m_i is the mass of the particle. In a typical MD simulation, initial positions $\boldsymbol{r}_i(\tau_0)$ and velocities $\boldsymbol{v}_i(\tau_0)$ are assigned to each atom of the system[27]. Then, forces acting over each particle are calculated from a given set of potential energy function and parameters known as *force-field* by applying Equation 29, and Newton's equation of motion are solved to determine how the particles respond to the acceleration produced by the force $\boldsymbol{F}$. To obtain the time evolution of the position, it is necessary to solve the differential Equation 30 by integration using different methods. This integral is computed numerically, so a 'small' time-step $\Delta\tau$ (ideally, the time-step size should be significantly shorter than the fastest motion on the simulation) is chosen. Finally, the particles are moved to their new positions, the simulation time is increased by $\Delta\tau$, and the cycle is repeated again until the total number of MD simulation steps is reached [165] (see Figure 31).

Different algorithms used to perform the integration of the classical equation of motion have been proposed [168], but basically all assume that positions and velocities of the particles can be described as Taylor series expansions:

$$\begin{gathered}\boldsymbol{r}_i(\tau+\Delta\tau) = \boldsymbol{r}_i(\tau) + \boldsymbol{v}_i(\tau)\Delta\tau + \frac{1}{2}\boldsymbol{a}(\tau)\Delta\tau^2 + \frac{1}{6}\boldsymbol{b}(\tau)\Delta\tau^3 \\ + \cdots \\ \boldsymbol{v}_i(\tau+\Delta\tau) = \boldsymbol{v}_i(\tau) + \boldsymbol{a}_i(\tau)\Delta\tau + \frac{1}{2}\boldsymbol{b}(\tau)\Delta\tau^2 + \cdots \\ \boldsymbol{a}_i(\tau+\Delta\tau) = \boldsymbol{a}_i(\tau) + \boldsymbol{b}(\tau)\Delta\tau + \cdots\end{gathered} \tag{31}$$

where the terms $\boldsymbol{v}$, $\boldsymbol{a}$ and $\boldsymbol{b}$ (and the subsequent) are the velocity (first derivative of the position with time), acceleration (second derivative), and the third derivative terms respectively (and so on). The *order* of the Taylor

[27] The initial coordinates of each particle of the system may come from experimental data, such as X-ray crystallographic structures, or from a theoretical model. Initial velocities can be assigned by random selection from a Maxwell-Boltzmann distribution at the required temperature.

expansion is typically truncated after the acceleration term, since higher order terms are difficult to compute and contribute much less in the expansions than those of lower order.

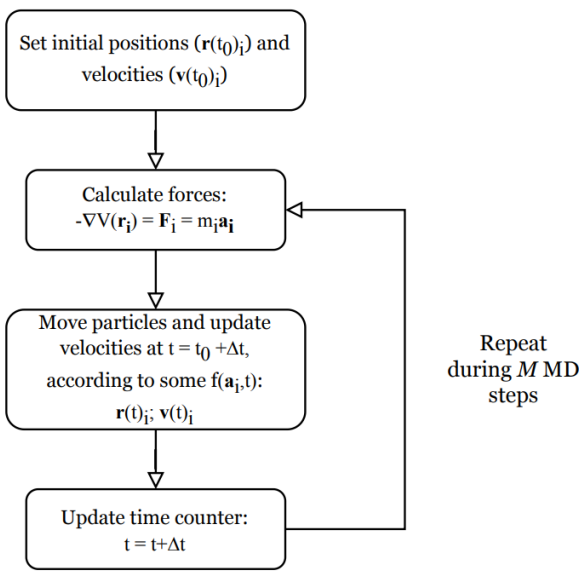


Figure 31. Schematic view of a typical MD simulation program.

This way, a MD simulation generates a *trajectory* that describes the evolution of a system along the simulation time. Specifically, it produces a collection of possible configurations of the system (i.e, a collection of points $\Gamma(r,p)$ in phase space) for a given set of macroscopic variables. If MD simulations were capable of exploring all the points of phase space, then the partition function could be exactly determined by summing up every individual contribution from each microstate (Equation 28). In that case, the system would be ergodic. However, in practice, for a system of the size typically studied by computational simulation methods, the magnitude of the phase space is incommensurable, and thus attempting to obtain an ergodic trajectory to calculate the partition function is virtually infeasible.

The alternative approach followed by MD (or MC) simulations is to visit a representative large enough subset of the most informative microstates (i.e, points in phase space) to approximate the partition function properly. This process of exploration of phase space in a simulation is dubbed *sampling*. Obviously 'good' sampling is an essential requirement to obtain accurate physical properties of the simulated system. It turns out that, for *mechanical properties*[28] of the system (such as pressure or internal energy), the low energy

[28] More specifically, *mechanical properties* are those that depends on the derivative of the partition function. On the other hand, *thermal properties* depends on the partition function itself.

states contribute more importantly to the partition function, and indeed those states are the ones preferentially generated by MD simulations. Thus, these thermodynamic properties are well defined and can be properly predicted in simulations. In contrast, the partition function for *thermal properties* (such as entropy) is also dependent on high energy states and, consequently, their estimation using MD generated configurations will be poor and inaccurate [165,166]. For such calculations, advanced simulation approaches and sampling methods must be employed.

Under the previous assumption, estimated average values of basic thermodynamic properties can be easily obtained from MD trajectories. The link between the theoretical framework presented in the previous section for a continuous ensemble, and the ensemble generated by MD simulations, arises from the discretization of Equation 26:

$$\langle A \rangle_{ens} = \frac{1}{M} \sum_{i=1}^{M} A_i \qquad 32$$

where M is the number of *simulation steps* (Figure 31) –i.e, system configurations- and A_i is the value of the observable A at the i step.

4. FORCE FIELDS

Force-field is the term used to refer to a particular set of potential energy function and its parameters, which describes the interaction between every particle in the system. These functions are built on the basis of the Born-Oppenheimer approximation. Thus the electronic motions are ignored and the energy depends only on the nuclear positions. Therefore, MD simulations use intrinsically classical mechanics to model a wide variety of phenomena, which allows to address the study of systems that are prohibitively large to be evaluated by other methods that explicitly represents electrons in the calculations (*quantum mechanical* methods).

Any force-field typically describes the molecular features as a sum of different energy terms: bond lengths and angles between particles are represented as harmonic potentials, bond rotations or dihedral angles are modelled using periodic functions, and electrostatic and van der Waals interactions (also known as *non-bonded* terms of the functional) are accounted by using Coulomb's law and Lennard-Jones potentials, respectively. A general force field function is shown below:

$$V(\boldsymbol{r^N}) = \sum_{bonds} K_r(r_{ij} - r_{eq})^2 + \sum_{angles} K_\theta(\theta_{ijk} - \theta_{eq})^2 + \sum_{dihedrals} \sum_{n} k_{\varphi,n}[1 + cos(n\varphi_{ijkl} + \delta_n)] + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} [\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}] \quad (33)$$

where $\boldsymbol{r^N}$ refers to the positions of all particles of the system[29]. Briefly, the first summatory term extends to all covalent bonds in pairs of connected atoms ij, the second term involves the bond angle between sets of ijk contiguous atoms, and the third is a sum over all torsional angles defined by sets of $ijkl$ atoms. Lastly, the fourth term, which is the most computationally demanding to calculate, corresponds to the intra and inter-molecular non-bonded interactions [165,166].

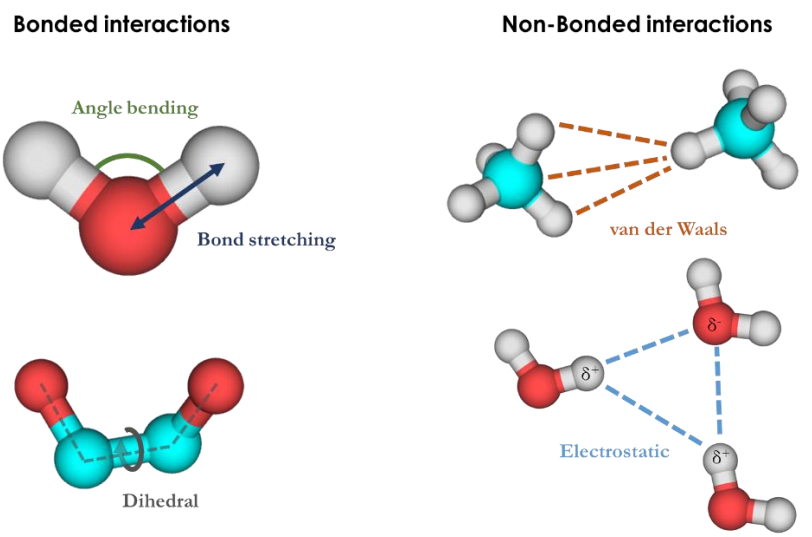


Figure 32. Representation of the 4 basic terms typically found in molecular force fields.

Force-fields differ from each other not only by their functional form but also by the parameters they use (the constants in Equation 33). These functions and parameters are obtained from laboratory experiments and

[29] Notice that we here emphasize that $\boldsymbol{r^N}$ corresponds to the position vector of each particle for clarity. We have previously used $\boldsymbol{r}$ referring to all the particles as well in the previous discussion, without the superscript for simplicity.

from precise quantum mechanical calculations. Force-fields are in principle designed to reproduce the structural features of the molecules, but they are also devised to predict other molecular properties as well, for which they are *parametrized* accordingly. Therefore, force-fields are purely empirical in the sense that there is not any single or 'correct' procedure to develop them, but rather they are built with the aim of properly reproducing molecular features. Ideally, a force field should be able to replicate not only specific experimental features for the molecules used to its parametrization (known as *training set*), but also for other molecules of the same kind (i.e, proteins, DNA, lipids, etc.). The success or failure of a force field in predicting how the particles in the system behave or how well a specific property is characterized, is what ultimately determines their validity for studying a particular system.

There are a variety of force-fields that might be used to model distinct molecular systems in different situations. Some of them are summarized in Table 3. All the MD simulations described in this essay has been performed using the AMBER suite of biomolecular simulation programs [169,170]. The force-fields GAFF2 [171], ff14SB [172] and GLYCAM_06j [173] have been employed to parametrize organic molecules (except non-modified sugars, in general), proteins and carbohydrates.

Table 3. Some examples of force fields employed in computational simulations. The force fields shown here are periodically updated and extended.

Force field	Application
ff14SB	Proteins
GAFF	Small organic molecules
OPLS	Small organic molecules
GLYCAM_06	Carbohydrates
LIPID17	Lipids
CHARMM36	Proteins, DNA, RNA, lipids
AMOEBA	Polarizable FF
MARTINI	Coarse-grained proteins, lipids, carbohydrates, nanoparticles
REF15	Proteins, DNA, RNA, small molecules

5. MD SIMULATIONS IN EXPLICIT SOLVENT

A common problem when performing MD simulations of molecules in explicit solvent, is the treatment of those molecules placed at the boundaries of the container, that necessarily experience different forces from those molecules of the bulk liquid. This problem may be overcome by using a method known as *periodic boundary conditions* (PBC), which aims at representing the properties of a macroscopic system with a much smaller but infinitely periodic one [166].

When PBC are employed in a simulation, the system is placed in a 3D container called 'primary cell' or 'box'. This primary system cell is then replicated infinitely (Figure 34). The simplest cell to repeat in all directions has a cubic form, but other shapes are also used in practice. Whenever a particle moves out of the main cell during a simulation, one of its images enters through the exact opposite face with the same velocity. Therefore, the number of molecules in the unit cell is conserved, and only their coordinates must be computed. However, the forces acting over each particle are calculated considering every other in the infinite array that is relevant to the calculation.

The success of PBC depends importantly on the range of the non-bonded interactions that are accounted for. Since every particle could actually interact with its other images from neighboring boxes, the risk of introducing a certain symmetry in a fluid that should otherwise be isotropic appears. This flaw is easily circumvented by using a sufficiently large box to minimize the Lennard-Jones interactions (since $V(\boldsymbol{r}_i) \propto r^{-6}$), but can have important consequences for the long-range electrostatic potential ($V(\boldsymbol{r}_i) \propto r^{-1}$). (Figure 33). Moreover, the energy evaluation in a system with PBC would require the calculation of an infinite number of interaction terms, arising from the infinite repetition of the original system. Even for a single cell, the calculation of non-bonded interactions scales with N^2, which means that they are computationally very expensive.

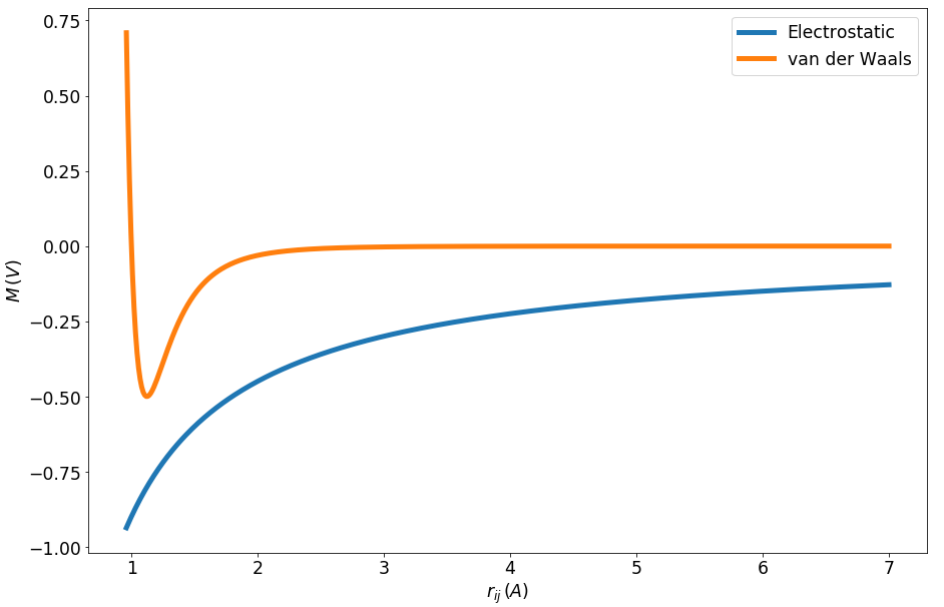


Figure 33. Variation of electrostatic (colored blue) and van der Waals (*VdW*, orange) potential energy functions with the distance between a pair of atoms r_{ij}. The van der Waals term approaches to zero much faster than the electrostatic one.

In order to reduce the cost of the simulations, the two following strategies are typically applied to biomolecular simulations for the non-bonded interactions: the *minimum image convention* and the use of a *non-bonded cutoff*. The minimum image convention considers the interaction of a particular molecular entity with every other molecule of the infinitely repeated system only once. To further reduce the computational cost, a non-bonded cutoff can be applied as well so that only the molecules whose geometric center lies in a specific region around the main molecule are considered in the force calculations (Figure 34). This implies that, when a cubic box is employed for PBC, the size of the cutoff necessarily must not be larger than half the length of the cell to be consistent with the minimum image convention.

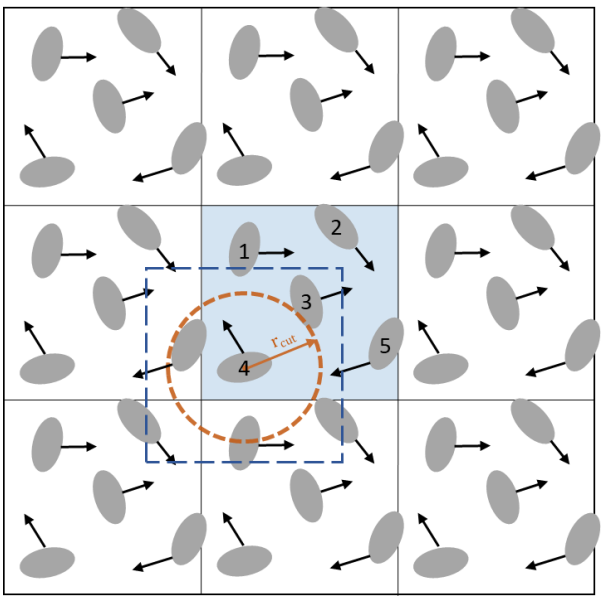


Figure 34. A 2D representation of a subset of an infinite replicated system formed by 5 molecules, using periodic boundary conditions with a cubic cell. For clarity, the primary cell at the center is shown in blue. Linear momenta (black vectors), minimum image convention (dashed blue square) and a cutoff for non-bonded interactions (dashed orange circle or radii r_{cut}) for molecule 4 are represented as well.

Although the use of distance cutoffs is computationally efficient and easy to implement, it entails an important limitation: the potential function finds a discontinuity at the cutoff distance. This, in turn, results in simulations in which the total energy of the system is not conserved. For this reason, the previous approach is especially inadequate for treating long-range forces, as the electrostatic potential (Figure 33). To circumvent this problem, the *Particle Mesh Ewald* (PME) method (which is a particular implementation of a more general procedure known as *Ewald summation* [174]) was proposed by Darden [175,176]. It is now one of the most popular and accurate approximations to estimate the electrostatic energy. While a full description of the procedure is beyond the scope of this Thesis, it essentially consists on splitting the contribution of the total electrostatic potential into two terms: a short-range term calculated in real space, and a long-range contribution computed in the reciprocal space (Fourier space). This approach has the advantage that both terms of the summation converges quickly in their respective spaces, solving the mentioned problem of the potential truncation. The key difference of the PME with respect to the standard Ewald method is that the computation of the long-range part can be carried out using the so-called fast *Fourier transform* (FFT) algorithm (ref), which allows to drastically reduce the computation time. In summary, the methodology allows the calculation of the full electrostatic energy (i.e, the potential is not

truncated at all) with high accuracy and efficiency, as has been proved even for highly charged polar systems [176,177].

6. SIMULATIONS IN THE NVT/NPT ENSEMBLES

In the standard MD simulation procedure described above, the integration of Newton's equation of motion implies that the total energy is conserved. Therefore, the system is isolated and strictly the simulation takes place in the microcanonical NVE ensemble. However, inaccuracies in the numerical integration and the use of the different cutoffs normally translate into the total energy not being conserved and the system being heated. Moreover, it is typically more interesting to run MD simulations in different conditions than NVE. In fact, a much more common practice is to use constant temperature settings, since this is the usual scenario in which laboratory experiments are conducted. For that reason the *canonical* (NVT) or *isobaric-isothermal* (NPT) are typically the most used ensembles for simulation of biomolecules in liquid phase. As mentioned above, while all these ensembles are only rigorously equivalent in the limit of an infinite system, thermodynamic properties calculated from statistical averages in any ensemble are very similar for systems of a few hundred particles.

Some of the methods typically employed to obtain constant temperature or pressure simulations, with special focus on those used in this dissertation, will be discussed in the following lines and in the corresponding *Appendix*.

6.1. Temperature control

Several methods have been developed in order to control the temperature of the system during a simulation, which can be divided into two main categories: *stochastic* and *non-stochastic methods*[30].

Non-stochastic methods simply use *velocity scaling* to update the velocities of each particle in order to be consistent with the user required system temperature (Appendix A.III). One of the main drawbacks of these methods is that, since the only condition imposed to the system is restricting the total kinetic energy to agree with the required temperature, solutes and solvent can indeed have different velocities still fulfilling that condition. This phenomenon of having an inhomogeneous distribution of temperature throughout a heterogeneous system leads to the 'hot-solvent, cold-solute' [178,179] or the 'flying-ice-cube' [180] problems frequently observed in the literature. Furthermore, they do not sample from any known ensemble.

[30] There also exists the so-called *extended system methods*, such as the Nosé-Hoover scheme, that will not be covered here.

Stochastic methods, on the contrary, alleviate many of the issues encountered with the previous methods. First, they act on a particle-centered instead of a system-as-a-whole basis for temperature control, circumventing the 'hot-solvent, cold-solute' and the 'flying-ice-cube' artifacts. Second, they do strictly generate canonical ensembles.

One of the most broadly employed stochastic methods is the scheme proposed by Andersen [181], which is discussed in Appendix A.IV. Another common approach, and the one preferentially applied in the MD simulations performed in this dissertation, is the *Langevin* thermostat [166,182,183]. In this scheme, two forces are applied to every particle in the system at each time step, simulating the contact with a *heat bath*. The temperature is controlled via these two force terms, which modify the equation of motion described in Equation 29 in the following manner:

$$\boldsymbol{F}_i = -\nabla V(\boldsymbol{r}_i) - m\boldsymbol{v}_i\gamma + \boldsymbol{W}_i \qquad 34$$

where γ is a friction constant (i.e, a *collision frequency*) and $\boldsymbol{W}_i$ is a random force, uncorrelated with the positions and momenta of the particles, that follows a Gaussian distribution of zero mean and variance:

$$\boldsymbol{W}_i = \sqrt{2m\gamma k_B T/\Delta t} \qquad 35$$

with Δt being the MD time step and k_B the Boltzmann constant.

Conveniently, the Andersen and Langeving methods sample correctly the NVT and NPT ensemble probabilities, and indeed are effective for temperature control. Thus, the estimation of thermodynamic averages can be this way calculated in the pursued laboratory conditions. Additionally, the Langevin thermostat (and to a certain extent, the Andersen thermostat as well) allows using larger time steps to integrate the equations of motion so that simulation times can be importantly extended (in combination with the use of SHAKE, see below), in comparison with the standard time steps used when simulations are conducted in the microcanonical ensemble. As a downside, the use of stochastics methods impede the calculation of time-dependent properties such as diffusion coefficients, as a result of the discontinuity introduced in the trajectory with respect to the Newtonian dynamics.

6.2. Pressure control

Simulations in the isothermal-isobaric ensemble (NPT) relies on similar stochastic and non-stochastic algorithms to control the pressure of the system. The general details of how the pressure is computed in MD simulations can be found in Appendix A.V. Most simulations described

herein have been computed using the Berendsen scheme, which will be presented below. An alternative stochastic approach is discussed as well in Appendix A.VI.

To maintain a constant pressure during a simulation, the volume must fluctuate. This can be accomplished by simply scaling the volume, similarly to the velocity scaling for temperature control before mentioned. Alternatively, the pressure of the system can be modulated according to the same principle of the scaling method used by Berendsen to maintain a constant temperature (see Appendix A.III) [165,184]. In this case, the system can be imagined to be coupled with a 'pressure bath', and the volume of the box is adapted by a factor λ to preserve the desired pressure:

$$\lambda = 1 - \beta \frac{\Delta t}{\tau_P}(P_f - P_0) \qquad 36$$

where τ_P is the 'coupling constant' of the bath, Δt is the MD time step used, β is the isothermal compressibility[31], and P_0 and P_f are the current and the target pressure of the system, respectively. As it can be observed in Equation 36, τ_P controls the strength of the coupling, and thus can be used to regulate the smoothness of the scaling.

Assuming the system is placed in a cubic box (according to PBC) of volume $V_{box,0}$, the volume is then scaled as:

$$V_{box,f} = \lambda V_{box,0} \qquad 37$$

and the center of mass coordinates of each molecule is thus scaled as:

$$\boldsymbol{r}_f = \lambda^{1/3} \boldsymbol{r}_0 \qquad 38$$

being $\boldsymbol{r}_0$ and $\boldsymbol{r}_f$ the current and scaled center of mass vector position, respectively. In this example the scaling is performed isotropically, but anisotropic scaling can be chosen as well. Although this Berendsen scaling method is widely employed in MD simulations and proved to work properly when a sufficiently large τ_P (> 0.1 ps) is used [184], it does not strictly sample from any know ensemble.

[31]It can be defined as the fractional change in volume of a fluid due to a change in pressure: $\beta = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_T$. It is typically set to the experimental of the solvent used in the simulation.

7. CONSTRAINED MD SIMULATIONS

A common practice in current MD simulations of large biomolecular systems is the use of *constrains*. These are imposed requirements that the system must fulfill. For instance, it is a common practice in MD simulations to use constraints on bond lengths, so that they are forced to adopt specific values. Thus, the equations of motion are solved taking into account the modified constrained terms.

Probably, the most widely employed algorithm to perform constrained MD simulations is SHAKE [185] (or others based on it [186]). It works by solving each constraint equation in an iterative manner, up to a point in which all have been satisfied within a certain tolerance. This process is performed at each time step to avoid drifting in the constraint values along the simulation. When using constrains, it is important to ensure that the influence of the algorithm over the trajectory is small. For example, the fluctuations introduced should be smaller than those caused by other sources (such as the already mentioned non-bonded cutoffs), and the affected degrees of freedom must be weakly coupled to other degrees of freedom of the system. Obviously, the constrained degrees of freedom should not be participating in any of the important phenomena under study. In practice, the use of SHAKE is typically restricted to stretching of bonds involving hydrogen atoms, since these are the ones with higher vibration frequency.

The main advantage of setting constrains in a simulation is that larger time steps can be used, which allows notably extending simulation times (up to the millisecond timescale) and thus –hopefully- improving the exploration of phase space. Currently, the use of constrains in combination with temperature control schemes (such as the Langevin or Andersen thermostats) is a widespread protocol in MD simulations of large biomolecular systems, and as such it has been employed throughout most simulations described in the following chapters.

8. MD SIMULATIONS OF CARBOHYDRATES

Carbohydrates are in general very flexible molecules. From a conformational perspective, crystallographic structures are clearly insufficient to explain behavior in solution of sugars apart from the simplest cases. In contrast, solution NMR spectroscopy is especially suitable to address their study [109,112]. The observed NMR data, however, is averaged over all conformational states that takes place in solution sufficiently fast. This is the case, for instance, of the rotameric populations about the C5-C6 bond in pyranoses (defined by ω angle, as seen before). Therefore, the

deconvolution of the observed NMR properties into the individual three-dimensional structures contributing to them is mandatory to properly interpret the experimental results.

MD simulations in explicit solvent have become the preferred computational approach to tackle this problem. The first attempts to correlate NMR derived properties with those calculated *in vacuo* by first principles or molecular mechanics failed, especially due to the overestimation of intramolecular hydrogen bonds, and only seem to work when those hydrogen bonds were diminished somehow [13,187,188]. On the contrary, MD simulations have proved to accurately predict the conformational states of simple sugars present in solution [173,187,189–191], thanks to the remarkable improvements in classical mechanical force fields in the later years, together with the increased simulation times accessible nowadays. Still, the question of how much sampling is sufficient has been an issue since the beginning of MD simulations for all kind of molecules. In the case of carbohydrates, it seems that, when several glycosidic bonds are present in the studied molecule, convergence may be reached, if ever, well into the microsecond time scale [192].

Nevertheless, it is not always necessary to get a complete prediction of rotametic populations. For instance, when trying to derive populations from observed scalar couplings (J coupling) it is sufficient that all relevant states be identified during the simulation. Then, the MD generated ensembles can be combined with quantum mechanical predicted J values and experimentally observed J couplings to yield to populations of each conformer [189]. Other NMR observables can be combined as well with the MD ensembles, including residual dipolar couplings (RDCs) [193,194], NOEs [195] or recently STD effects (this, in particular, is addressed in *Chapter 5*).

The interactions of proteins and carbohydrates can be scrutinized as well by MD simulations, which can provide an atomic resolved interpretation to the NMR experiments performed. The combination of different but compatible force fields for proteins and sugars (and other type of molecules) has permitted to obtain significant insights into how these protein-glycan interactions occur. Moreover, computational simulations can help disentangle the energy components involved in the molecular recognition process [49,196].

REFERENCES

1. Lis, H.; Sharon, N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. *Chem. Rev.* **1998**, *98*, 637–674.

2. *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., Schnaar, R.L., Seeberger, P.H., Eds.; 3rd editio.; 2017;

3. McNaught, A.D. International Union of Pure and Applied Chemistry and International Union of Biochemistry and Molecular Biology Joint Commission on Biochemical Nomenclature Nomenclature of Carbohydrates. *Adv. Carbohydr. Chem. Biochem.* **1997**, *68*, 43–171.

4. Kamerling, J.; Boons, G.-J.; Lee, Y.C.; Suzuki, A.; Taniguchi, N.; Voragen, A.G.J. *Comprehensive Glycoscience (From Chemistry to Systems Biology)*; Kamerling, J.P., Ed.; Utrecht, The Netherlands, 2008;

5. Clayden, J.; Greeves, N.; Warren, S. *Organic Chemistry*; OUP Oxford, 2012; ISBN 9780199270293.

6. Schwart, J.C.P. Rules for Conformation Nomenclature for Five- and Six-Membered Rings in Monosaccharides and Their Derivatives. *J. Chem. Soc. Perkin Trans. 1* **1973**, 1–4.

7. Dixon, B.F.; Jeannin, Y.; Loening, K.L.; Moss, G.P. Conformational Nomenclature for Five and Six- Membered Ring Forms of Monosaccharides and Their Derivatives: Recommendations 1980. *Eur. J. Biochem.* **1980**, *111*, 295–298.

8. Martin, A.; Arda, A.; Désiré, J.; Martin-Mingot, A.; Probst, N.; Sinaÿ, P.; Jiménez-Barbero, J.; Thibaudeau, S.; Blériot, Y. Catching Elusive Glycosyl Cations in a Condensed Phase with HF/SbF5 Superacid. *Nat. Chem.* **2016**, *8*.

9. Edward; T., J. Stability of Glycosides to Acid Hydrolysis. *Chem. Ind.* **1955**, *3*.

10. Lemieux, R.U.; Chu, P. Conformations and Relative Stabilities of Acetylated Sugars as Determined by NMR Spectroscopy and Anomerization Equilibria. *Abstr. Pap. Am. Chem. Soc.* **1958**, *133*.

11. Bock, K.; Duus, J. A Conformational Study of Hydroxymethyl Groups in Carbohydrates Investigated by H NMR Spectroscopy. *J. Carbohydr. Chem.* **1994**, *13*, 513–543.

12. Nóbrega, C.; Vázquez, J.T. Conformational Study of the Hydroxymethyl Group in α-D-Mannose Derivatives. *Tetrahedron Asymmetry* **2003**, *14*, 2793–2801.

13. Rockwell, G.D.; Grindley, T.B. Effect of Solvation on the Rotation of Hydroxymethyl Groups in Carbohydrates. *J. Am. Chem. Soc.* **1998**, *120*, 10953–10963.

14. Hori, H.; Nishida, Y.; Ohrui, H.; Meguro, H. Conformational Analysis of Hydroxymethyl Group of D-Mannose Derivatives Using (65)- and (6R)-(6-2H1)-D-Mannose. *J. Carbohydr. Chem.* **1990**, *9*.

15. Nishida, Y.; Ohrui, H.; Meguro, H. 1H-NMR Studies of (6r)- and (6s)-Deuterated d-Hexoses: Assignment of the Preferred Rotamers about C5□C6 Bond of D-Glucose and D-Galactose Derivatives in Solutions. *Tetrahedron Lett.* **1984**, *25*, 1575–1578.

16. Ohrui, H.; Nishida, Y.; Higuchi, H.; Hori, H.; Meguro, H. The Preferred Rotamer about the C 5 —C 6 Bond of D -Galactopyranoses and the Stereochemistry of Dehydrogenation by D -Galactose Oxidase. *Can. J. Chem.* **1987**, *65*, 1145–1153.

17. Cornish-Bowden Symbols for Specifying the Conformation of Polysaccharide Chains. *Pure Appl. Chem* **1983**, *55*, 1269–1272.

18. *Glycoscience: Chemistry and Chemical Biology I–III*; Fraser-Reid, B.O., Tatsuta, K., Thiem, J., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; ISBN 978-3-540-67764-2.

19. *Anomeric Effect*; Szarek, W.A., Horton, D., Eds.; ACS Symposium Series; AMERICAN CHEMICAL SOCIETY: WASHINGTON, D. C., 1979; Vol. 87; ISBN 9780841204706.

20. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in Health and Disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366.

21. Pinho, S.S.; Reis, C.A. Glycosylation in Cancer: Mechanisms and Clinical Implications. *Nat. Rev. Cancer* **2015**, *15*, 540–555.

22. Dwek, R.A. Glycobiology: "Towards Understanding the Function of Sugars." *Biochem. Soc. Trans.* **1995**, *23*, 1–25.

23. Varki, A. Biological Roles of Glycans. *Glycobiology* **2017**, *27*, 3–49.

24. Ernst, B.; Magnani, J.L. From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Discov.* **2009**, *8*, 661–677.

25. Hevey, R. Strategies for the Development of Glycomimetic Drug Candidates. *Pharmaceuticals* **2019**, *12*.

26. Tamburrini, A.; Colombo, C.; Bernardi, A. Design and Synthesis of Glycomimetics: Recent Advances. *Med. Res. Rev.* **2020**, *40*, 495–531.

27. Gillis, E.P.; Eastman, K.J.; Hill, M.D.; Donnelly, D.J.; Meanwell, N.A. Applications of Fluorine in Medicinal Chemistry. *J. Med. Chem.* **2015**, *58*, 8315–8359.

28. Bratteby, K.; Torkelsson, E.; L'Estrade, E.T.; Peterson, K.; Shalgunov, V.; Xiong, M.; Leffler, H.; Zetterberg, F.R.; Olsson, T.G.; Gillings, N.; et al. In Vivo Veritas: 18F-Radiolabeled Glycomimetics Allow Insights into the Pharmacological Fate of Galectin-3 Inhibitors. *J. Med. Chem.* **2020**, *63*.

29. Vulpetti, A.; Hommel, U.; Landrum, G.; Lewis, R.; Dalvit, C. Design and NMR-Based Screening of LEF, a Library of Chemical Fragments with Different Local Environment of Fluorine. *J. Am. Chem. Soc.* **2009**, *131*, 12949–12959.

30. Chen, H.; Viel, S.; Ziarelli, F.; Peng, L. F NMR: A Valuable Tool for Studying Biological Events. *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.

31. Linclau, B.; Ardá, A.; Reichardt, N.-C.; Sollogoub, M.; Unione, L.; Vincent, S.P.; Jiménez-Barbero, J. Fluorinated Carbohydrates as Chemical Probes for Molecular Recognition Studies. Current Status and Perspectives. *Chem. Soc. Rev.* **2020**, *49*, 3863–3888.

32. Hoffmann, M.; Rychlewski, J. When, in the Context of Drug Design, Can a Fluorine Atom Successfully Substitute a Hydroxyl Group? *Int. J. Quantum Chem.* **2002**, *89*, 419–427.

33. Hunter, L. The C-F Bond as a Conformational Tool in Organic and Biological Chemistry. *Beilstein J. Org. Chem.* **2010**, *6*, 1–14.

34. O'Hagan, D. Understanding Organofluorine Chemistry. An Introduction to the C-F Bond. *Chem. Soc. Rev.* **2008**, *37*, 308–319.

35. Denavit, V.; Lainé, D.; St-Gelais, J.; Johnson, P.A.; Giguère, D. A Chiron Approach towards the Stereoselective Synthesis of Polyfluorinated Carbohydrates. *Nat. Commun.* **2018**, *9*.

36. Withers, S.G.; Street, J.P.; Rettig, S.J. The Preferred Conformation of 2-Fluoro-2-Deoxy β-D-Mannopyranosyl Fluoride. An X-Ray Crystallographic and 2-Dimensional Proton Nuclear Magnetic

Resonance Study. *Can. J. Chem.* **1986**, *64*.

37. Diercks, T.; Infantino, A.S.; Unione, L.; Jiménez-Barbero, J.; Oscarson, S.; Gabius, H.-J. Fluorinated Carbohydrates as Lectin Ligands: Synthesis of OH/F-Substituted N -Glycan Core Trimannoside and Epitope Mapping by 2D STD-TOCSYreF NMR Spectroscopy. *Chem. - A Eur. J.* **2018**, *24*, 15761–15765.

38. Kerins, L.; Byrne, S.; Gabba, A.; Murphy, P. V. Anomer Preferences for Glucuronic and Galacturonic Acid and Derivatives and Influence of Electron-Withdrawing Substituents. *J. Org. Chem.* **2018**, *83*.

39. Linclau, B.; Wang, Z.; Compain, G.; Paumelle, V.; Fontenelle, C.Q.; Wells, N.; Weymouth-Wilson, A. Investigating the Influence of (Deoxy)Fluorination on the Lipophilicity of Non-UV-Active Fluorinated Alkanols and Carbohydrates by a New Log P Determination Method. *Angew. Chemie Int. Ed.* **2016**, *55*, 674–678.

40. Sharon, N. Lectins: Past, Present and Future. *Biochem. Soc. Trans.* **2008**, *36*, 1457–1460.

41. Gupta, G.S. *Animal Lectins: Form, Function and Clinical Applications*; Springer Vienna: Vienna, 2012; ISBN 978-3-7091-1064-5.

42. Brown, G.D.; Willment, J.A.; Whitehead, L. C-Type Lectins in Immunity and Homeostasis. *Nat. Rev. Immunol.* 2018, *18*, 374–389.

43. Mammen, M.; Choi, S.K.; Whitesides, G.M. Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chemie - Int. Ed.* **1998**, *37*, 2754–2794.

44. Quiocho, F.A. Carbohydrate-Binding Proteins: Tertiary Structures and Protein-Sugar Interactions. *Annu. Rev. Biochem.* **1986**, *VOL. 55*, 287–315.

45. del Carmen Fernández-Alonso, M.; Díaz, D.; Berbis, M.Á.; Marcelo, F.; Cañada, J.; Jiménez-Barbero, J. Protein-Carbohydrate Interactions Studied by NMR: From Molecular Recognition to Drug Design. *Curr. Protein Pept. Sci.* **2012**, *13*, 816–30.

46. Spiwok, V. CH/π Interactions in Carbohydrate Recognition. *Molecules* **2017**, *22*, 1–11.

47. Darby, J.F.; Hopkins, A.P.; Shimizu, S.; Roberts, S.M.; Brannigan, J.A.; Turkenburg, J.P.; Thomas, G.H.; Hubbard, R.E.; Fischer, M. Water Networks Can Determine the Affinity of Ligand Binding to

Proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15818–15826.

48. Breiten, B.; Lockett, M.R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C.M.; Heroux, A.; Krilov, G.; Whitesides, G.M. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding. *J. Am. Chem. Soc.* **2013**, *135*, 15579–15584.

49. Kadirvelraj, R.; Foley, B.L.; Dyekjær, J.D.; Woods, R.J. Involvement of Water in Carbohydrate−Protein Binding: Concanavalin A Revisited. *J. Am. Chem. Soc.* **2008**, *130*, 16933–16942.

50. Sager, C.P.; Eriş, D.; Smieško, M.; Hevey, R.; Ernst, B. What Contributes to an Effective Mannose Recognition Domain? *Beilstein J. Org. Chem.* **2017**, *13*, 2584–2595.

51. Brewer, C.F.; Bhattacharyya, L.; Brown, R.D.; Koenig, S.H. Preparation and Properties of Metal Ion Derivatives of the Lentil and Pea Lectins. *Biochemistry* **1985**, *24*.

52. Bhattacharyya, L.; Brewer, C.F. Lectin-Carbohydrate Interactions. **1988**, *212*, 207–212.

53. Schwarz, F.P.; Puri, K.D.; Bhat, R.G.; Surolia, A. Thermodynamics of Monosaccharide Binding to Concanavalin A, Pea (Pisum Sativum) Lectin Lentil (Lens Culinaris) Lectin. *J. Biol. Chem.* **1993**, *268*, 7668–7677.

54. Ruzeinikov, S.N.; Mikhailova, I.Y.; Tsygannik, I.N.; Pangborn, W.D.; W. Pletnev, V.. The Structure of the Pea Lectin-D-Mannopyranose Complex at a 2.1 A Resolution. *RUSS.J.BIOORGANIC CHEM* **1998**, *24*, 277–279.

55. Pletnev, V.Z.; Ruzheinikov, S.N.; Tsygannik, I.N.; Mikhailova, I.Y.; Duax, W.; Ghosh, D.; Pangborn, W. The Structure of Pea Lectin-D-Glucopyranose Complex at a 1.9 Å Resolution. *Russ. J. Bioorganic Chem.* 1997, *23*.

56. Singh, S.K.; Streng-Ouwehand, I.; Litjens, M.; Weelij, D.R.; García-Vallejo, J.J.; van Vliet, S.J.; Saeland, E.; van Kooyk, Y. Characterization of Murine MGL1 and MGL2 C-Type Lectins: Distinct Glycan Specificities and Tumor Binding Properties. *Mol. Immunol.* **2009**, *46*, 1240–1249.

57. Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E.J.; Gabius, H.-J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural

Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2014**, *20*, 16147–16155.

58. van Vliet, S.J.; van Liempt, E.; Saeland, E.; Aarnoudse, C.A.; Appelmelk, B.; Irimura, T.; Geijtenbeek, T.B.H.; Blixt, O.; Alvarez, R.; van Die, I.; et al. Carbohydrate Profiling Reveals a Distinctive Role for the C-Type Lectin MGL in the Recognition of Helminth Parasites and Tumor Antigens by Dendritic Cells. *Int. Immunol.* **2005**, *17*, 661–669.

59. Diniz, A.; Coelho, H.; Dias, J.S.; van Vliet, S.J.; Jiménez-Barbero, J.; Corzana, F.; Cabrita, E.J.; Marcelo, F. The Plasticity of the Carbohydrate Recognition Domain Dictates the Exquisite Mechanism of Binding of Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **2019**, *25*.

60. Higashi, N.; Fujioka, K.; Denda-Nagai, K.; Hashimoto, S.I.; Nagai, S.; Sato, T.; Fujita, Y.; Morikawa, A.; Tsuiji, M.; Miyata-Takeuchi, M.; et al. The Macrophage C-Type Lectin Specific for Galactose/N-Acetylgalactosamine Is an Endocytic Receptor Expressed on Monocyte-Derived Immature Dendritic Cells. *J. Biol. Chem.* **2002**, *277*.

61. Davidson, A.L.; Dassa, E.; Orelle, C.; Chen, J. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. *Microbiol. Mol. Biol. Rev.* **2008**, *72*.

62. Hazelbauer, G.L. Bacterial Chemotaxis: The Early Years of Molecular Studies. *Annu. Rev. Microbiol.* **2012**, *66*, 285–303.

63. Sooriyaarachchi, S.; Ubhayasekera, W.; Boos, W.; Mowbray, S.L. X-Ray Structure of Glucose/Galactose Receptor from Salmonella Typhimurium in Complex with the Physiological Ligand, (2R)-Glyceryl-??-D- Galactopyranoside. *FEBS J.* **2009**, *276*, 2116–2124.

64. Ortega, G.; Diercks, T.; Millet, O. Carbohydrate Affinity for the Glucose − Galactose Binding Protein Is Regulated by Allosteric Domain Motions. **2012**.

65. Unione, L.; Ortega, G.; Mallagaray, A.; Corzana, F.; Pérez-Castells, J.; Canales, A.; Jiménez-Barbero, J.; Millet, O. Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations. *ACS Chem. Biol.* **2016**, *11*.

66. Geijtenbeek, T.B.H.; Torensma, R.; Van Vliet, S.J.; Van Duijnhoven,

G.C.F.; Adema, G.J.; Van Kooyk, Y.; Figdor, C.G. Identification of DC-SIGN, a Novel Dendritic Cell-Specific ICAM-3 Receptor That Supports Primary Immune Responses. *Cell* **2000**, *100*, 575–585.

67. Mitchell, D.A.; Fadden, A.J.; Drickamer, K. A Novel Mechanism of Carbohydrate Recognition by the C-Type Lectins DC-SIGN and DC-SIGNR. *J. Biol. Chem.* **2001**, *276*, 28939–28945.

68. Geijtenbeek, T.B.H.; Kwon, D.S.; Torensma, R.; Van Vliet, S.J.; Van Duijnhoven, G.C.F.; Middel, J.; Cornelissen, I.L.M.H.A.; Nottet, H.S.L.M.; KewalRamani, V.N.; Littman, D.R.; et al. DC-SIGN, a Dendritic Cell-Specific HIV-1-Binding Protein That Enhances Trans-Infection of T Cells. *Cell* **2000**, *100*, 587–597.

69. Valverde, P.; Martínez, J.D.; Cañada, F.J.; Ardá, A.; Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* 2020, *21*.

70. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D.A.; Alvarez, R.; Blixt, O.; Taylor, M.E.; Weis, W.I.; Drickamer, K. Structural Basis for Distinct Ligand-Binding and Targeting Properties of the Receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **2004**, *11*, 591–598.

71. Feinberg, H.; Taylor, M.E.; Razi, N.; McBride, R.; Knirel, Y.A.; Graham, S.A.; Drickamer, K.; Weis, W.I. Structural Basis for Langerin Recognition of Diverse Pathogen and Mammalian Glycans through a Single Binding Site. *J. Mol. Biol.* **2011**, *405*, 1027–1039.

72. Bloch, F, Hanse W.W, P.M. The Nuclear Experiment. *Phys. Rev.* **1946**, *70*, 474–485.

73. Holdgate, G.A. Making Cool Drugs Hot: Isothermal Titration Calorimetry as a Tool to Study Binding Energetics. *Biotechniques* **2001**, *31*, 164–184.

74. Duff, M.R.; Grubbs, J.; Howell, E.E. Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. *J. Vis. Exp.* **2011**.

75. Nguyen, H.H.; Park, J.; Kang, S.; Kim, M. Surface Plasmon Resonance: A Versatile Technique for Biosensor Applications. *Sensors (Switzerland)* 2015, *15*.

76. Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T.L.; Homans, S.W.; Kessler, H.; Luchinat, C.; et al. Perspectives on NMR in Drug Discovery: A Technique Comes of

Age. *Nat. Rev. Drug Discov.* **2008**, *7*, 738–745.

77. Li, Q.; Kang, C.B. A Practical Perspective on the Roles of Solution NMR Spectroscopy in Drug Discovery. *Molecules* **2020**, *25*, 1–19.

78. Dalvit, C.; Flocco, M.; Knapp, S.; Mostardini, M.; Perego, R.; Stockman, B.J.; Veronesi, M.; Varasi, M. High-Throughput NMR-Based Screening with Competition Binding Experiments. *J. Am. Chem. Soc.* **2002**, *124*, 7702–7709.

79. Wang, Y. Sen; Liu, D.; Wyss, D.F. Competition STD NMR for the Detection of High-Affinity Ligands and NMR-Based Screening. *Magn. Reson. Chem.* **2004**, *42*, 485–489.

80. Zheng, H.; Handing, K.B.; Zimmerman, M.D.; Shabalin, I.G.; Almo, S.C.; Minor, W. X-Ray Crystallography over the Past Decade for Novel Drug Discovery -Where Are We Heading Next? *Expert Opin. Drug Discov.* 2015, *10*.

81. Keeler, J. *Understanding NRM Spectroscopy*; Second edi.; John Wiley & Sons, Ltd: Cambridge, 2010; ISBN 978-0-470-74608-0.

82. Levitt, M.H. *Spin Dynamics*; 2nd Editio.; John Wiley & Sons, Ltd: Southampton, 2008; ISBN 978-0-470-51117-6.

83. Hore, P.; Jones, J.A.; Wimperis, S. *NMR: The Toolkit. How Pulse Sequences Work*; 2nd Revise.; Oxford University Press: Oxford, 2015; ISBN 0198703422.

84. Jacobsen, N.E. *NMR Spectroscopy Explained*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007; ISBN 9780470173350.

85. S. Rule, G.; Hitchens, T.K. *Fundamentals of Protein NMR Spectroscopy*; Springer: Netherlands, 2006; ISBN 1402034997.

86. Peng, J.W.; Moore, J.; Abdul-Manan, N. NMR Experiments for Lead Generation in Drug Discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *44*, 225–256.

87. Claridge, T.D. *High-Resolution NMR Techniques in Organic Chemistry*; 2009; ISBN 978-0-08-054628-5.

88. Wang, G.F.; Li, C.; Pielak, G.J. Probing the Micelle-Bound Aggregation-Prone State of α-Synuclein With19F NMR Spectroscopy. *ChemBioChem* **2010**.

89. Li, H.; Frieden, C. Observation of Sequential Steps in the Folding of Intestinal Fatty Acid Binding Protein Using a Slow Folding Mutant and 19F NMR. *Proc. Natl. Acad. Sci. U. S. A.* **2007**.

90. Holzberger, B.; Rubini, M.; Möller, H.M.; Marx, A. A Highly Active DNA Polymerase with a Fluorous Core. *Angew. Chemie - Int. Ed.* **2010**.

91. Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. Ribose 2'-F Labeling: A Simple Tool for the Characterization of RNA Secondary Structure Equilibria by 19F NMR Spectroscopy. *J. Am. Chem. Soc.* **2005**.

92. Li, Q.; Chen, J.; Trajkovski, M.; Zhou, Y.; Fan, C.; Lu, K.; Tang, P.; Su, X.; Plavec, J.; Xi, Z.; et al. 4'-Fluorinated RNA: Synthesis, Structure, and Applications as a Sensitive 19F NMR Probe of RNA Structure and Function. *J. Am. Chem. Soc.* **2020**.

93. Kreutz, C.; Kählig, H.; Konrat, R.; Micura, R. A General Approach for the Identification of Site-Specific RNA Binders by 19F NMR Spectroscopy: Proof of Concept. *Angew. Chemie - Int. Ed.* **2006**.

94. Li, C.; Wang, G.F.; Wang, Y.; Creager-Allen, R.; Lutz, E.A.; Scronce, H.; Slade, K.M.; Ruf, R.A.S.; Mehl, R.A.; Pielak, G.J. Protein 19F NMR in Escherichia Coli. *J. Am. Chem. Soc.* **2010**, *132*, 321–327.

95. Unione, L.; Alcalá, M.; Echeverria, B.; Serna, S.; Ardá, A.; Franconetti, A.; Cañada, F.J.; Diercks, T.; Reichardt, N.; Jiménez-Barbero, J. Fluoroacetamide Moieties as NMR Spectroscopy Probes for the Molecular Recognition of GlcNAc-Containing Sugars: Modulation of the CH-π Stacking Interactions by Different Fluorination Patterns. *Chem. - A Eur. J.* **2017**, *23*, 3957–3965.

96. Calle, L.P.; Echeverria, B.; Franconetti, A.; Serna, S.; Fernández-Alonso, M.C.; Diercks, T.; Cañada, F.J.; Ardá, A.; Reichardt, N.-C.; Jiménez-Barbero, J. Monitoring Glycan-Protein Interactions by NMR Spectroscopic Analysis: A Simple Chemical Tag That Mimics Natural CH-π Interactions. *Chem. - A Eur. J.* **2015**, *21*, 11408–11416.

97. N'Go, I.; Golten, S.; Ardá, A.; Cañada, J.; Jiménez-Barbero, J.; Linclau, B.; Vincent, S.P. Tetrafluorination of Sugars as Strategy for Enhancing Protein-Carbohydrate Affinity: Application to UDP-Galp Mutase Inhibition. *Chem. - A Eur. J.* **2014**, *20*, 106–112.

98. Aretz, J.; Anumala, U.R.; Fuchsberger, F.F.; Molavi, N.; Ziebart, N.; Zhang, H.; Nazaré, M.; Rademacher, C. Allosteric Inhibition of a Mammalian Lectin. *J. Am. Chem. Soc.* **2018**, *140*, 14915–14925.

99. Denavit, V.; Lainé, D.; Bouzriba, C.; Shanina, E.; Gillon, É.; Fortin, S.; Rademacher, C.; Imberty, A.; Giguère, D. Stereoselective Synthesis of Fluorinated Galactopyranosides as Potential Molecular Probes for Galactophilic Proteins: Assessment of Monofluorogalactoside–LecA Interactions. *Chem. – A Eur. J.* **2019**, *25*, 4478–4490.

100. Xu, B.; Unione, L.; Sardinha, J.; Wu, S.; Ethève-Quelquejeu, M.; Pilar Rauter, A.; Blériot, Y.; Zhang, Y.; Martín-Santamaría, S.; Díaz, D.; et al. Gem-Difluorocarbadisaccharides: Restoring the Exo-Anomeric Effect. *Angew. Chemie - Int. Ed.* **2014**, *53*.

101. Dalvit, C.; Fagerness, P.E.; Hadden, D.T.A.; Sarver, R.W.; Stockman, B.J. Fluorine-NMR Experiments for High-Throughput Screening: Theoretical Aspects, Practical Considerations, and Range of Applicability. *J. Am. Chem. Soc.* **2003**, *125*, 7696–7703.

102. Dalvit, C.; Ardini, E.; Flocco, M.; Fogliatto, G.P.; Mongelli, N.; Veronesit, M. A General NMR Method for Rapid, Efficient, and Reliable Biochemical Screening. *J. Am. Chem. Soc.* **2003**, *125*.

103. Taylor, J.D.; Gilbert, P.J.; Williams, M.A.; Pitt, W.R.; Ladbury, J.E. Identification of Novel Fragment Compounds Targeted against the PY Pocket of V-Src SH2 by Computational and NMR Screening and Thermodynamic Evaluation. *Proteins Struct. Funct. Genet.* **2007**, *67*.

104. Dalvit, C. Ligand- and Substrate-Based19F NMR Screening: Principles and Applications to Drug Discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243–271.

105. Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62*, 2218–2244.

106. Dalvit, C. Theoretical Analysis of the Competition Ligand-Based NMR Experiments and Selected Applications to Fragment Screening and Binding Constant Measurements. *Concepts Magn. Reson. Part A* **2008**, *32A*, 341–372.

107. Meyer, B.; Peters, T. NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors. *Angew. Chemie Int. Ed.* **2003**, *42*, 864–890.

108. Gimeno, A.; Valverde, P.; Ardá, A.; Jiménez-Barbero, J. Glycan Structures and Their Interactions with Proteins. A NMR View. *Curr. Opin. Struct. Biol.* 2020, *62*.

109. Valverde, P.; Quintana, J.I.; Santos, J.I.; Ardá, A.; Jiménez-Barbero, J. Novel NMR Avenues to Explore the Conformation and Interactions of Glycans. *ACS Omega* **2019**, *4*.

110. Burz, D.S.; Dutta, K.; Cowburn, D.; Shekhtman, A. Mapping Structural Interactions Using In-Cell NMR Spectroscopy (STINT-NMR). *Nat. Methods* **2006**, *3*.

111. Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M.A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F.J.; et al. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chemie Int. Ed.* **2019**, *58*, 7268–7272.

112. Roldós, V.; Cañada, F.J.; Jiménez-Barbero, J. Carbohydrate-Protein Interactions: A 3D View by NMR. *ChemBioChem* **2011**, *12*, 990–1005.

113. Angulo, J.; Rademacher, C.; Biet, T.; Benie, A.J.; Blume, A.; Peters, H.; Palcic, M.; Parra, F.; Peters, T. NMR Analysis of Carbohydrate-Protein Interactions. *Methods Enzymol.* 2006, *416*.

114. Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. A Solution NMR Study of the Interactions of Oligomannosides and the Anti-HIV-1 2G12 Antibody Reveals Distinct Binding Modes for Branched Ligands. *Chem. - A Eur. J.* **2011**, *17*, 1547–1560.

115. Mayer, M.; Meyer, B. Group Epitope Mapping by Saturation Transfer Difference NMR to Identify Segments of a Ligand in Direct Contact with a Protein Receptor. *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.

116. Fiege, B.; Rademacher, C.; Cartmell, J.; Kitov, P.I.; Parra, F.; Peters, T. Molecular Details of the Recognition of Blood Group Antigens by a Human Norovirus as Determined by STD NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **2012**, *51*.

117. Zuiderweg, E.R.P. Mapping Protein-Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry* **2002**, *41*, 1–7.

118. Vaynberg, J.; Qin, J. Weak Protein-Protein Interactions as Probed by NMR Spectroscopy. *Trends Biotechnol.* 2006, *24*.

119. Yamaoki, Y.; Nagata, T.; Sakamoto, T.; Katahira, M. Recent Progress of In-Cell NMR of Nucleic Acids in Living Human Cells. *Biophys. Rev.* 2020, *12*.

120. Walinda, E.; Morimoto, D.; Sugase, K. Resolving Biomolecular Motion and Interactions by R2 and R1ϱ Relaxation Dispersion NMR. *Methods* **2018**, *148*, 28–38.

121. Lucas, L.H.; Larive, C.K. Measuring Ligand-Protein Binding Using NMR Diffusion Experiments. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **2004**, *20*.

122. Johnson, C.S. Diffusion Ordered Nuclear Magnetic Resonance Spectroscopy: Principles and Applications. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 203–256.

123. Monaco, S.; Tailford, L.E.; Juge, N.; Angulo, J. Differential Epitope Mapping by STD NMR Spectroscopy To Reveal the Nature of Protein–Ligand Contacts. *Angew. Chemie - Int. Ed.* **2017**, *56*, 15289–15293.

124. Angulo, J.; Nieto, P.M. STD-NMR: Application to Transient Interactions between Biomolecules-a Quantitative Approach. *Eur. Biophys. J.* **2011**, *40*, 1357–1369.

125. Williamson, M.P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **2013**, *73*, 1–16.

126. Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62*, 2218–2244.

127. Matei, E.; Gronenborn, A.M. ^{19}F Paramagnetic Relaxation Enhancement: A Valuable Tool for Distance Measurements in Proteins. *Angew. Chemie - Int. Ed.* **2016**, *55*, 150–154.

128. Liang, B.; Bushweller, J.H.; Tamm, L.K. Site-Directed Parallel Spin-Labeling and Paramagnetic Relaxation Enhancement in Structure Determination of Membrane Proteins by Solution NMR Spectroscopy. *J. Am. Chem. Soc.* **2006**, *128*, 4389–4397.

129. Jahnke, W.; Perez, L.B.; Paris, C.G.; Strauss, A.; Fendrich, G.; Nalin, C.M. Second-Site NMR Screening with a Spin-Labeled First Ligand [5]. *J. Am. Chem. Soc.* **2000**, *122*, 7394–7395.

130. Jahnke, W.; Rüdisser, S.; Zurini, M. Spin Label Enhanced NMR Screening. *J. Am. Chem. Soc.* **2001**, *123*, 3149–3150.

131. Ludwig, C.; Michiels, P.J.A.; Wu, X.; Kavanagh, K.L.; Pilka, E.; Jansson, A.; Oppermann, U.; Günther, U.L. SALMON: Solvent

Accessibility, Ligand Binding, and Mapping of Ligand Orientation by NMR Spectroscopy. *J. Med. Chem.* **2008**, *51*, 1–3.

132. Vulpetti, A.; Caronni, D.; Veronesi, M.; Dalvit, C.; Mongelli, N. NMR-Based Quality Control Approach for the Identification of False Positives and False Negatives in High Throughput Screening. *Curr. Drug Discov. Technol.* **2006**, *3*, 115–124.

133. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-.).* **1996**, *274*, 1531–1534.

134. Dalvit, C. Ligand- and Substrate-Based 19F NMR Screening: Principles and Applications to Drug Discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243–271.

135. Carr, H.Y.; Purcell, E.M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **1954**, *94*.

136. Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, *29*, 688–691.

137. Martínez, J.D.; Manzano, A.I.; Calviño, E.; Diego, A. De; Rodriguez De Francisco, B.; Romanò, C.; Oscarson, S.; Millet, O.; Gabius, H.J.; Jiménez-Barbero, J.; et al. Fluorinated Carbohydrates as Lectin Ligands: Simultaneous Screening of a Monosaccharide Library and Chemical Mapping by 19F NMR Spectroscopy. *J. Org. Chem.* **2020**.

138. Martínez, J.D.; Valverde, P.; Delgado, S.; Romanò, C.; Linclau, B.; Reichardt, N.C.; Oscarson, S.; Ardá, A.; Jiménez-Barbero, J.; Cañada, F.J. Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates. *Molecules* **2019**, *24*, 2337.

139. Swift, T.J.; Connick, R.E. NMR-Relaxation Mechanisms of O^{17} in Aqueous Solutions of Paramagnetic Cations and the Lifetime of Water Molecules in the First Coordination Sphere. *J. Chem. Phys.* **1962**, *37*, 307–320.

140. Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **1999**, *38*, 1784–1788.

141. Bernardi, A.; Arosio, D.; Potenza, D.; Sánchez-Medina, I.; Mari, S.; Cañada, F.J.; Jiménez-Barbero, J. Intramolecular Carbohydrate-Aromatic Interaction and Intermolecular van Der Waals Interactions Enhance the Molecular Recognition Ability of GM1 Glycomimetics

for Cholera Toxin. *Chem. - A Eur. J.* **2004**, *10*.

142. Yuan, Y.; Wen, X.; Sanders, D.A.R.; Pinto, B.M. Exploring the Mechanism of Binding of UDP-Galactopyranose to UDP-Galactopyranose Mutase by STD-NMR Spectroscopy and Molecular Modeling. *Biochemistry* **2005**, *44*.

143. Vogtherr, M.; Peters, T. Application of NMR Based Binding Assays to Identify Key Hydroxy Groups for Intermolecular Recognition. *J. Am. Chem. Soc.* **2000**, *122*.

144. Haselhorst, T.; Weimar, T.; Peters, T. Molecular Recognition of Sialyl Lewisx and Related Saccharides by Two Lectins. *J. Am. Chem. Soc.* **2001**, *123*.

145. Angulo, J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Rama Krishna, N.; Peters, H.; Palcic, M.M.; Peters, T. Blood Group B Galactosyltransferase: Insights into Substrate Binding from NMR Experiments. *J. Am. Chem. Soc.* **2006**, *128*, 13529–13538.

146. Mari, S.; Serrano-Gómez, D.; Cañada, F.J.; Corbí, A.L.; Jiménez-Barbero, J. 1D Saturation Transfer Difference NMR Experiments on Living Cells: The DC-SIGN/Oligomannose Interaction. *Angew. Chemie - Int. Ed.* **2004**, *44*.

147. Blume, A.; Fitzen, M.; Benie, A.J.; Peters, T. Specificity of Ligand Binding to Yeast Hexokinase PII Studied by STD-NMR. *Carbohydr. Res.* **2009**, *344*.

148. Hall, J.L.; Sohail, A.; Cabrita, E.J.; Macdonald, C.; Stockner, T.; Sitte, H.H.; Angulo, J.; MacMillan, F. Saturation Transfer Difference NMR on the Integral Trimeric Membrane Transport Protein GltPh Determines Cooperative Substrate Binding. *Sci. Rep.* **2020**, *10*.

149. Monaco, S.; Walpole, S.; Doukani, H.; Nepravishta, R.; Martínez-Bailén, M.; Carmona, A.T.; Ramos-Soriano, J.; Bergström, M.; Robina, I.; Angulo, J. Exploring Multi-Subsite Binding Pockets in Proteins: DEEP-STD NMR Fingerprinting and Molecular Dynamics Unveil a Cryptic Subsite at the GM1 Binding Pocket of Cholera Toxin B. *Chem. - A Eur. J.* **2020**, *26*.

150. Zhang, Q.; Gimeno, A.; Santana, D.; Wang, Z.; Valdés-Balbin, Y.; Rodríguez-Noda, L.M.; Hansen, T.; Kong, L.; Shen, M.; Overkleeft, H.S.; et al. Synthetic, Zwitterionic Sp1 Oligosaccharides Adopt a Helical Structure Crucial for Antibody Interaction. *ACS Cent. Sci.* **2019**, *5*.

151. Mayer, M.; James, T.L. NMR-Based Characterization of Phenothiazines as a RNA Binding Scaffoldt. *J. Am. Chem. Soc.* **2004**, *126*, 4453–4460.

152. Peñalver, P.; Marcelo, F.; Jiménez-Barbero, J.; Vicent, C. Carbohydrate Recognition at the Minor-Groove of the Self-Complementary Duplex d(CGCGAATTCGCG)2 by a Synthetic Glyco-Oligoamide. *Chem. - A Eur. J.* **2011**, *17*.

153. Souard, F.; Muñoz, E.; Peñalver, P.; Badía, C.; Del Villar-Guerra, R.; Asensio, J.L.; Jiménez-Barbero, J.; Vicent, C. Sugar-Oligoamides: Bound-State Conformation and DNA Minor-Groove-Binding Description by TR-NOESY and Differential-Frequency Saturation-Transfer- Difference Experiments. *Chem. - A Eur. J.* **2008**, *14*.

154. Rademacher, C.; Krishna, N.R.; Palcic, M.; Parra, F.; Peters, T. NMR Experiments Reveal the Molecular Basis of Receptor Recognition by a Calicivirus. *J. Am. Chem. Soc.* **2008**, *130*.

155. Rademacher, C.; Guiard, J.; Kitov, P.I.; Fiege, B.; Dalton, K.P.; Parra, F.; Bundle, D.R.; Peters, T. Targeting Norovirus Infection - Multivalent Entry Inhibitor Design Based on NMR Experiments. *Chem. - A Eur. J.* **2011**, *17*.

156. Haselhorst, T.; Garcia, J.M.; Islam, T.; Lai, J.C.C.; Rose, F.J.; Nicholls, J.M.; Peiris, J.S.M.; Von Itzstein, M. Avian Influenza H5-Containing Virus-like Particles (VLPs): Host-Cell Receptor Specificity by STD NMR Spectroscopy. *Angew. Chemie - Int. Ed.* **2008**, *47*.

157. Viegas, A.; Manso, J.; Nobrega, F.L.; Cabrita, E.J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ.* **2011**, *88*, 990–994.

158. Jayalakshmi, V.; Krishna, N.R. Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand–Receptor Complexes. *J. Magn. Reson.* **2002**, *155*, 106–118.

159. Di Micco, S.; Bassarello, C.; Bifulco, G.; Riccio, R.; Gomez-Paloma, L. Differential-Frequency Saturation Transfer Difference NMR Spectroscopy Allows the Detection of Different Ligand-DNA Binding Modes. *Angew. Chemie - Int. Ed.* **2005**, *45*, 224–228.

160. Diercks, T.; Ribeiro, J.P.; Cañada, F.J.; Andre, S.; Jiménez-Barbero, J.; Gabius, H.J. Fluorinated Carbohydrates as Lectin Ligands: Versatile

Sensors in 19f-Detected Saturation Transfer Difference NMR Spectroscopy. *Chem. - A Eur. J.* **2009**, *15*, 5666–5668.

161. André, S.; Cañada, F.J.; Shiao, T.C.; Largartera, L.; Diercks, T.; Bergeron-Brlek, M.; El Biari, K.; Papadopoulos, A.; Ribeiro, J.P.; Touaibia, M.; et al. Fluorinated Carbohydrates as Lectin Ligands: Biorelevant Sensors with Capacity to Monitor Anomer Affinity in 19F-NMR-Based Inhibitor Screening. *European J. Org. Chem.* **2012**, 4354–4364.

162. Amaro, R.E.; Mulholland, A.J. Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat. Rev. Chem.* **2018**, *2*.

163. Bottaro, S.; Lindorff-Larsen, K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? *Science (80-.).* **2018**, *361*, 355–360.

164. Huggins, D.J.; Biggin, P.C.; Dämgen, M.A.; Essex, J.W.; Harris, S.A.; Henchman, R.H.; Khalid, S.; Kuzmanic, A.; Laughton, C.A.; Michel, J.; et al. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, 1–23.

165. Leach, A. *Molecular Modelling: Principles and Applications*; 2nd Editio.; Pearson, 2001; ISBN 978-0582382107.

166. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; 2nd Editio.; Oxford University Press: Oxford, 2017; Vol. 1; ISBN 9780198803195.

167. McQuarrie, D.A. *Statistical Mechanics*; HARPER & ROW: New York, 1976; ISBN 06-044366-9.

168. Leimkuhler, B.J.; Reich, S.; Skeel, R.D. Integration Methods for Molecular Dynamics. In; 1996.

169. Salomon-Ferrer, R.; Case, D.A.; Walker, R.C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198–210.

170. Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.

171. Wang, J.M.; Wolf, R.M.; Caldwell, J.W.; Kollman, P. a; Case, D. a Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

172. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.; Simmerling, C.; Hauser, K.E. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *Just Accept. Manuscr.* • *Publ. Date* **2015**, 7.

173. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.

174. Ewald, P.P. Die Berechnung Optischer Und Elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *369*, 253–287.

175. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

176. Darden, T.; Perera, L.; Li, L.; Lee, P. New Tricks for Modelers from the Crystallography Toolkit: The Particle Mesh Ewald Algorithm and Its Use in Nucleic Acid Simulations. *Structure* **1999**, *7*, 55–60.

177. Cheatham, T.E.; Miller, J.L.; Fox, T.; Darden, T.A.; Kollman, P.A. Molecular Dynamics Simulations on Solvated Biomolecular Systems: The Particle Mesh Ewald Method Leads to Stable Trajectories of DNA, RNA, and Proteins. *J. Am. Chem. Soc.* **1995**, *117*, 4193–4194.

178. Guenot, J.; Kollman, P.A. Molecular Dynamics Studies of a DNA-binding Protein: 2. An Evaluation of Implicit and Explicit Solvent Models for the Molecular Dynamics Simulation of the Escherichia Coli Trp Repressor. *Protein Sci.* **1992**, *1*, 1185–1205.

179. Cheng, A.; Merz, K.M. Application of the Nosé-Hoover Chain Algorithm to the Study of Protein Dynamics. *J. Phys. Chem.* **1996**, *100*, 1927–1937.

180. Harvey, S.C.; Tan, R.K.-Z.; Cheatham, T.E. The Flying Ice Cube: Velocity Rescaling in Molecular Dynamics Leads to Violation of Energy Equipartition. *J. Comput. Chem.* **1998**, *19*, 726–740.

181. Andersen, H.C. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.

182. Loncharich, R.J.; Brooks, B.R.; Pastor, R.W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523–535.

183. Pastor, R.W.; Brooks, B.R.; Szabo, A. An Analysis of the Accuracy of Langevin and Molecular Dynamics Algorithms. *Mol. Phys.* **1988**, *65*, 1409–1419.

184. Berendsen, H.J.C.; Postma, J.P.M.; Van Gunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

185. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

186. Andersen, H.C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*.

187. Kirschner, K.N.; Woods, R.J. Solvent Interactions Determine Carbohydrate Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*.

188. French, A.D.; Dowd, M.K. Exploration of Disaccharide Conformations by Molecular Mechanics. *J. Mol. Struct. THEOCHEM* **1993**, *286*.

189. Gonzalez-Outeiriño, J.; Kirschner, K.N.; Thobhani, S.; Woods, R.J. Reconciling Solvent Effects on Rotamer Populations in Carbohydrates - A Joint MD and NMR Analysis. *Can. J. Chem.* **2006**, *84*, 569–579.

190. Martínez, J.D.; Infantino, A.S.; Valverde, P.; Diercks, T.; Delgado, S.; Reichardt, N.C.; Ardá, A.; Cañada, F.J.; Oscarson, S.; Jiménez-Barbero, J. The Interaction of Fluorinated Glycomimetics with Dc-Sign: Multiple Binding Modes Disentangled by the Combination of NMR Methods and MD Simulations. *Pharmaceuticals* **2020**, *13*, 1–18.

191. Corzana, F.; Motawia, M.S.; Hervé du Penhoat, C.; Perez, S.; Tschampel, S.M.; Woods, R.J.; Engelsen, S.B. A Hydration Study of (1→4) and (1→6) Linked α-Glucans by Comparative 10 Ns Molecular Dynamics Simulations and 500-MHz NMR. *J. Comput. Chem.* **2004**, *25*.

192. Sattelle, B.M.; Almond, A. Shaping up for Structural Glycomics: A Predictive Protocol for Oligosaccharide Conformational Analysis Applied to N-Linked Glycans. *Carbohydr. Res.* **2014**, *383*, 34–42.

193. Almond, A.; Bunkenborg, J.; Franch, T.; Gotfredsen, C.H.; Duus, J. Comparison of Aqueous Molecular Dynamics with Nmr Relaxation and Residual Dipolar Couplings Favors Internal Motion in a Mannose Oligosaccharide. *J. Am. Chem. Soc.* **2001**, *123*, 4792–4802.

194. Martín-Pastor, M.; Canales, A.; Corzana, F.; Asensio, J.L.; Jiménez-Barbero, J. Limited Flexibility of Lactose Detected from Residual Dipolar Couplings Using Molecular Dynamics Simulations and Steric Alignment Methods. *J. Am. Chem. Soc.* **2005**, *127*.

195. Chalmers, G.; Glushka, J.N.; Foley, B.L.; Woods, R.J.; Prestegard, J.H. Direct NOE Simulation from Long MD Trajectories. *J. Magn. Reson.* **2016**, *265*, 1–9.

196. Bryce, R.A.; Hillier, I.H.; Naismith, J.H. Carbohydrate-Protein Recognition: Molecular Dynamics Simulations and Free Energy Analysis of Oligosaccharide Binding to Concanavalin A. *Biophys. J.* **2001**, *81*.

197. Tsai, D.H. The Virial Theorem and Stress Calculation in Molecular Dynamics. *J. Chem. Phys.* **1979**, *70*, 1375–1382.

198. Overhauser, A.W. Polarization of Nuclei in Metals. *Phys. Rev.* **1953**, *92*.

CHAPTER 2

Objectives & Route Map of the Thesis

This section aims at providing a general vision of the projects addressed during the predoctoral research period, which are gathered in the following chapters. The main objectives of these investigations will be exposed, as well as the link between the following chapters presented in this dissertation, including a brief introduction to each of them.

OBJECTIVES

The main objective pursued in this Thesis has been extending the scope of the applications of ligand-based NMR strategies to the study of molecular recognition processes involving carbohydrates and lectins. In particular, we were interested on the use of ^{19}F-NMR spectroscopy (although complemented by ^{1}H, ^{13}C and/or ^{15}N-NMR experiments).

Then, as a complementary objective, we aimed at revealing important features of the mechanisms implicated in different recognition processes taking place between lectins and glycomimetics, and how the chemical modifications of the natural sugar could affect the binding event.

Finally, it was also our goal to conceive and test the scope of new computational protocols, based on existing methods (for example, MD simulations and CORCEMA-ST), to deepen into the mechanisms of the molecular recognitions processes and fitting our experimental designs.

These scientific objectives have been intimately related to the training aims, focused on the acquisition of high-level knowledge in state-of-the-art NMR methods and computational chemistry approaches, and their applications to problems of biological interest.

ROUTE MAP OF THE THESIS

Carbohydrate-mediated interactions are involved in a variety of recognition events in nature, either related with normal physiological functions occurring in bio-organisms or in disease. Therefore, a profound understanding of the mechanisms taking part in these interactions is of paramount importance to eventually be able to modulate them. With this in mind, we embarked on their study by means of different ^{19}F-NMR ligand-based techniques and computational methods to complement the experimental data. The results and conclusions reached during these investigations are gathered in the following chapters. Herein, a brief summary of their content is provided.

In Chapter 3, entitled *Fluorinated Carbohydrates as Lectin Ligands: Simultaneous Screening of a Monosaccharide Library and Chemical Mapping by ^{19}F NMR Spectroscopy*, it is described a strategy based on the use of simple ^{19}F-NMR

relaxation filtered experiments that allows the simultaneous screening of a library of mono-fluorinated monosaccharides against different lectins, and assessing which are the sugar hydroxyl groups crucial for binding. The methodology relies on the rational design of the fluorinated-sugar library, which permits unequivocally deducing the key OH groups in the overall interaction with the lectin. Thus, the methodology extends the information gained with standard screening procedures, in which only the binders in a pool of compounds are identified, but the fine details of their interactions with the receptor remains unveiled.

The application of this methodology to study DC-SIGN interactions is shown in Chapter 4. DC-SIGN is an important C-type lectin receptor which specifically recognizes highly-glycosylated patterns at the surface of pathogens and host glycoprotein, including HIV and Ebola viruses. The results described in Chapter 4, entitled *Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates*, revealed a new binding mode for Man moieties to DC-SIGN, which was not observed before in X-Ray crystallographic structure. Strikingly, we found that this binding mode imitates the binding architecture of Fuc with DC-SIGN, which is the preferred ligand of the lectin.

Lastly, in Chapter 5, named *The Interaction of Fluorinated Glycomimetics with DC-SIGN: Multiple Binding Modes Disentangled by the Combination of NMR Methods and MD Simulations*, the study of the interaction between a branched trifluorinated derivative (and its two forming di-fluorinated branches individually) of the ubiquitous trimannoside core with DC-SIGN is adressed. It had been previously reported that the ligand binding occurred preferentially through the terminal D-Man unit of the 1→3 branch, and the possibility of multiple binding modes taking place in solution proposed. Although ^{1}H-STD-NMR seemed an attractive technique for this task, unfortunately the huge signal overlapping of the three Man units prevented obtaining reliable STD profiles. Thanks to the novel 2D-STD-TOCSYreF experiment, it was possible to acquire STD spectra with exquisite signal dispersion, although the STD data obtained were still difficult to interpret solely in this qualitative manner. Therefore, a protocol in which the theoretical STDs of MD simulation trajectories of the different ligand-lectin complexes are estimated by using CORCEMA-ST was proposed. The calculated STDs were then submitted, together with the experimental STD data, to BM-Mixer calculations. BM-Mixer is an in-house program designed to find the proportion of each ligand-protein complex in solution that best fit the experimental STD.

CHAPTER 3

Fluorinated carbohydrates as lectin ligands: simultaneous screening of a monosaccharide library and chemical mapping by ^{19}F NMR spectroscopy

FLUORINATED CARBOHYDRATES AS LECTIN LIGANDS: SIMULTANEOUS SCREENING OF A MONOSACCHARIDE LIBRARY AND CHEMICAL MAPPING BY ^{19}F NMR SPECTROSCOPY

Reproduced with permission from the authors: Ana I. Manzano, Eva Calviño, Ana de Diego, Borja Rodriguez de Francisco, Cecilia Romanò, Stefan Oscarson, Oscar Millet, Hans-Joachim Gabius, Jesús Jiménez-Barbero, Francisco J. Cañada. 'Fluorinated Carbohydrates as Lectin Ligands: Simultaneous Screening of a Monosaccharide Library and Chemical Mapping by ^{19}F NMR Spectroscopy', *The Journal of Organic Chemistry* **2020** *85* (24), 16072-16081.

Abstract: Molecular recognition of carbohydrates is a key step in essential biological processes. Carbohydrate receptors can distinguish monosaccharides even if they only differ in a single aspect of the orientation of the hydroxyl groups or harbor subtle chemical modifications. Hydroxyl-by-fluorine substitution has proven its merits for chemically mapping the importance of hydroxyl groups in carbohydrate-receptor interactions. ^{19}F NMR spectroscopy could thus be adapted to allow contact mapping together with screening in compound mixtures. Using a library of fluorinated glucose (Glc), mannose (Man) and galactose (Gal) derived by systematically exchanging every hydroxyl group by a fluorine atom, we developed a strategy combining chemical mapping and ^{19}F NMR T_2 filtering-based screening. By testing this strategy on the proof-of-principle level with a library of 13 fluorinated monosaccharides to a set of three carbohydrate receptors of diverse origin, i,e the human macrophage galactose-type lectin, a plant lectin, *Pisum sativum* agglutinin, and the bacterial Gal-/Glc-binding protein from *Escherichia coli*, it became possible to simultaneously define their monosaccharide selectivity and identify the essential hydroxyls for interaction.

1. INTRODUCTION

Molecular recognition events are at the heart of health and disease. From the chemical perspective, understanding the details of interactions for the underlying functional pairings may provide key information for innovative drug discovery and design. In this context, carbohydrate oligomers (saccharides, glycans) are ubiquitous in nature, commonly presented on cell surfaces by protein and lipid scaffolds [1-4]. Structurally, an exceptionally large diversity can be generated by simply exploiting permutations of linkage points and anomeric position at each glycosidic linkage [5]. As a consequence, glycans are "ideal for generating compact units with explicit informational properties"[6], and this information is being disclosed to be 'read' and 'translated' into (patho)physiological processes by lectins [4,5]. Thus, the analysis of glycan-lectin recognition has become a topic with biomedically promising perspective [7,8] and a fructiferous ground to enhance the symbiosis of Chemistry and Biology as Lemieux asked for [9].

From the molecular recognition perspective, different approaches have been tested to examine the relevance of hydroxyl groups from saccharide units in binding to receptors. One of them extensively applied rests on screening a given set of available closely related saccharides that display different stereochemistry and/or substitutions at a certain site within the sugar ring [10-19]. This protocol synthetically eliminates or modifies hydroxyl groups (deoxygenation, methylation, exchange by halogens)[20]. In particular, hydroxyl-by-fluorine substitution has been used to trace key hydroxyl groups for contact with either lectins, antibodies, transporters or enzymes [21].

Fluorine can be considered as an isosteric mimic of the hydroxyl group, although without the capacity to act as hydrogen-bond donor and with a diminished hydrogen-bond acceptor competence [22-24]. Additionally, its particular physicochemical properties [25] introduce electronic and polar-hydrophobic effects [26]. Indeed, fluorine modulates the population of the conformational space [27,28] and lipophilicity of fluorine-containing carbohydrates [29]. Smart use of these features has already allowed developing new molecules that efficiently act as substrates [30] and inhibitors of glycosidases [31,32]. Advances for fluorine introduction into organic molecules have made available a large variety of mono- [33] and polyfluorinated saccharides, [33-40] which are highly attractive as chemical probes from different point of views [21]. It is also well known that fluorine-containing molecules are extensively used in bioorganic and medicinal chemistry [23,41]. Many of these studies have driven the development of ^{19}F

NMR-spectroscopy methodologies as valuable tools to study molecular recognition events or to screen compounds libraries [42,43].

In this context, we and others have applied ^{19}F-observed NMR strategies to study glycan-protein interactions by means of saturation transfer difference (STD NMR-spectroscopy) measurements using 1D [44] and 2D [34] experimental designs, by monitoring chemical shifts perturbations and exchange kinetics [45,46], by observing line broadening of the ^{19}F NMR signals, [47,49] or by employing relaxation filtering protocols [50,51].

Herein, we propose a robust and general method to efficiently pick up and study the interactions of a library of fluorinated sugars with a given receptor. Taking advantage of the large chemical shift range of the ^{19}F nucleus and its sensitivity, monitoring sugar-protein interactions by a panel of 13 different monofluorinated sugars (with up to 26 well resolved ^{19}F NMR signals considering the presence of the α and β-anomers for each sugar, Figure 1, Table S1) provides information on the selectivity of the binding event in a single setup. This methodology extends the applicability of the reported T_2-filtering strategy, and overcomes the limits of ^{1}H NMR resolution (see spectrum in Figure 1b).

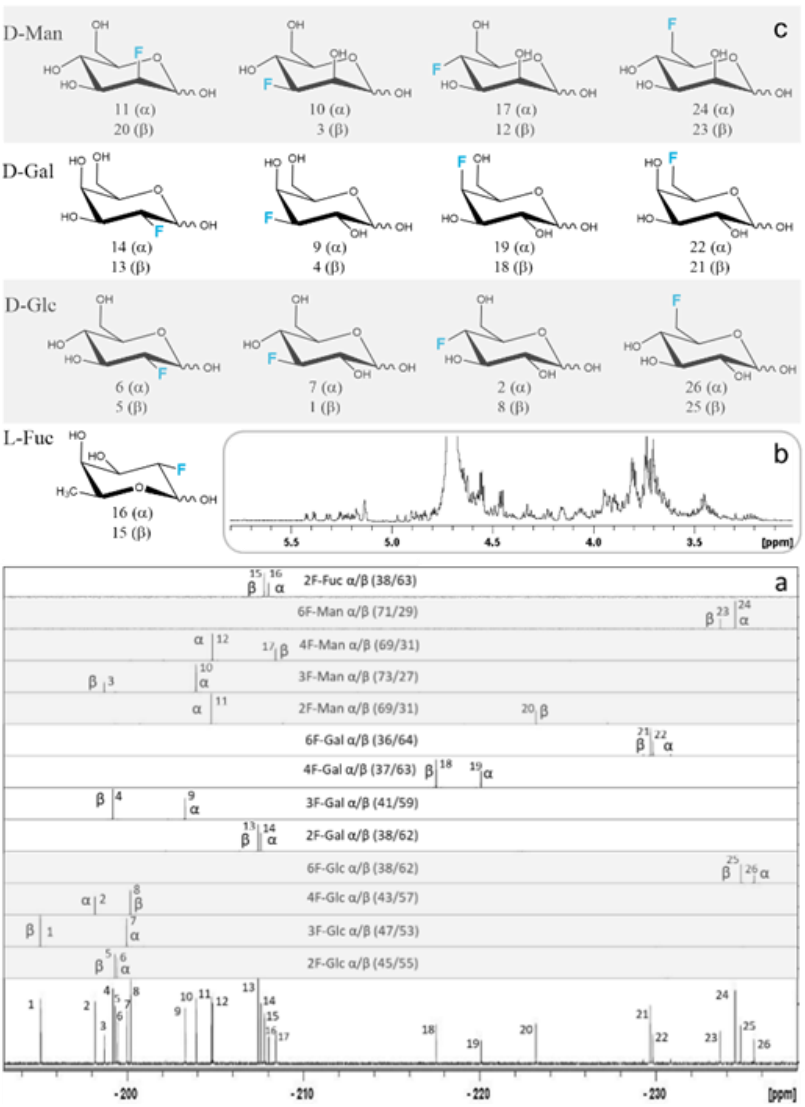


Figure 1. a) ^{19}F NMR (^{1}H decoupled) spectra recorded for each individual monofluorinated monosaccharide as anomeric mixture. The α/β anomeric ratios are given between brackets. Lower panel, ^{19}F-NMR spectrum of the full library. Each peak is numbered from lower to higher field. b) ^{1}H NMR spectrum of the mixture of the 13 monosaccharides. c) Representation of the structures of the different monodeoxy-monofluorinated monosaccharides present in the library. The corresponding peak number for each anomer in the ^{19}F NMR spectrum is indicated.

As proof-of-concept, two lectins and a sugar transporter of diverse origins and selectivities have been chosen: the human Macrophage Galactose-type Lectin (MGL, CLEC10A, CD301), a C-type lectin binding *N*-acetylgalactosamine in *O*-glycans (T_n antigen, CD175) and in *N*-glycans (LacdiNAc) [52-54]; *Pisum sativum* agglutinin (PSA), a plant lectin selective for α-mannopyranosides and -glucopyranosides [15,55]; and the

glucose/galactose-binding protein (GGBP) [56-58], a bacterial sensor for free monosaccharides. From the analysis of data from simple 1D ^{19}F NMR experiments by applying transversal relaxation filters, screening and chemical mapping are simultaneously achieved. In essence, information on the monosaccharide selectivity for a particular sugar receptor is obtained (screening) together with the direct identification of hydroxyls that are essential for binding and those that can be chemically substituted or modified without critically compromising the binding event (chemical mapping).

2. RESULTS

Three different types of sugar receptors are deliberately selected herein to illustrate broad applicability, i.e. a human lectin involving Ca^{2+} for direct ligand contact, a plant agglutinin, and a bacterial sugar transporter.

MGL belongs to the C-type lectin family characterized by containing a calcium cation at the binding site, directly involved in carbohydrate recognition by coordination bonding [52,53]. MGL, like the hepatic asialoglycoprotein receptor, is a transmembrane protein with the carbohydrate recognition domain (CRD) on top of its extracellular stalk that oligomerizes to trimers [59]. GalNAc in α/β linkage are the preferred ligands (K_D = 12 μM for Me α-*N*-acetyl-galactosaminide, Me α-GalNAc), galactose being a weaker binder (K_D = 0.9 mM for Me α-Gal) [53]. To perform the recognition studies, the soluble extracellular ectodomain containing the CRD was used. It is known that the Ca^{2+} in the binding site makes contact to its Gal/GalNAc ligands through the equatorial/axial OH-3 and OH-4 groups [53]. In order to study the importance of each hydroxyl group of the Gal moiety, the binding of the four possible monodeoxy-monofluorinated Gal analogues (at positions 2, 3, 4, and 6), keeping free the anomeric position, was tested to detect those hydroxyl-to-fluorine substitutions that impair binding. A similar strategy, using an extended mixture of mono- and polyfluorinated galactopyranosides and applying a diversity of techniques, has allowed to identify OH-3 and 4 as the coordinating groups in a calcium-dependent bacterial galactophilic lectin [33].

Since every monosaccharide exists as a mixture of its α and β anomers in equilibrium, eight different molecules are present in solution. The ^{19}F NMR spectrum of the mixture is pleasingly simple, just showing eight individual ^{19}F NMR signals (Figure 2a), one for each monofluorinated Gal anomer in the mixture. Their intensities are governed by the anomer ratio at equilibrium [60].

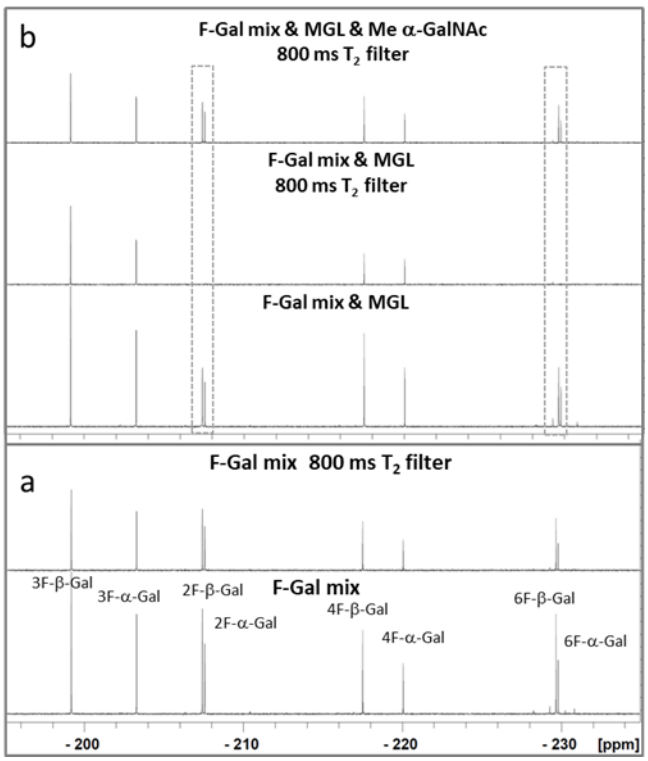


Figure 2. a) ^{19}F NMR (^{1}H decoupled) spectra of the F-Galactose mixture (2F-Gal (0.5 mM); 3F-Gal (0.68mM), 4F-Gal (0.37 mM) and 6F-Gal (0.45 mM) in absence of the protein, without T_2 relaxation filter (lower panel), and after applying a 800 ms T_2 relaxation filter (upper panel). b) ^{19}F NMR (^{1}H decoupled) spectrum of the same mixture in the presence of 0.015 mM MGL before (lower panel) and after (mid panel) the application of a 800 ms T_2 relaxation filter. The upper panel shows the spectrum after the addition of 0.5 mM Me α-GalNAc with the same relaxation filter.

The transverse relaxation time (T_2) for each compound was measured in the free state in the absence of lectin, ranging between 1.2 to 1.8 s for Gal derivatives (Table S1). The monofluorinated Gal mixture was added to a solution of MGL and the T_2 filtering strategy [50,61,62] was applied to identify the binders. Briefly, those molecules that bind to the protein drastically change their hydrodynamic behavior in the bound state, and thus their rotational motion correlation time increases towards that of the large protein, with a concomitant reduction in T_2. Additionally, the effective transverse relaxation is also affected by the kinetics of the chemical exchange process between the free and bound states, further reducing the observed T_2, especially if the system no longer follows the fast chemical exchange regime. This reduction in T_2, which is in first instance manifested in standard 1D NMR spectra as signal broadening, can be easily transformed into a signal-intensity reduction by the application of a standard Carr–Purcell–Meiboom–Gill (CPMG) spin echo pulse train sequence before acquisition. The filtered

NMR spectrum displays the NMR signals of the binders significantly reduced or even suppressed, compared to those of the unbound compounds.

Figure 2b shows the comparison of the ^{19}F NMR spectra recorded for the mixture of monofluorinated galactose derivatives in the presence of MGL, (lower panel) with that obtained by applying a spin-echo filter of 800 ms (central panel). The drastic reduction of the intensity of signals in presence of protein (Figure 2b central panel) relative to the experiment in its absence (Figure 2a upper panel) corresponding to 2F-Gal (β = 2% , α = 4%) and 6F-Gal (β = 1% , α = 7%) is clearly observed, while the NMR signals obtained for 3F-Gal (β = 57% , α = 73%) and 4F-Gal (β = 98% , α = 80%) are significantly less altered.

To confirm that this selective signal reduction is due to the binding of the ^{19}F-containing Gal entities to the CRD, a competition experiment was performed by adding Me α-GalNAc to the mixture. The recovery of the 2F-Gal and 6F-Gal signals was evident, indicating that they are displaced from the binding site by the strong competitor (Figure 2b, upper panel).

Drawing a conclusion from chemical mapping [30,49] is straightforward: the modification of either hydroxyl at 3 or 4 eliminates a coordination bond in the interaction between sugar and Ca^{2+}. Therefore, the signals corresponding to 3F-Gal and 4F-Gal are not affected by the lectin and do not show significant signal reduction. On the contrary, the hydroxyls at positions 2 and 6 can be substituted by fluorine. Their ^{19}F NMR signals are clearly reduced in the presence of the MGL due to binding.

Since the broad dispersion of ^{19}F NMR chemical shifts of the four anomeric pairs of the monofluorinated Gal analogues is more than 30 ppm (between -199 and -230 ppm), the feasibility to test a broad panel of monofluorinated monosaccharides was envisioned. Thus, the four monodeoxy-monofluorinated D-glucoses (2F-Glc, 3F-Glc, 4F-Glc and 6F-Glc) and D-mannoses (2F-Man, 3F-Man, 4F-Man and 6F-Man) together with 2-deoxy-2-fluoro-L-fucose (2F-Fuc) were added to provide a library with 13 different anomeric pairs of monofluorinated monosaccharide (Figure 1). All compounds were available from commercial sources except 6F-Man that was chemically synthesized (see experimental section).

This mixture with the 13 fluorinated monosaccharide anomeric pairs gives a very crowded ^{1}H-NMR spectrum (Figure 1b). In contrast, its proton-decoupled ^{19}F NMR spectrum presents well-resolved individual signals for each of the 26 different molecules in the sample, which are spread over 40 ppm (Figure 1a). Thus, the extended compound library, including the monofluorinated Gal, Glc and Man analogues, was now tested with MGL, applying again the T_2-filtering strategy (Figure 3). For qualitative visualization of the NMR signals affected by the lectin, the obtained filtered ^{19}F NMR

spectrum was subtracted from the non-filtered one following the protocol described in the experimental section (a correction factor *f* was applied to the filtered spectrum to account for the signal reduction due to transversal relaxation unrelated with the presence of the protein). Only the ^{19}F NMR signals corresponding to binders should appear in the difference spectrum. Indeed, the peaks corresponding to 2F-Gal (peaks 13 and 14) and 6F-Gal (peaks 21 and 22) are clearly displayed in the difference spectrum (Figure 3) in accordance with the results of the experiment described above for the smaller sized Gal library.

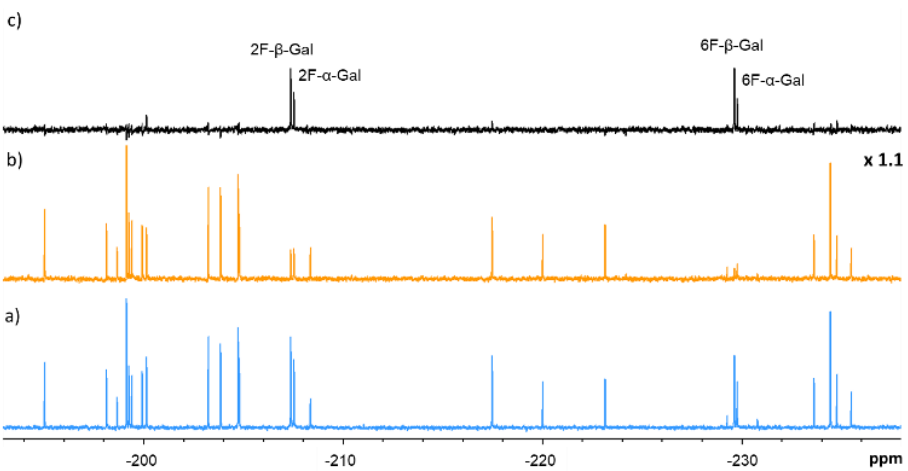


Figure 3. ^{19}F NMR (^{1}H-decoupled) T_2-filtered spectra recorded for the fluorinated monosaccharide library (Man, Glc, and Gal analogues) in the presence of MGL (30 μM). a) Spectrum acquired with a short 8 ms T_2 filter. b) spectrum acquired with long 160 ms T_2 filter factored 1.1 times. c) Difference spectrum.

Given the encouraging results for the first system, the suitability of the monofluorinated monosaccharide library for simultaneous ligand screening and chemical mapping was further tested with two other types of carbohydrate receptors with different sugar selectivities.

PSA. *Pisum sativum* agglutinin (PSA), a leguminous lectin with a "jelly roll" fold [63], was also tested. PSA is selective for Man/Glc-containing oligosaccharides without involvement of Ca^{2+} in contacts to the sugar, but displays weak affinity for single monosaccharides: 0.53 and 1.15 mM for methyl α-mannopyranoside (Me α-Man) and methyl α-glucopyranoside (Me α-Glc), respectively [55].

Several ^{19}F NMR peaks diminished (Figure 4a,b) when the T_2 filter was applied. Those present in the difference spectrum (Figure 4c) correspond to 2F-Glc (peaks 5 and 6) and 2F-Man (11 and 20), 3F-Glc (1 and 7), 3F-Man (3 and 10), 6F-Glc (25 and 26) and 6F-Man (23 and 24). Neither 4F-Glc nor any Gal derivatives were observed in the difference ^{19}F NMR spectrum. However, a limitation of the method was detected. The difference ^{19}F NMR

spectrum also displays "false positives" corresponding to fast relaxing signals (see below), especially when the applied T_2 filter is long enough (720 ms in this experiment). This was the case for 4F-α-Man (peak 12, $^{19}F\text{-}T_{2,free} = 0.7$ s), whose signal relaxation was significantly faster than that of the other signals of the molecules present in the mixture (Table S1).

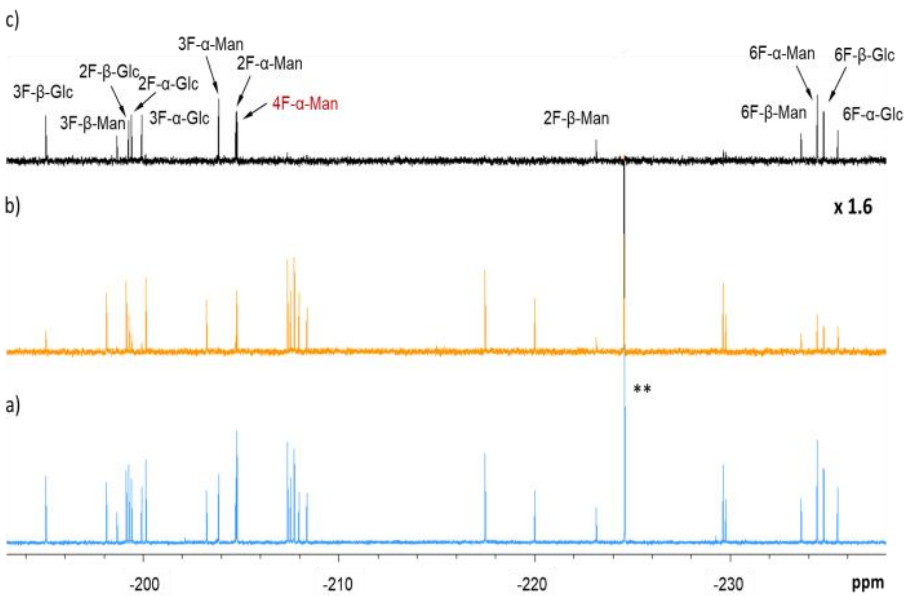


Figure 4. ^{19}F NMR (^{1}H-decoupled) T_2-filtered spectra recorded for the fluorinated monosaccharide library in the presence of PSA (25 μM, ligand/protein ratio around 36:1). a) Spectrum acquired with a short 8 ms T_2 filter. b) spectrum acquired with 720 ms of T_2 filter factored 1.6 times. c) Difference spectrum. The peak corresponding to 4F-α-Man, in red, is a "false positive" (see text). **2-fluoroethanol added as internal reference.

To confirm specific binding, the difference ^{19}F NMR spectrum was again complemented with competition experiments (Figure 5) in the presence of a known ligand (Me α-Man). The signals corresponding to the 2F- and 3F-Man/Glc derivatives were now clearly observed, indicating that 2F/3F-Man/Glc are indeed displaced from the binding site by Me α-Man. On the contrary, no difference in the intensities of the 4F-α-Man signal was observed in the absence and presence of the competitor (Figure 5), indicating that this molecule is not a binder of the lectin. Very likely, the signal observed in the difference spectrum described above is due to the intrinsic fast relaxation of 4F-α-Man ($^{19}F\text{-}T_{2,free} = 0.70$ s, Table S1). Analogous results were observed in the absence of the lectin, i.e, T_2 of 4F-Man is not affected by the presence of PSA. A mixed situation took place for the 6F-Man and 6F-Glc derivatives (Figure 5). In these cases, the observed signals in the difference experiment are due to ligand binding and to fast relaxation. In fact, the intrinsic $T_{2,free}$ for the corresponding ^{19}F signals of the 6F-Man ($^{19}F\text{-}T_{2,free} = 1.00$ s and $^{19}F\text{-}T_{2,free}$

= 0.82 s for the β and α anomers, respectively) and 6F-Glc (^{19}F-$T_{2,free}$ = 0.92 s and ^{19}F-$T_{2,free}$ = 0.90 s for the β and α anomers, respectively) derivatives is also rather short. On the other side, indeed, the initial decrease in signal intensity induced by the presence of the protein was subtle, but recovery was almost complete after addition of a small concentration of competitor, thus also confirming affinity, although likely weaker.

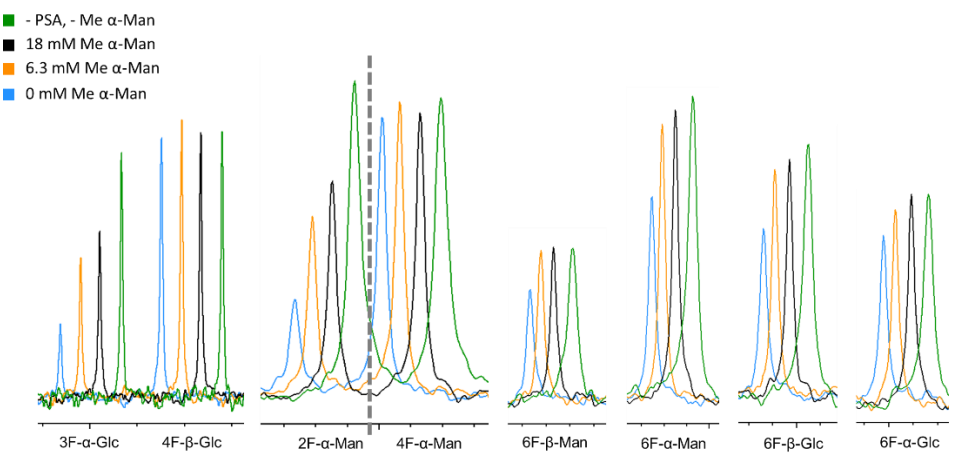


Figure 5. Close-up view of selected ^{19}F NMR peaks recorded in the ^{19}F NMR T_2-filtered spectra (720 ms) of the fluorinated monosaccharide library in the absence of lectin (green), in the presence of PSA (blue), and when adding different concentrations of Me α-Man.

Interestingly, it has been described that hydroxyls at positions 2 and 3 of glucose can be substituted by fluorine while retaining binding by PSA; however, when F is at the 6 position the reported binding was minimal and modifications at OH-4 abolished the binding [15,64], supporting the results presented here regarding PSA selectivity.

GGBP. The third receptor is the bacterial GGBP. It is involved in chemotaxis and sugar transport in bacteria and has a very high affinity for Glc (0.04 μM) and Gal (0.13 μM) [65,66] typical for bacterial binders of free monosaccharides. Its structure consists in two globular Rossman fold domains and, differently to the tested lectins, GGBP presents a deep binding pocket at the hinge connecting and closing both domains around the monosaccharide ligand [56,57]. When GGBP was added to the monofluorinated monosaccharide library, the signals belonging to Glc and Gal molecules with F atoms at positions 4 or 6 showed reduced peaks intensities in the T_2-filtered spectrum (Figure 6). On the contrary, those signals corresponding to Glc and Gal moieties substituted at either positions 2 or 3 were not affected by the presence of GGBP. These evidences indicate that the OH groups at those 2 and 3 positions are required for the binding

to take place and cannot be substituted by a fluorine atom. In the difference ^{19}F NMR spectrum, signals for 2F-α-Man, 4F-α-Man and 6F-α-Man also appear (Figure 6c). However, when Glc was added to the library/receptor mixture as competitor, the corresponding signals of those fluoromannoses were not recovered (Figure 7). In the cases of 4F-α-Man and 6F-α-Man, as for PSA, this behavior is again due to the intrinsic fast relaxation of the 4F-α-Man and 6F-α-Man ^{19}F signals. Interestingly, 2F-α-Man is a special case, its signal reduction only takes place in the presence of the protein and it is not affected by glucose (Figure 7). This result suggests 2F-α-Man interacts with GGBP but at a location different of the canonical sugar-binding site.

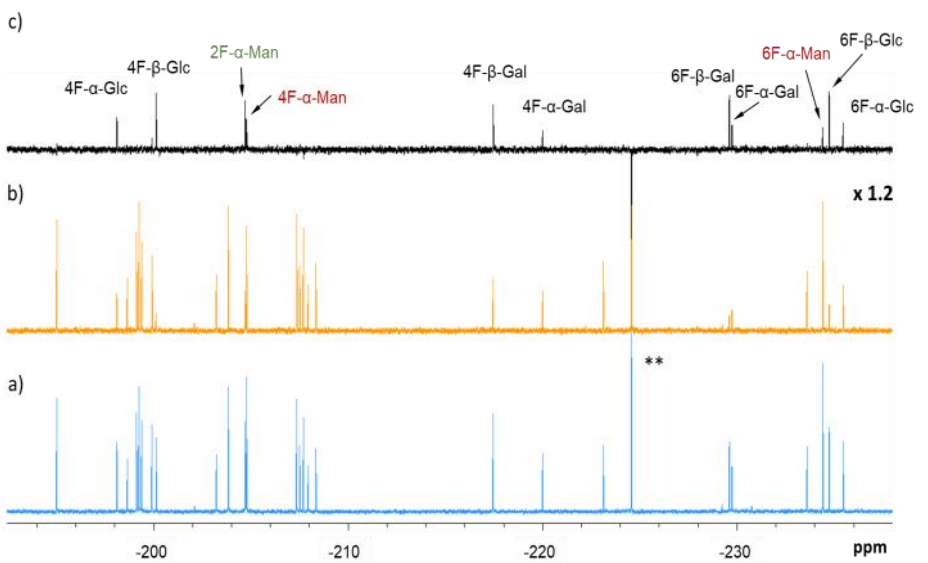


Figure 6. ^{19}F NMR (^{1}H-decoupled) T_2-filtered spectra recorded for the fluorinated monosaccharide library in the presence of GGBP (25 μM, ligand/protein ratio ca. 36:1). a) Spectrum acquired with a short 8 ms T_2 filter, b) spectrum acquired with long 400 ms of T_2 filter factored 1.3 times, c) difference spectrum. For peaks labeled in green and red see text. ** 2-fluoroethanol added as internal reference.

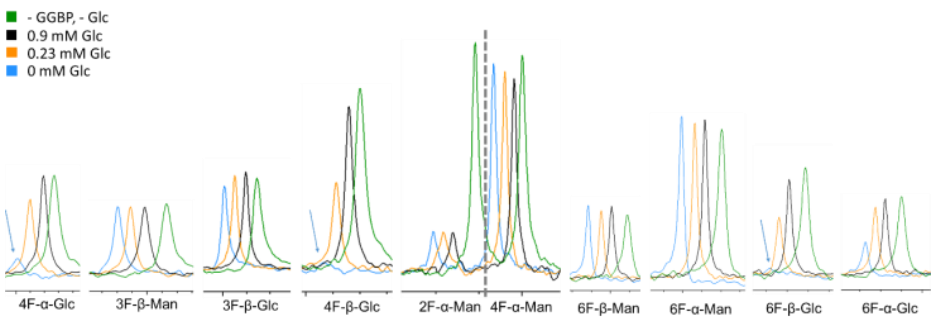


Figure 7. Close-up view of selected ^{19}F NMR peaks recorded in the ^{19}F NMR T_2-filtered spectra of the fluorinated monosaccharide library in the absence of receptor (green), in the presence of GGBP (blue), and when adding different concentrations of the Glc competitor (0.23 mM orange and 0.9 mM black). In all cases, the NMR experiments were acquired using 720 ms of T_2 filter.

3. DISCUSSION

The tested screening method is based on the dramatic differences in transverse relaxation observed for binders within a library of fluorinated monosaccharides, when acquiring NMR spectra in the absence or presence of a carbohydrate-binding protein. The transversal relaxation time is related to the rotational motion correlation time of the molecule, and it sharply decreases as the correlation time increases. When monosaccharides interact with a large receptor, they adopt the correlation time of the macromolecule during the time the complex is associated, and thus undergo a critical decrease of their T_2. This change in T_2 may be followed in a straightforward manner under fast exchange conditions between bound and free states. Thus, only a single ^{19}F NMR signal appears in the spectrum at the averaged chemical shift of the exchanging states weighted by their corresponding molar fractions. In fact, the observed effective T_2 also depends on the kinetics of the exchange between the free and bound forms. Both rotational motion and exchange effects add together in the T_2 filtering strategy and allow the efficient detection of medium to low-affinity binders (from low micromolar to millimolar K_D), even using high ligand/protein ratios [62]. The application of the CPMG-based T_2 filtering scheme is fairly straightforward and usually a reasonable number of spin-echo loops before acquisition is sufficient to obtain highly sensitive NMR spectra with the required information discriminating binders from non-binders. From the practical perspective, the current library renders very well resolved ^{19}F NMR spectra with separated signals for all different monosaccharide moieties in the mixture. Obviously, other fluorinated saccharides could well be added to the mixture increasing the screening power of the concept. As an added value

for the ^{19}F observation, the experiments do not require any deuterated buffer, thus simplifying the experimental setup.

Regarding the screening process, in first instance, and assuming that all ^{19}F nuclei in the library have similar T_2 relaxation times when free in solution, it should be possible to qualitatively visualize those signals that are affected by the protein. To do so, a difference NMR spectrum is obtained by substracting the spectrum recorded using a short spin-echo delay from a second one measured employing a longer delay. However, the ^{19}F NMR signals of some molecules, such as 4F-α-Man, 6F-α/β-Man and 6F-α/β-Glc (Table 1), relax significantly faster ($T_2 < 1$s) than the others ($T_2 > 1.2$ s) and their peaks consistently appear in the difference spectrum when long spin-echo relaxation delays are used. Therefore, to unambiguously assess the existence of specific binders at the carbohydrate-binding site, the difference ^{19}F spectrum should be complemented with the information provided by additional competition experiments carried out by adding a known ligand of the lectin. The comparison of the recovered ^{19}F NMR signals in the presence of an excess of the competitor in the lectin/library sample can be expressed as the ratio of signal intensities, $I_{t(+C/-C)}$, measured in spectra acquired with a relaxation filter t in presence (+C) and absence (-C) of competitor (C), thus highlighting the specific binders (Figure 8).

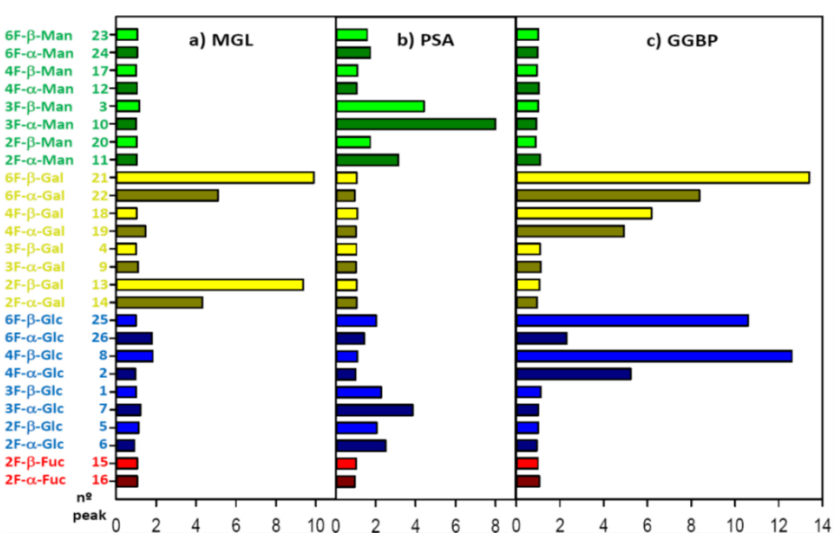


Figure 8. Competition experiments and ^{19}F NMR T_2-filtered spectra for the analysis of the interaction of the fluorinated monosaccharides library with the different lectins. The x axis corresponds to the signal intensity recovery, $I_{t(+C/-C)}$, expressed as the ratio between the relative decay at time t in the presence (+) and absence (-) of competitor (C). From left to right: a) MGL, T_2 filter t = 720 ms. MGL (30 μM) and competitor Me α-GalNAc (0.9 mM)b) PSA, T_2 filter t = 720 ms, PSA (25 μM) and competitor Me α-Man (18 mM) and c) GGBP. T_2 filter t = 400 ms, GGBP (25 μM) and competitor Glc, at (0.9 mM). In all cases the mixture with 0.9±0.3 mM of each monosaccharide was used.

At least qualitatively, these signal recovery data allow distinguishing the specificities of the three sugar receptors, correlating them with their known monosaccharide selectivity: MGL only recognize Gal moieties, PSA binds Glc and Gal analogues, while GGBP interacts with Glc and Gal monosaccharides. Additionally, information on the selectivity for the anomeric configuration can be gleaned from the signal recovery data in the T_2-filtered competition experiments. For instance, for PSA, the α-anomers show a higher recovery ratio than their corresponding β-anomers in accordance with previous reports. On the contrary, based on X-ray and NMR structural data, GGBP has been described to display specificity for β-anomers[57]. The recovery ratio data here presented are in agreement with that selectivity but also show that the α-anomers are binders, as previously suggested by means of ligand-binding kinetic experiments [65,66].

As mentioned above, the OH by F substitution has been extensively used in carbohydrate chemistry to map the key hydroxyl groups of a given sugar that are involved in their recognition by lectins, antibodies, transporters or enzymes [10-20]. The methodology presented herein, which employs a rationally assembled collection of monofluorinated monossaccharides for which their hydroxyl groups have been systematically substituted by fluorine atoms, allows dissecting chemical mapping information regarding the importance of each individual hydroxyl group in the interaction with its receptor. For MGL, the same experiment allows identifying its selectivity for Gal moieties and simultaneously shows that hydroxyls at positions 3 and 4 are essential to keep the interaction ability of the Gal analogue, while hydroxyls 2 and 6 can be modified still maintaining binding to MGL. PSA can recognize Man and Glc, epimers at position 2. Thus, the orientation of OH-2, axial in Man, equatorial in Glc is not essential for binding and, consequently, both fluorinated epimers 2F-Man and 2F-Glc are recognized. Moreover, it can be inferred that modifications at OH-3 are tolerated, as F to OH substitution at this position does not block binding to the lectin. On the contrary, OH-4 is essential for binding while the modification at position 6 still sustains a weak interaction. Finally, for GGBP, which also recognizes two monosaccharides that share the equatorial configuration at C2, i.e. Glc and Gal, OH-2 and OH-3 are necessary for binding, while OH-4 (either axial in Gal or equatorial in Glc) and OH-6 can be modified. Thus, the binding pattern is completely opposite to that observed for PSA. Additionally, for GGBP, the possibility of a secondary binding site has been deduced, given the existence of binding to 2F-α-Man (see Figure 6, 2F-α-Man is marked in green) and the fact that this interaction is not abolished by Glc (see Figure 7, the signal intensity of 2F-α-Man is not recovered after addition of Glc), the

canonical ligand of GGBP. The implications of this result remain to be explored.

Although T_2 filtering has been merely applied herein from a qualitative perspective, the obtained data clearly pave the way to perform further quantitative affinity studies. In fact, such values could be in principle deduced for each isolated monosaccharide from competition experiments, using a competitor with a known affinity constant [67].

In summary, using this ^{19}F NMR-based T_2-filtering strategy using a library of fluorinated monosaccharides generated through systematic OH-to-F substitutions allows i) defining sugar selectivity of the tested receptor, ii) detecting its anomer preference and iii) identifying the key hydroxyls for binding, distinguishing them from those that can be chemically modified in the quest to find new binders. Extending this approach to other saccharides (aminosugars and sialosides) and to synthetic libraries of disaccharides, will especially be attractive to screen a variety of carbohydrate-receptor families, on the way "from biology to drug target" [68]. In this sense, Siglecs, sialoside receptors proposed to act as "immune cell checkpoints in disease" [69], or the multifunctional galectins [70,71], look like exciting targets to start with. The versatility of the described strategy is evident: it shows applicability to lectins and sensor/transport proteins, and proved to be suitable to cover diverse selectivities and wide-ranging affinities, from sub-microM (40 nM for the Glc-GGBP complex) to over mM (1.15 mM for the Me β-Glc/PSA complex) dissociation constants. Thus, the method is robust and envisioned to find wide application.

4. EXPERIMENTAL SECTION

4.1. Materials

PSA was from commercial source (Sigma Aldrich-Merck) and dissolved in phosphate-buffered saline at pH 7.5. MGL ectodomain was recombinantly produced in *E.coli*, and routinely checked for purity and activity as previously described, including ascertaining GalNAc-inhibitable histochemical staining[53,72], the samples for NMR were prepared in deuterated Tris buffer (10 mM), containing $CaCl_2$ (1 mM) and NaCl (75 mM) at pH 7.5 by means of five ultrafiltration-dilution buffer exchange steps with a 10 kDa cut-off membrane. GGBP was expressed in *E. coli* and purified as previously described[56] and the samples prepared in 20 mM Tris, containing 150 mM NaCl and 10 mM $CaCl_2$ at pH 7.0. Protein concentrations were measured by UV spectrometry.

The monofluorinated monosaccharide mixtures were prepared from concentrated stock solutions of each individual monosaccharide depending

on their availability, either commercial or from synthesis. Final concentrations in the mixtures were centered around 0.5 mM or 0.9 mM, depending on the experiment, with variations in ±35% range. Given the intrinsic different equilibrium populations of the different anomers for a given monosaccharide, it is impossible to use the same concentration for each individual species.

Fluorinated monosaccharides: 2-deoxy-2-fluoro-glucose, 3-deoxy-3-fluoro-glucose, 4-deoxy-4-fluoro-glucose, 6-deoxy-6-fluoro-glucose, 2-deoxy-2-fluoro-galactose, 3-deoxy-3-fluoro-galactose, 4-deoxy-4-fluoro-galactose, 6-deoxy-6-fluoro-galactose, 2-deoxy-2-fluoro-mannose, 3-deoxy-3-fluoro-mannose, 4-deoxy-4-fluoro-mannose, and 2-deoxy-2-fluoro-fucose, 2-fluoroethanol, Me α-*N*-acetylgalactosaminide and Me α-mannopyranoside were from commercial sources (Sigma-Aldrich Merck, Spain; Carbosynth, UK). 6-Deoxy-6-F-mannose was synthetized as described in the supporting information, and its analytical data were consistent with literature values [73]. Characterization of the intermediates in reaction steps in the synthesis are described below. NMR peak assignments were made using correlation spectroscopy (COSY) and heteronuclear single-quantum coherence (HSQC).

Methyl 2,3,4-tri-*O*-benzoyl-6-deoxy-6-fluoro-α-D-manno pyranoside (2). Compound **1** [74] (50 mg, 0.098 mmol) was dissolved in dry CH_2Cl_2 (1.2 mL), then the solution was cooled to -78 °C and DAST (98 µL, 0.74 mmol) was slowly added drop-wise. The reaction mixture was kept at -78 °C for 30 min, then warmed to RT and left stirring overnight. The solution was then cooled to -20 °C and the reaction quenched with MeOH. The solvents were evaporated and the residue was purified by silica gel column chromatography (toluene/EtOAc, 98:2→8:2, v/v) to give 2 as a yellowish solid (37 mg, 0.07 mmol, 74%). R_f= 0.8, toluene/EtOAc 8:2; ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.14 – 8.05 (m, 2H, H_{Bz}), 8.02 – 7.94 (m, 2H, H_{Bz}), 7.86 – 7.78 (m, 2H, H_{Bz}), 7.64 – 7.59 (m, 1H, H_{Bz}), 7.56 – 7.46 (m, 3H, H_{Bz}), 7.46 – 7.36 (m, 3H, H_{Bz}), 7.29 – 7.23 (m, 2H, H_{Bz}), 5.94 – 5.85 (m, 2H, H-3, H-4), 5.68 (dd, *J* = 3.0, 1.8 Hz, 1H, H-2), 5.02 (d, *J* = 1.8 Hz, 1H, H-1), 4.64 (dt, *J* = 46.9, 3.7 Hz, 2H, H-6ab), 4.34 – 4.21 (m, 1H, H-5), 3.55 (s, 3H, OCH_3). ^{19}F NMR (376 MHz, $CDCl_3$): δ -231.70 (td, *J* = 47.2, 22.9 Hz). All analytical data were consistent with literature values [74].

1-*O*-acetyl-2,3,4-tri-*O*-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (3). Compound 2 (180 mg, 0.35 mmol) was dissolved in Ac_2O/AcOH (2:1, 3.5 mL). H_2SO_4 (4 µL, 0.07 mmol) was slowly added drop-wise at 0 °C and the mixture was stirred for 5 h. The reaction was then diluted with AcOEt and washed with sat. $NaHCO_3$ (aqueous). The organic layer was dried over $MgSO_4$, filtered and concentrated. The residue was purified by silica gel flash column chromatography (Toluene/EtOAc,

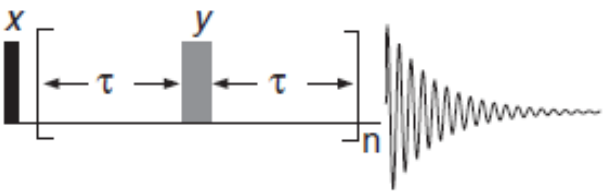
98:2→8:2, v/v) to give 3 as a white powder (150 mg, 0.28 mmol, 80%). R_f= 0.62, Tol/AcOEt 9:1; $[\alpha]_d^{20}$= -64.1 (c=0.6, $CHCl_3$); ^{1}H NMR (500 MHz, $CDCl_3$): δ 8.13 – 8.06 (m, 2H, H_{Bz}), 8.02 – 7.94 (m, 2H, H_{Bz}), 7.86 – 7.77 (m, 2H, H_{Bz}), 7.66 – 7.60 (m, 1H, H_{Bz}), 7.56 – 7.52 (m, 1H, H_{Bz}), 7.50 (t, *J* = 7.8 Hz, 2H, H_{Bz}), 7.47 – 7.43 (m, 1H, H_{Bz}), 7.40 (t, *J* = 7.8 Hz, 2H, H_{Bz}), 7.28 (t, *J* = 7.9 Hz, 2H, H_{Bz}), 6.39 (d, *J* = 2.0 Hz, 1H, H-1), 5.99 (t, *J* = 10.1 Hz, 1H, H-4), 5.92 (dd, *J* = 10.1, 3.3 Hz, 1H, H-3), 5.72 (dd, *J* = 3.4, 2.0 Hz, 1H, H-2), 4.69 – 4.56 (m, 2H, H-6ab), 4.34 (ddt, *J* = 23.3, 10.1, 3.2 Hz, 1H, H-5), 2.28 (s, 3H, $OCOCH_3$). ^{13}C{^{1}H} NMR (126 MHz, $CDCl_3$): δ 168.25 ($O\underline{C}OCH_3$), 165.70 (CO_{Bz}), 165.40 (CO_{Bz}), 165.35 (CO_{Bz}), 133.86 (C_{Bz}), 133.78 (C_{Bz}), 133.53 (C_{Bz}), 130.13 ($2C_{Bz}$), 129.94 ($2C_{Bz}$), 129.89 ($2C_{Bz}$), 129.02 (C_{Bz}), 128.89 (C_{Bz}), 128.85 (C_{Bz}), 128.81 ($2C_{Bz}$), 128.67 ($2C_{Bz}$), 128.51 ($2C_{Bz}$), 90.81 (C-1), 81.34 (d, *J* = 176.4 Hz, C-6), 71.92 (d, *J* = 19.3 Hz, C-5), 69.72 (C-3), 69.30 (C-2), 65.78 (d, *J* = 7.0 Hz, C-4), 21.11 ($OCO\underline{C}H_3$). ^{19}F NMR (376 MHz, $CDCl_3$): δ -232.65 (td, *J* = 47.0, 23.3 Hz). HRMS (ESI-TOF) m/z $[M+Na]^+$ calcd for $C_{29}H_{25}FO_9Na$ 559.1380; found 559.1396.

6-Deoxy-6-fluoro-D-mannose (4). Compound 3 (150 mg, 0.28 mmol) was dissolved in dry MeOH (2 mL), then solid sodium methoxide was added until pH=10-11. The reaction was stirred for 3 h, then quenched with Dowex 50WX8 H^+ form, filtered and concentrated. The residue was purified by silica gel flash column chromatography (CH_2Cl_2:MeOH, 98:5, v/v) to give 4 as a white solid (40 mg, 0.22 mmol, 78%, α:β 9:1). R_f= 0.2, CH_2Cl_2/MeOH 9:1; ^{1}H NMR (500 MHz, CD_3OD): δ 5.11 (d, *J* = 1.7 Hz, 1H, H-1α), 4.73 – 4.54 (m, 2H, H-6a, H-6b), 3.91 (dddd, *J* = 26.0, 10.0, 4.7, 2.0 Hz, 1H, H-5), 3.83 (dd, *J* = 3.4, 1.7 Hz, 1H, H-2), 3.80 (dd, *J* = 9.2, 3.4 Hz, 1H, H-3), 3.69 (t, *J* = 9.6 Hz, 1H, H-4). ^{19}F NMR (470 MHz, CD_3OD) δ -233.99 (td, J = 47.9, 23.8 Hz), -234.81 (td, J = 47.8, 25.8 Hz). All analytical data were consistent with literature values [73].

4.2. NMR experiments

All NMR spectra were recorded on a 500 MHz Bruker spectrometer (470.56 MHz for fluorine) equipped with a ^{19}F probe (^{19}F, ^{1}H SEF from Bruker) at 298 K in D_2O unless otherwise is indicated. Standard pulse sequences 1D ^{1}H with and without decoupling ^{19}F and 1D-^{19}F with and without decoupling 1H included in Topspin acquisition software were used. For measuring transversal relaxation times, T_2, the CPMG pulse sequence was used [75,76]. It was as follows: [*D*-90_x-(τ-180_y-τ)$_n$-acquire], with a prescan delay of 4 s and a pre and post 180° pulse echo delay τ of 2ms. The number ***n*** of echo loops varies from 2 to 2000, depending on the experiment. The 90_x and 180_y pulse durations were calculated for each sample. .Total time

used for the relaxation filter corresponds to $\boldsymbol{n}$ times the spin echo pulse applied: $n(2\tau + 180_y)$, (typically between 8ms to 8 s).



^{19}F was set as the observe nucleus, and proton decoupling was carried out during acquisition using WALTZ-16 scheme.

4.3. Transverse relaxation time of F-monosaccharides

To carry out the relaxation filtered experiments, individual stock solutions around 50 mM in deuterated water of each fluorinated monosaccharide (glucose, galactose, mannose and L-fucose) were prepared. These stock solutions were appropriately mixed and diluted to the final concentration used in each experiment. The concentrations were estimated by integrating the corresponding signals in the ^{19}F-spectrum. 2-Fluoroethanol was added to the mixture as internal reference. To prepare the samples of the monosaccharide library in the presence of proteins, 0.2 mL or 0.5 mL (for using 2 mm and 5 mm NMR tubes, respectively) aliquots of the mixture with 0.9 mM of each fluorinated monosaccharide were dried in a speed-vac and the resulting powder reconstituted with the same volume of the corresponding buffer with and without protein. T_2 values were obtained from a series of CPMG experiments recorded with increasing number n (spin echo loops). Experiments with up to 16 different spin echo total relaxation times ranging from 8 to 8000 ms were determined.

4.4. Detection of ligand binding by T_2-filtered experiments

A general protocol was followed using a protein-containing solution with a concentration between 10 μM and 30 μM. The mixtures of monofluorinated monosaccharides were prepapred by mixing aliquots of each monosaccharide from highly concentrated. The final concentration of each monosaccharide in the mixture was around 0.9 mM ([α]+[β]) ranging between 0.6 to 1.2 mM depending their availability. The ligand to protein ratio (L/P) was maintained between 20 to 150 fold excess, optimized in each case to yield comparable T_2 decay responses between the three systems (PSA, GGBP and MGL). CPMG experiments were carried out as previously described, but recording an initial reference experiment with 2 CPMG loops

with τ= 2ms (8 ms total relaxation time) and one to five additional experiments with CPMG filters between 16-400 CPMG echo loops (64 to 1600 ms, respectively; the exact values of the spectra selected are indicated in each experiment).

In order to obtain the difference spectrum, the T_2 filtered spectra were multiplied by a factor f to correct the signal decay in the absence of protein. *f* is defined as the mean value of the ratio of ^{19}F signal intensities after the first (I_1) and the last (I_t) CPMG experiments for all the fluorinated monosaccharides in the mixture: $f = \overline{I_1/I_t}$. The first CPMG spectrum is acquired with $t_1 = 8$ ms, and the last one at $t = 160, 720$ and 400 ms in each case, yielding a factor *f* of 1.1, 1.6 and 1.2 as shown in Figures 3, 4 and 6, respectively.

4.5. Detection of ligand binding by competition experiments

Competition (displacement) experiments were performed by adding an excess of a cognate sugar to the lectin/monofluorinated monosaccharide mixtures. In particular, Me α-*N*-acetylgalactosaminide (12 μM Kd) up to 1 mM for MGL [53], Me α-mannopyranoside (530 μM [55] up to 18 mM for PSA and glucose (0.04 μM) [65] up to 1 mM for GGBP. Equivalent experiments, with the same CPMG relaxation filter parameters to those used for detection of ligand binding, were carried out to observe recovery in signals that had previously diminished as a consequence of binding. Each experiment was repeated upon sequential addition of the competing ligand.

The signal recovery ratio represented in Figure 8 for each fluorinated monosaccharide in the presence of the lectin after relaxation time t_i, $I_{t_i(+C/-C)}$, with (+C) or without (-C) competitor was calculated from the ratio of relative signal decays in presence $(I_{t_i}/I_{t_1})_{+C}$ and in absence $(I_{t_i}/I_{t_1})_{-C}$ of competitor for t_1 = 8 ms and t_i = 720, 720 and 400 ms for MGL, PSA and GGBP, respectively.

5. REFERENCES

1. Ginsburg, V.; Neufeld, E. F. Complex Heterosaccharides of Animals *Ann. Rev. Biochem.* **1969**, *38*, 371.
2. Varki, A. Biological roles of glycans *Glycobiology* **2017**, *27*, 3.
3. Gagneux, P.; Aebi, M.; Varki, A. In *Essentials of Glycobiology. 3rd edition* ; Varki A, C. R., Esko JD, et al. , Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor (NY), 2017, p 253.
4. Kaltner, H.; Abad-Rodriguez, J.; Corfield, A. P.; Kopitz, J.; Gabius, H.-J. The sugar code: letters and vocabulary, writers, editors and readers and biosignificance of functional glycan-lectin pairing *Biochem. J.* **2019**, *476*, 2623.
5. Böhm, M.; Bohne-Lang, A.; Frank, M.; Loss, A.; Rojas-Macias, M. A.; Lütteke, T. Glycosciences.DB: an annotated data collection linking glycomics and proteomics data (2018 update) *Nucleic Acids Res.* **2018**, *47*, D1195.
6. Winterburn, P. J.; Phelps, C. F. The Significance of Glycosylated Proteins *Nature* **1972**, *236*, 147.
7. Solís, D.; Bovin, N. V.; Davis, A. P.; Jiménez-Barbero, J.; Romero, A.; Roy, R.; Smetana, K., Jr.; Gabius, H.-J. A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code *Biochim. Biophys. Acta* **2015**, *1850*, 186.
8. Agre, P.; Bertozzi, C.; Bissell, M.; Campbell, K. P.; Cummings, R. D.; Desai, U. R.; Estes, M.; Flotte, T.; Fogleman, G.; Gage, F.; Ginsburg, D.; Gordon, J. I.; Hart, G.; Hascall, V.; Kiessling, L.; Kornfeld, S.; Lowe, J.; Magnani, J.; Mahal, L. K.; Medzhitov, R.; Roberts, R. J.; Sackstein, R.; Sarkar, R.; Schnaar, R.; Schwartz, N.; Varki, A.; Walt, D.; Weissman, I. Training the next generation of biomedical investigators in glycosciences *J. Clin. Invest.* **2016**, *126*, 405.
9. Lemieux, R. U. Toward Enhanced symbiosis of Chemistry and Biology *FASEB J.* **1994**, *8*, 273.
10. Bessell, E. M.; Foster, A. B.; Westwood, J. H. The use of deoxyfluoro-D-glucopyranoses and related compounds in a study of yeast hexokinase specificity *Biochem. J.* **1972**, *128*, 199.

11. Glaudemans, C. P. J. Mapping of subsites of monoclonal anti-carbohydrate antibodies using deoxy and deoxyfluorosugars *Chem. Rev.* **1991**, *91*, 25.

12. Fernández, P.; Jiménez-Barbero, J.; Martín-Lomas, M. Synthesis of all the possible monomethyl ethers and several deoxyhaloanalogs of methyl β-lactoside as ligands for the *Ricinus communis* lectins *Carbohydr. Res.* **1994**, *254*, 61.

13. Hevey, R. Bioisosteres of Carbohydrate Functional Groups in Glycomimetic Design *Biomimetics* **2019**, *4*, 53.

14. Barnett, J. E. G.; Holman, G. D.; Munday, K. A. Structural Requirements for Binding to Sugar-Transport System of Human Erythrocyte *Biochem. J.* **1973**, *131*, 211.

15. Vanwauwe, J. P.; Loontiens, F. G.; Debruyne, C. K. Carohydrate binding specificity of lectin from pea (pisum-sativum) *Biochim. Biophys. Acta* **1975**, *379*, 456.

16. Bhattacharyya, L.; Brewer, C. F. Lectin-carbohydrate interactions - studies of the nature of hydrogen-bonding between D-galactose and certain D-galactose-specific lectins, and between D-mannose and concanavalin-A *Eur. J. Biochem.* **1988**, *176*, 207.

17. Withers, S. G.; Street, I. P.; Percival, M. D. Fluorinated carbohydrates as probes of enzyme specificity and mechanism *ACS Symp. Ser.* **1988**, *374*, 59.

18. Nycholat, C. M.; Bundle, D. R. Synthesis of monodeoxy and mono-O-methyl congeners of methyl beta-D-mannopyranosyl-(1 -> 2)-beta-D-mannopyranoside for epitope mapping of anti-Candida albicans antibodies *Carbohydr. Res.* **2009**, *344*, 555.

19. McCarter, J. D.; Adam, M. J.; Withers, S. G. Binding-Energy and Catalysis. Fluorinated and Deoxygenated Glycosides as Mechanistics Probes of *Escherichia-coli* (LacZ) Beta-Galactosidase *Biochem. J.* **1992**, *286*, 721.

20. Lemieux, R. U. The origin of the specificity in the recognition of oligosaccharides by proteins. *Chem. Soc. Rev.* **1989**, *18*, 347.

21. Linclau, B.; Ardá, A.; Reichardt, N.-C.; Sollogoub, M.; Unione, L.; Vincent, S. P.; Jiménez-Barbero, J. Fluorinated carbohydrates as chemical probes for molecular recognition studies. Current status and perspectives *Chem. Soc. Rev.* **2020**, *49*, 3863.

22. Bégué, J.-P.; Bonnet-Delpon, D. l. *Bioorganic and medicinal chemistry of fluorine*; John Wiley & Sons: Hoboken, N.J., 2008.

23. Gillis, E. P.; Eastman, K. J.; Hill, M. D.; Donnelly, D. J.; Meanwell, N. A. Applications of Fluorine in Medicinal Chemistry *J. Med. Chem.* **2015**, *58*, 8315.

24. Dalvit, C.; Invernizzi, C.; Vulpetti, A. Fluorine as a hydrogen-bond acceptor: experimental evidence and computational calculations *Chem. Eur. J.* **2014**, *20*, 11058.

25. Dunitz, J. D. Organic fluorine: odd man out *ChemBioChem* **2004**, *5*, 614.

26. Biffinger, J. C.; Kim, H. W.; DiMagno, S. G. The polar hydrophobicity of fluorinated compounds *Chembiochem* **2004**, *5*, 622.

27. Unione, L.; Xu, B.; Díaz, D.; Martín-Santamaría, S.; Poveda, A.; Sardinha, J.; Rauter, A. P.; Bleriot, Y.; Zhang, Y.; Cañada, F. J.; Sollogoub, M.; Jiménez-Barbero, J. Conformational Plasticity in Glycomimetics: Fluorocarbamethyl-L-idopyranosides Mimic the Intrinsic Dynamic Behaviour of Natural Idose Rings *Chem. Eur. J.* **2015**, *21*, 10513.

28. St-Gelais, J.; Côté, É.; Lainé, D.; Johnson, P. A.; Giguère, D. Addressing the Structural Complexity of Fluorinated Glucose Analogues: Insight into Lipophilicities and Solvation Effects *Chem. Eur. J. 2020, doi: 10.1002/chem.202002825*.

29. Linclau, B.; Wang, Z.; Compain, G.; Paumelle, V.; Fontenelle, C. Q.; Wells, N.; Weymouth-Wilson, A. Investigating the Influence of (Deoxy) fluorination on the Lipophilicity of Non-UV-Active Fluorinated Alkanols and Carbohydrates by a New log P Determination Method *Angew. Chem., Int. Ed.* **2016**, *55*, 674.

30. Williams, S. J.; Withers, S. G. Glycosyl fluorides in enzymatic reactions *Carbohydr. Res.* **2000**, *327*, 27.

31. Withers, S. G.; Street, I. P.; Bird, P.; Dolphin, D. H. 2-Deoxy-2-Fluoroglycosides. A Novel Class of Mechanism-Based Glucosidase Inhibitors *J. Am. Chem. Soc.* **1987**, *109*, 7530.

32. Tysoe, C.; Withers, S. G. Fluorinated Mechanism-Based Inhibitors: Common Themes and Recent Developments *Curr. Top. Med. Chem.* **2014**, *14*, 865.

33. Denavit, V.; Laine, D.; Bouzriba, C.; Shanina, E.; Gillon, E.; Fortin, S.; Rademacher, C.; Imberty, A.; Giguere, D. Stereoselective Synthesis of Fluorinated Galactopyranosides as Potential Molecular Probes for Galactophilic Proteins: Assessment of Monofluorogalactoside-LecA Interactions *Chem. Eur. J.* **2019**, *25*, 4478.

34. Diercks, T.; Infantino, A. S.; Unione, L.; Jiménez-Barbero, J.; Oscarson, S.; Gabius, H.-J. Fluorinated Carbohydrates as Lectin Ligands: Synthesis of OH/F-Substituted N-Glycan Core Trimannoside and Epitope Mapping by 2D STD-TOCSYreF NMR spectroscopy *Chem. Eur. J.* **2018**, *24*, 15761.

35. N'Go, I.; Golten, S.; Arda, A.; Cañada, F. J.; Jiménez-Barbero, J.; Linclau, B.; Vincent, S. P. Tetrafluorination of Sugars as Strategy for Enhancing Protein-Carbohydrate Affinity: Application to UDP-Galp Mutase Inhibition *Chem. Eur. J.* **2014**, *20*, 106.

36. Cheng, Y.; Guo, A.-L.; Guo, D.-S. Recent Progress in Synthesis and Applications of Fluorinated Carbohydrates *Curr. Org. Chem.* **2010**, *14*, 977.

37. Uhrig, M. L.; Lantano, B.; Postigo, A. Synthetic strategies for fluorination of carbohydrates *Org. Biomol. Chem.* **2019**, *17*, 5173.

38. Card, P. J. Synthesis of fluorinated carbohydrates *J. Carbohydr. Chem.* **1985**, *4*, 451.

39. Dax, K.; Albert, M.; Ortner, J.; Paul, B. J. Synthesis of deoxyfluoro sugars from carbohydrate precursors *Carbohydr. Res.* **2000**, *327*, 47.

40. Denavit, V.; Laine, D.; St-Gelais, J.; Johnson, P. A.; Giguere, D. A Chiron approach towards the stereoselective synthesis of polyfluorinated carbohydrates *Nat. Commun.* **2018**, *9*, 11.

41. Wang, J.; Sanchez-Rosello, M.; Acena, J. L.; del Pozo, C.; Sorochinsky, A. E.; Fustero, S.; Soloshonok, V. A.; Liu, H. Fluorine in pharmaceutical industry: fluorine-containing drugs introduced to the market in the last decade (2001-2011) *Chem Rev* **2014**, *114*, 2432.

42. Dalvit, C. Ligand- and substrate-based ^{19}F NMR screening: Principles and applications to drug discovery *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243.

43. Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects *J. Med. Chem.* **2019**, *62*, 2218.

44. Diercks, T.; Ribeiro, J. P.; Cañada, F. J.; André, S.; Jiménez-Barbero, J.; Gabius, H.-J. Fluorinated Carbohydrates as Lectin Ligands: Versatile Sensors in ^{19}F-Detected Saturation Transfer Difference NMR Spectroscopy *Chem. Eur. J.* **2009**, *15*, 5666.

45. Matei, E.; Andre, S.; Glinschert, A.; Infantino, A. S.; Oscarson, S.; Gabius, H.-J.; Gronenborn, A. M. Fluorinated Carbohydrates as Lectin Ligands: Dissecting Glycan-Cyanovirin Interactions by Using F-19 NMR Spectroscopy *Chem. Eur. J.* **2013**, *19*, 5364.

46. Shishmarev, D.; Fontenelle, C. Q.; Kuprov, I.; Linclau, B.; Kuchel, P. W. Transmembrane Exchange of Fluorosugars: Characterization of Red Cell GLUT1 Kinetics Using(19)F NMR *Biophys. J.* **2018**, *115*, 1906.

47. Unione, L.; Alcala, M.; Echeverría, B.; Serna, S.; Arda, A.; Franconetti, A.; Cañada, F. J.; Diercks, T.; Reichardt, N.; Jiménez-Barbero, J. Fluoroacetamide Moieties as NMR Spectroscopy Probes for the Molecular Recognition of GlcNAc-Containing Sugars: Modulation of the CH-p Stacking Interactions by Different Fluorination Patterns *Chem. Eur. J.* **2017**, *23*, 3957.

48. Katiyar, S.; Singh, A.; Surolia, A. The interaction of N-trifluoroacetylgalactosamine and its derivatives with winged bean (*Psophocarpus tetragonolobus*) basic agglutinin reveals differential mechanism of their recognition: a fluorine-19 nuclear magnetic resonance study *Glycoconjugate J.* **2014**, *31*, 537.

49. Krishna Sastry, M. V.; Swamy, M. J.; Surolia, A. Analysis of dynamics and mechanism of ligand binding to Artocarpus integrifolia agglutinin. A ^{13}C and ^{19}F NMR study *J. Biol. Chem.* **1988**, *263*, 14826.

50. Martínez, J. D.; Valverde, P.; Delgado, S.; Romano, C.; Linclau, B.; Reichardt, N. C.; Oscarson, S.; Arda, A.; Jiménez-Barbero, J.; Cañada, F. J. Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates *Molecules* **2019**, *24*, 2337.

51. Wamhoff, E. C.; Hanske, J.; Schnirch, L.; Aretz, J.; Grube, M.; Silva, D. V.; Rademacher, C. F-19 NMR-Guided Design of Glycomimetic Langerin Ligands *ACS Chem. Biol.* **2016**, *11*, 2407.

52. van Vliet, S. J.; Saeland, E.; van Kooyk, Y. Sweet preferences of MGL: carbohydrate specificity and function *Trends in immunology* **2008**, *29*, 83.

53. Marcelo, F.; García-Martín, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; Andre, S.; Cabrita, E. J.; Gabius, H.-J.; Nishimura, S.-I.; Jiménez-Barbero, J.; Cañada, F. J. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin *Chem. Eur. J.* **2014**, *20*, 16147.

54. Pirro, M.; Rombouts, Y.; Stella, A.; Neyrolles, O.; Burlet-Schiltz, O.; van Vliet, S. J.; de Ru, A. H.; Mohammed, Y.; Wuhrer, M.; van Veelen, P. A.; Hensbergen, P. J. Characterization of Macrophage Galactose-type Lectin (MGL) ligands in colorectal cancer cell lines *Biochim. Biophys. Acta* **2020**, *1864*.

55. Schwarz, F. P.; Puri, K. D.; Bhat, R. G.; Surolia, A. Thermodynamics of monosaccharide binding to concanavalin A, pea (*Pisum-sativum*) lectin, and lentil (*Lens-culinaris*) lectin *J. Biol. Chem.* **1993**, *268*, 7668.

56. Ortega, G.; Castaño, D.; Diercks, T.; Millet, O. Carbohydrate Affinity for the Glucose-Galactose Binding Protein Is Regulated by Allosteric Domain Motions *J. Am. Chem. Soc.* **2012**, *134*, 19869.

57. Borrok, M. J.; Kiessling, L. L.; Forest, K. T. Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures *Protein Sci.* **2007**, *16*, 1032.

58. Aqvist, J.; Mowbray, S. L. Sugar Recognition by a Glucose/Galactose Receptor. Evaluation of Binding Energetics from Molecular Dynamics Simulations *J. Biol. Chem.* **1995**, *270*, 9978.

59. Jégouzo, S. A. F.; Quintero-Martínez, A.; Ouyang, X.; dos Santos, Á.; Taylor, M. E.; Drickamer, K. Organization of the extracellular portion of the macrophage galactose receptor: a trimeric cluster of simple binding sites for N-acetylgalactosamine *Glycobiology* **2013**, *23*, 853.

60. Kerins, L.; Byrne, S.; Gabba, A.; Murphy, P. V. Anomer Preferences for Glucuronic and Galacturonic Acid and Derivatives and Influence of Electron-Withdrawing Substituents *J. Org. Chem.* **2018**, *83*, 7714.

61. Hajduk, P. J.; Olejniczak, E. T.; Fesik, S. W. One-dimensional relaxation- and diffusion-edited NMR methods for screening compounds that bind to macromolecules *J. Am. Chem. Soc.* **1997**, *119*, 12257.

62. Dalvit, C.; Piotto, M. F-19 NMR transverse and longitudinal relaxation filter experiments for screening: a theoretical and experimental analysis *Magn. Reson. Chem.* **2017**, *55*, 106.

63. Srinivas, V. R.; Reddy, G. B.; Ahmad, N.; Swaminathan, C. P.; Mitra, N.; Surolia, A. Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization *Biochim. Biophys. Acta* **2001**, *1527*, 102.

64. Schwarz, F. P.; Misquith, S.; Surolia, A. Effect of substituent on the thermodynamics of D-glucopyranoside binding to concanavalin A, pea (*Pisum sativum*) lectin and lentil (*Lens culinaris*) lectin *Biochem. J.* **1996**, *316 (Pt 1)*, 123.

65. Miller, D. M.; Olson, J. S.; Quiocho, F. A. The Mechanism of Sugar binding to the Periplasmic Receptor for Galactose Chemotaxis and Transport in *Escherichia coli J. Biol. Chem.* **1980**, *255*, 2465.

66. Vyas, M. N.; Vyas, N. K.; Quiocho, F. A. Crystallographic analysis of the epimeric and anomeric specificity of the periplasmic transport/chemosensory protein receptor for D-glucose and D-galactose *Biochemistry* **1994**, *33*, 4762.

67. Dalvit, C.; Parent, A.; Vallée, F.; Mathieu, M.; Rak, A. Fast NMR Methods for Measuring in the Direct and/or Competition Mode the Dissociation Constants of Chemical Fragments Interacting with a Receptor *ChemMedChem* **2019**, *14*, 1115.

68. O'Sullivan, J. A.; Chang, A. T.; Youngblood, B. A.; Bochner, B. S. Eosinophil and mast cell Siglecs: From biology to drug target *J. Leukocyte Biol.* **2020**, *108*, 73.

69. Duan, S.; Paulson, J. C. . Siglecs as Immune Cell Checkpoints in Disease *Ann. Rev. Immunol.,* **2020**, 38, 365.

70. Garcia Caballero, G.; Kaltner, H.; Kutzner, T. J.; Ludwig, A.-K.; Manning, J. C.; Schmidt, S.; Sinowatz, F.; Gabius, H.-J. How galectins have become multifunctional proteins *Histol.Histopathol.* **2020**, *35*, 509.

71. Ludwig, A.-K.; Michalak, M.; Xiao, Q.; Gilles, U.; Medrano, F. J.; Ma, H.; FitzGerald, F. G.; Hasley, W. D.; Melendez-Davila, A.; Liu, M.; Rahimi, K.; Kostina, N. Y.; Rodriguez-Emmenegger, C.; Moeller, M.; Lindner, I.; Kaltner, H.; Cudic, M.; Reusch, D.; Kopitz, J.; Romero, A.; Oscarson, S.; Klein, M. L.; Gabius, H.-J.; Percec, V. Design-functionality relationships for adhesion/growth-regulatory galectins *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 2837.

72. Kaltner, H.; Manning, J. C.; Caballero, G. G.; Di Salvo, C.; Gabba, A.; Romero-Hernandez, L. L.; Knospe, C.; Wu, D.; Daly, H. C.; O'Shea, D. F.; Gabius, H.-J.; Murphy, P. V. Revealing biomedically relevant cell and lectin type-dependent structure-activity profiles for glycoclusters by using tissue sections as an assay platform *RSC Advances* **2018**, *8*, 28716.

73. Khedri, Z.; Muthana, M. M.; Li, Y.; Muthana, S. M.; Yu, H.; Cao, H.; Chen, X. Probe sialidase substrate specificity using chemoenzymatically synthesized sialosides containing C9-modified sialic acid *Chem. Commun.* **2012**, *48*, 3357.

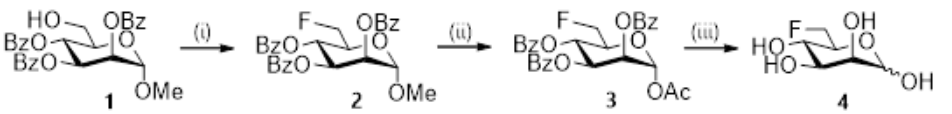
74. Esmurziev, A. M.; Simic, N.; Hoff, B. H.; Sundby, E. Synthesis and Structure Elucidation of Benzoylated Deoxyfluoropyranosides *J. Carbohydr. Chem.* **2010**, *29*, 348.

75. Carr, H. Y.; Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments *Physical Rev.* **1954**, *94*, 630.

76. Meiboom, S.; Gill, D. Modified Spin-Echo Method for Measuring Nuclear Relaxation Times *Rev. Sci. Instrum.* **1958**, *29*, 688

6. SUPPORTING INFORMATION

Synthesis of 6-deoxy-6-fluoro-D-mannose (4)



Scheme 1: Synthetic route to compound **4**. (i) DAST, CH_2Cl_2, -78 °C→RT, 74%; (ii) H_2SO_4, Ac_2O/AcOH, 0 °C, 80%; (iii) MeONa, MeOH, 78%, α:β, 9:1.

NMR Spectra, ^{1}H, ^{13}C and ^{19}F, of compounds 2, 3 and 4

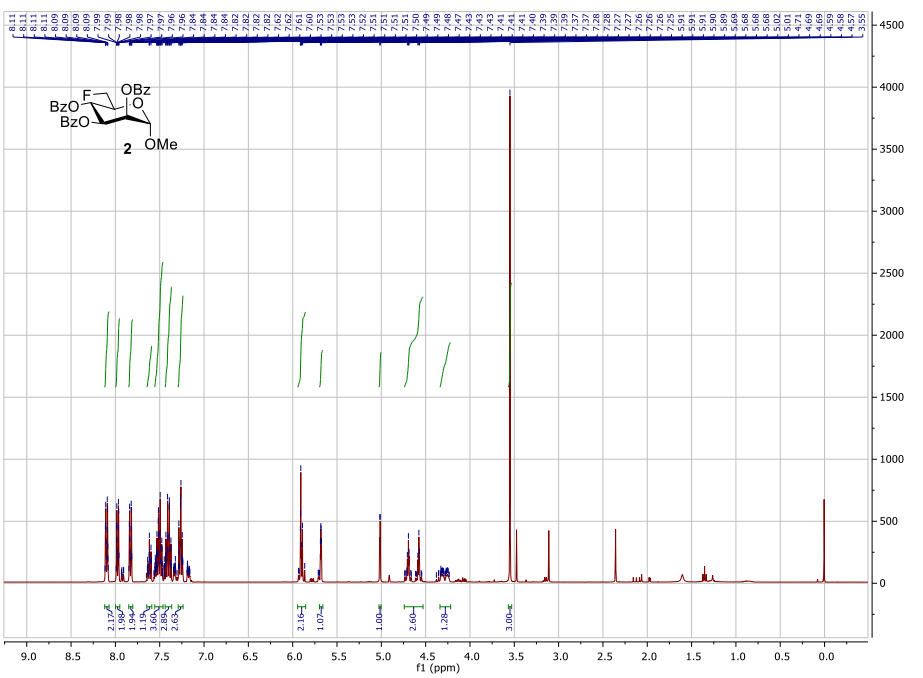


Figure S1. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of methyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (2)

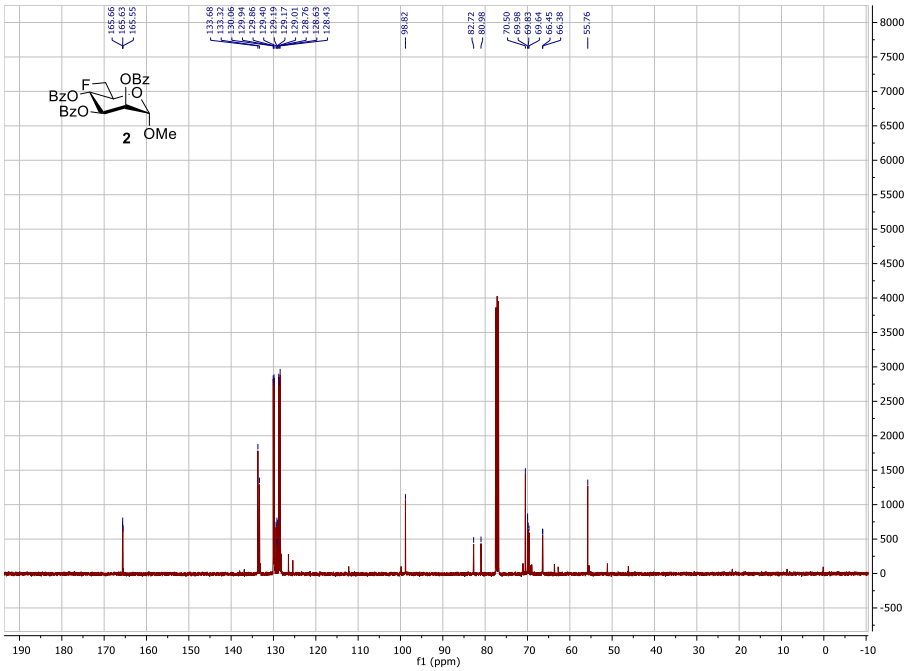


Figure S2. ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of methyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (2)

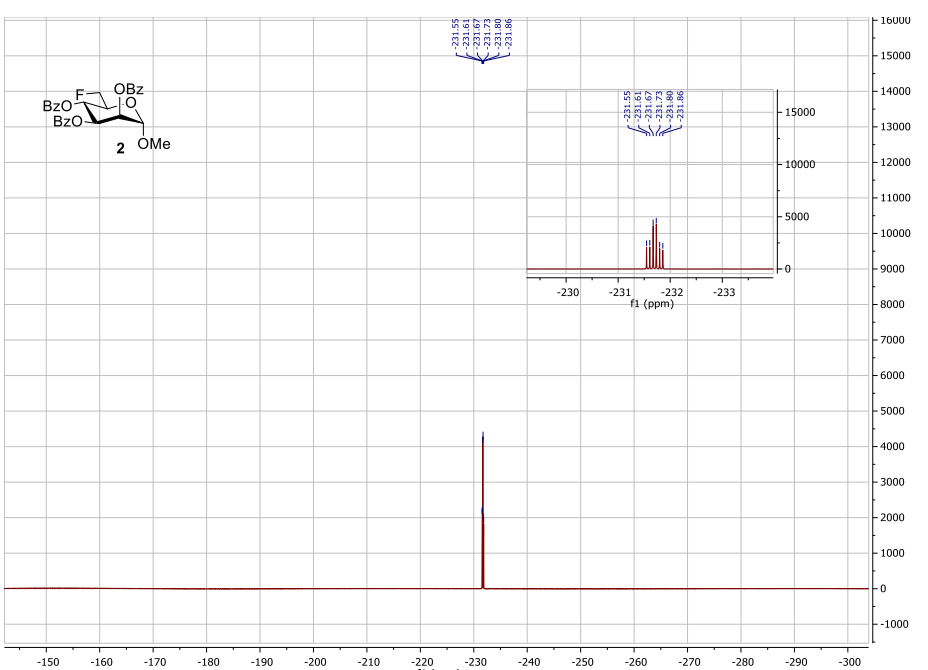


Figure S3. ^{19}F NMR (376 MHz, $CDCl_3$) of methyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (2)

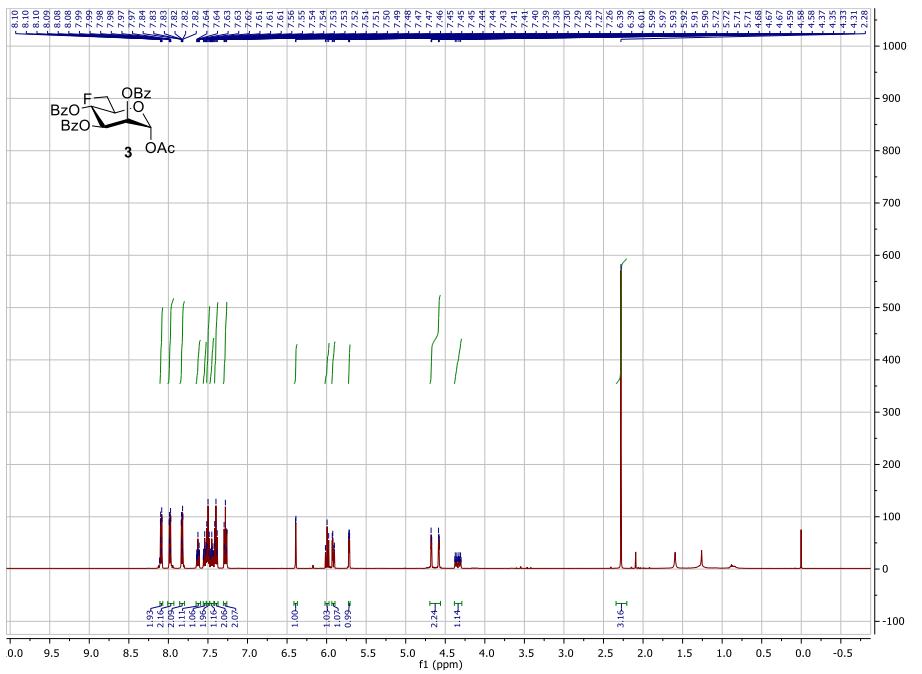


Figure S4. ^{1}H NMR (400 MHz, $CDCl_3$) spectrum of acetyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (3)

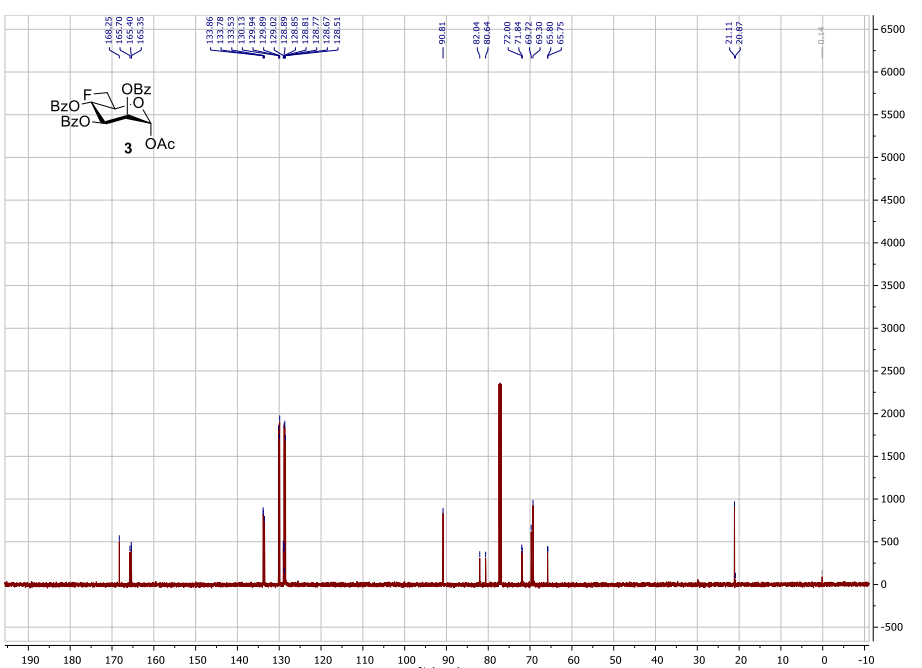


Figure S5. ^{13}C NMR (126 MHz, $CDCl_3$) spectrum of acetyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (3)

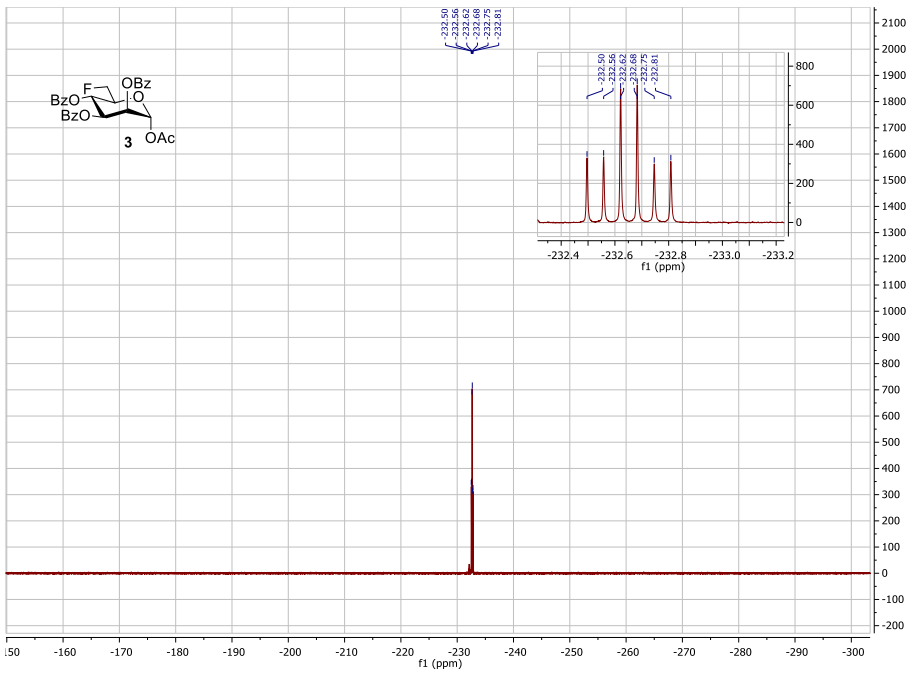


Figure S6. ^{19}F NMR (376 MHz, $CDCl_3$) spectrum of acetyl 2,3,4-tri-O-benzoyl-6-deoxy-6-fluoro-α-D-mannopyranoside (3)

Figure S7. ^{1}H NMR (400 MHz, CD_3OD) spectrum of 6-Deoxy-6-fluoro-D-mannose (4)

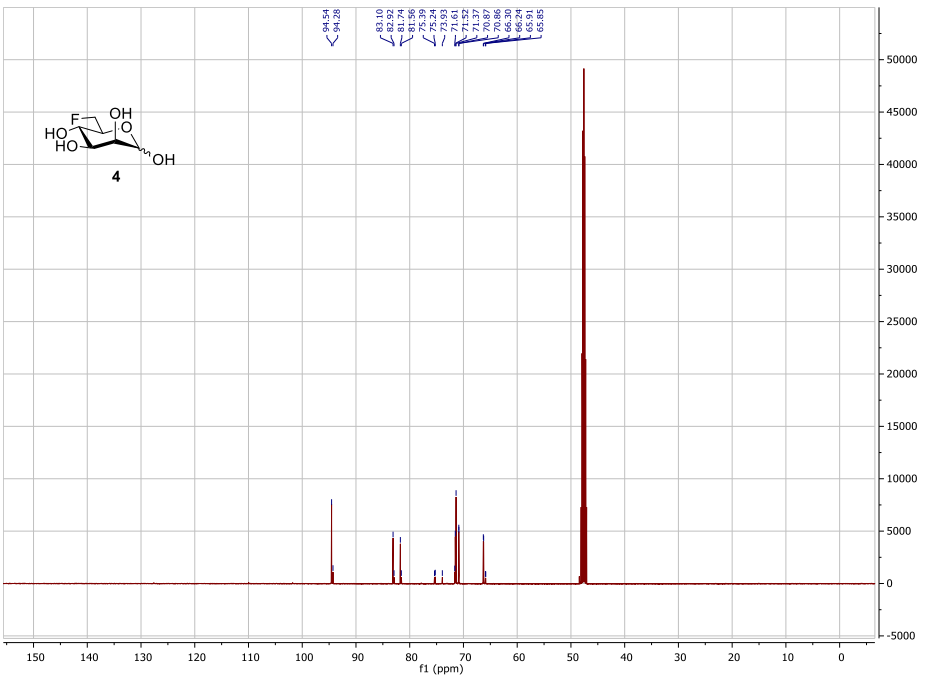


Figure S8. ^{13}C NMR (126 MHz, CD_3OD) spectrum of 6-Deoxy-6-fluoro-D-mannose (4)

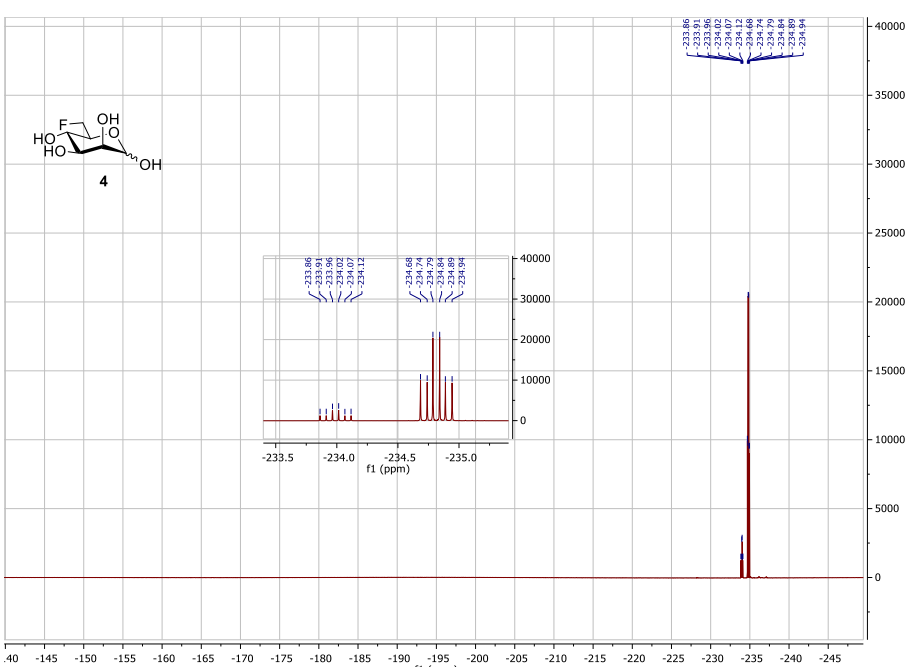


Figure S9. ^{19}F NMR (376 MHz, CD_3OD) spectrum of 6-Deoxy-6-fluoro-D-mannose (4)

^{19}F-NMR spectra of F-monosaccharides (470.55 MHz, D_2O)

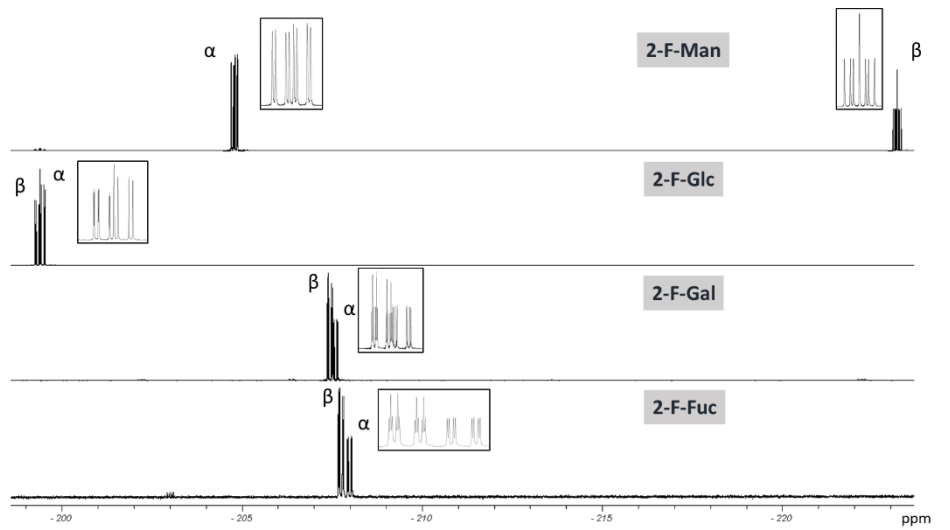


Figure S10. ^{19}F NMR spectrum of 2-F-monosaccharides.

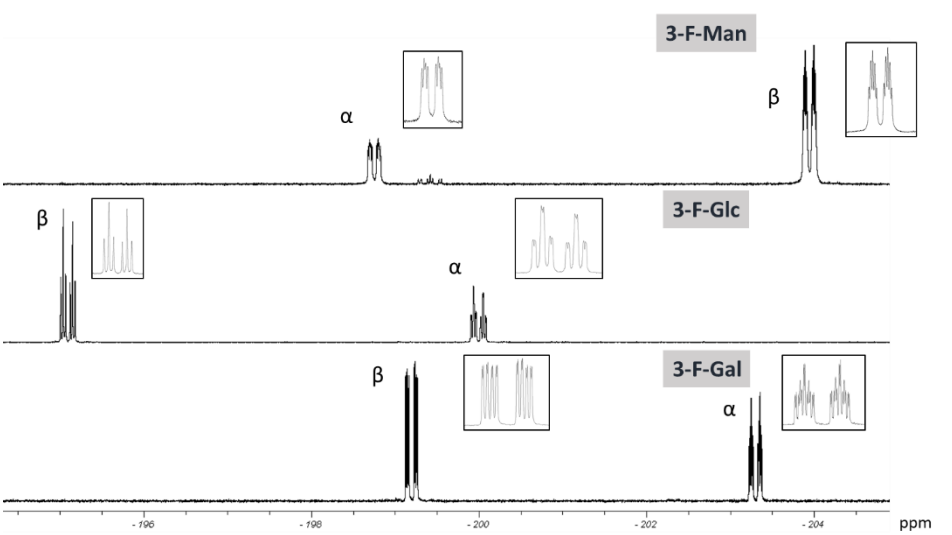


Figure S11. ^{19}F NMR spectrum of 3-F-monosaccharides.

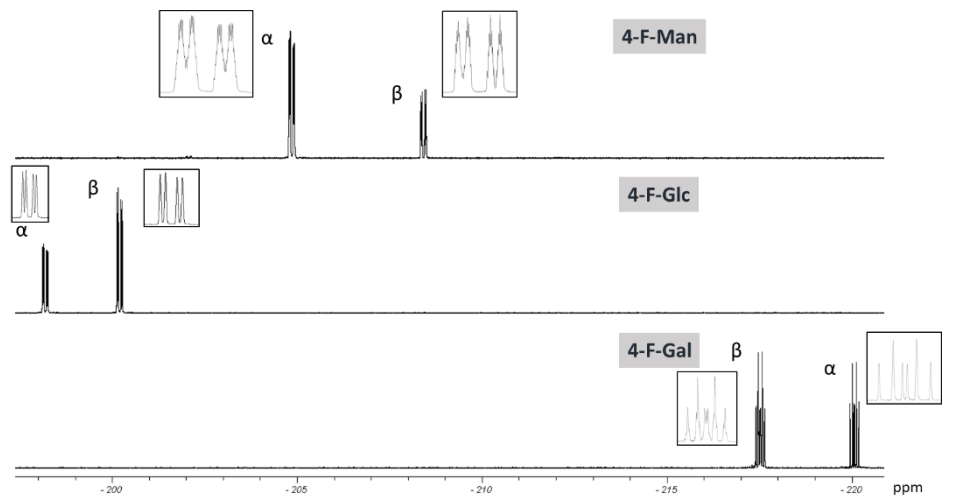


Figure S12. ^{19}F NMR spectrum of 4-F-monosaccharides.

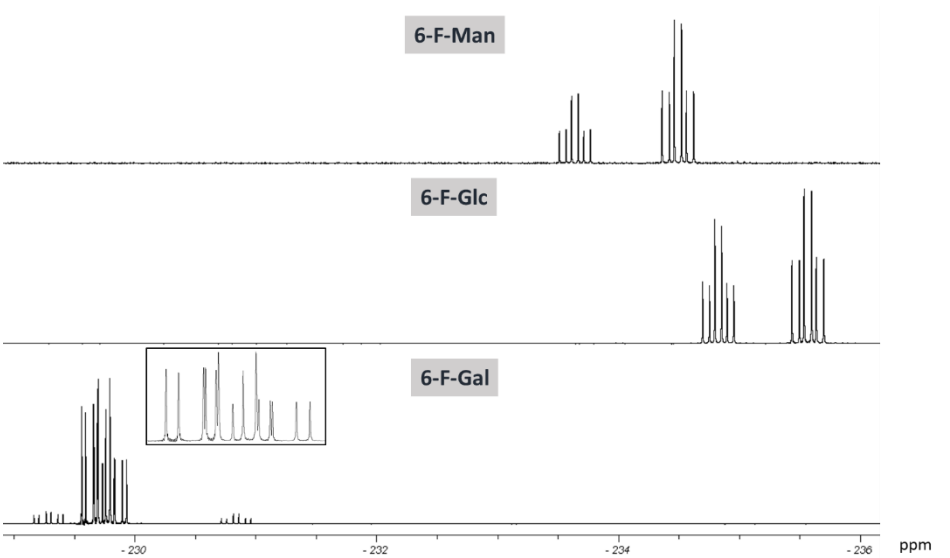


Figure S13. ^{19}F NMR spectrum of 6-F-monosaccharides.

CHAPTER 4

Unraveling sugar binding modes to DC-SIGN by employing fluorinated carbohydrates

UNRAVELING SUGAR BINDING MODES TO DC-SIGN BY EMPLOYING FLUORINATED CARBOHYDRATES

Reproduced with permission from the authors: Pablo Valverde, Sandra Delgado, Cecilia Romanò, Bruno Linclau, Niels C. Reichardt, Stefan Oscarson, Ana Ardá, Jesús Jiménez-Barbero, F. Javier Cañada. 'Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates', *Molecules* **2019**, *24*(12), 2337

Abstract: A fluorine nuclear magnetic resonance (^{19}F-NMR)-based method is employed to assess the binding preferences and interaction details of a library of synthetic fluorinated monosaccharides towards dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN), a lectin of biomedical interest, which is involved in different viral infections, including HIV and Ebola, and is able to recognize a variety of self- and non-self-glycans. The strategy employed allows not only screening of a mixture of compounds, but also obtaining valuable information on the specific sugar–protein interactions. The analysis of the data demonstrates that monosaccharides Fuc, Man, Glc, and Gal are able to bind DC-SIGN, although with decreasing affinity. Moreover, a new binding mode between Man moieties and DC-SIGN, which might have biological implications, is also detected for the first time. The combination of the ^{19}F with standard proton saturation transfer difference (^{1}H-STD-NMR) data, assisted by molecular dynamics (MD) simulations, permits us to successfully define this new binding epitope, where Man coordinates a Ca^{2+} ion of the lectin carbohydrate recognition domain (CRD) through the axial OH-2 and equatorial OH-3 groups, thus mimicking the Fuc/DC-SIGN binding architecture.

1. INTRODUCTION

Carbohydrates are ubiquitous in nature in different combinations, from single monosaccharides to extremely complex glycoconjugates. The possibility of finding different molecular structures, by variation of the stereochemistry of hydroxyl groups and the regiochemistry of the glycosidic linkages, grows geometrically as compared with other linear biopolymers. Ligand-based NMR experiments are widely used to study the interactions between carbohydrates and receptors in solution [1–4]. Among these methods that are especially powerful for screening purposes, ^{19}F-based approaches present a number of important advantages [5–8]. In this study, we applied a strategy that allows the screening of a library of synthetic fluorinated monosaccharides to study the interactions with the biologically relevant lectin dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN) [9] and, at the same time, obtain ligand-specific chemical mapping information.

DC-SIGN is a C-type lectin expressed by dendritic cells [10]. It acts as an adhesion receptor in cell–cell interactions, and plays crucial roles in the DC migration and adaptive immune response initiation [10]. It is extensively reported that DC-SIGN is able to bind the HIV-1 envelope glycoprotein, gp 120, which is exploited by the virus to enhance its infectivity of T cells [11,12]. The importance of DC-SIGN is, therefore, evident and a full understanding of the recognized mechanism and its different ligands remains a matter of interest. Glycan binding to DC-SIGN occurs through direct coordination with one of the structural Ca^{2+} ions of the lectin carbohydrate recognition domain (CRD) and the additional contacts with the surrounding amino acid residues define the sugar specificity. An evaluation of the monosaccharide binding affinity revealed Fuc as the preferred ligand, followed by Man (two-fold weaker) [13]. However, the binding specificity of DC-SIGN is remarkably broad, and its glycan binding promiscuity has been related to its different biological roles [14].

In this study, the sugar recognition profile of DC-SIGN is studied in detail. Employing the screening and ligand chemical mapping protocol with a rationally designed mixture of 26 monofluorinated monosaccharides, we confirm the previously reported preference for Fuc and Man moieties exhibited by DC-SIGN. Interestingly, evidence of the interaction between Gal moieties and the lectin is also found, as well as the existence of a new binding epitope for Man residues that involves direct contacts of Man O2/O3 with the Ca^{2+} of DC-SIGN CRD, which was not previously detected.

2. RESULTS AND DISCUSSION

2.1. ^{19}F-NMR-Based Chemical Mapping

The wide range of chemical shifts displayed by the ^{19}F nucleus permits the identification of every fluorine-containing monosaccharide within the complete mixture in a straightforward manner, using a standard proton decoupled ^{19}F spectra. Both α- and β-anomers are perfectly distinguished for every sugar. Therefore, the spectrum of the mixture contains 26 different ^{19}F-NMR signals (13 distinct fluorine-containing monosaccharides and two anomers of each, Figure 1).

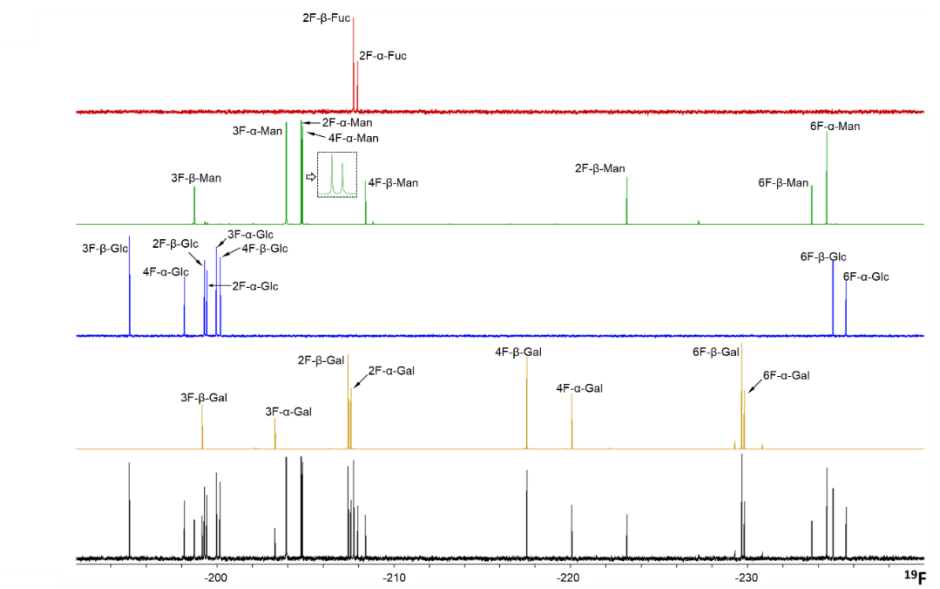


Figure 1. ^{19}F-NMR proton decoupled spectra of the monofluorinated monosaccharides grouped by sugar type: F-Gal (yellow), F-Glc (blue), F-Man (green), and F-Fuc (red). The mixture of all fluorine-containing monosaccharides employed herein is shown in black at the bottom. A close-up of the 2-F-α-Man and 4-F-α-Man signals is indicated for clarity. The concentration of the molecules was 0.8 mM ([α]+[β]). A total of 16 scans were acquired, with a repetition time of 3 s.

First, the transverse relaxation times (T_2) of the corresponding ^{19}F nuclei were quantitatively estimated using the conventional Carr-Purcell-Meiboom-Gill (CPMG) spin echo pulse sequence [15] as a reference for the following analysis (Table S1 in supporting information). The screening is based on monitoring changes in the transversal relaxation times of the fluoro-monosaccharides, before ($T_{2,free}$) and after ($T_{2,obs}$) addition of the lectin, which are indicative of binding. Although visual comparison of the signals in the absence and presence of the lectin for a specific relaxation filter time, or the

observation of the time evolution of the signals (Figure 2), allows a rapid hit identification, it is considerably more reliable to calculate the complete relaxation curves to estimate $T_{2,obs}$ of the ^{19}F nuclei in the mixture.

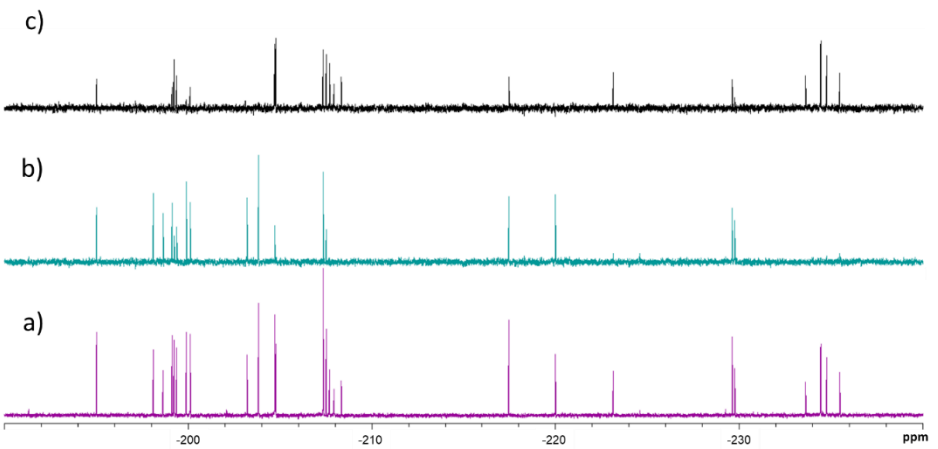


Figure 2. ^{19}F-NMR relaxation filter experiments performed for the fluorine-containing monosaccharide library in the presence of tetrameric dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin (DC-SIGN). The F-sugar/lectin ratio was 47:1. (**a**) Spectrum acquired without T_2 relaxation filter, (**b**) spectrum acquired after 2.4 s of T_2 relaxation filter, and (**c**) difference spectrum. Only the peaks of those molecules that display a significant decrease in their T_2 are present. These peaks correspond to DC-SIGN binders.

Thus, the interaction of tetrameric full-length DC-SIGN with the monosaccharide library was evaluated by computing the decrease of the T_2 values of the fluorinated molecules (Figure 3) in the presence of a given amount of the lectin, normalized for each peak with respect to its $T_{2,free}$ [8]:

$$\% T_{2,decrease} = \frac{T_{2,free} - T_{2,obs}}{T_{2,free}} \times 100 \quad (1)$$

In particular, controlled amounts of increasing DC-SIGN concentrations were added to the fluorinated sugar library and the observed T_2 values were calculated. The analysis of the data showed a systematic reduction of the signal intensities in specific cases (Figure 3), which were correlated with the presence of binding. Interestingly, even rather weak interactions of the molecules with the lectin were detected. Very few signals, corresponding to 3-, 4-, and 6-F-Gal, 3-F-Man, and 3- and 4-F-Glc, were very poorly affected by the presence of DC-SIGN (T_2 decay below 40%). The most perturbed signals corresponded to both anomers of 2-F-Fuc (about 90%), followed by Man moieties with fluorine substitution at positions 2, 4, and 6 (60–80%). As well, 2-, 6-F-Glc, and 2-F-Gal underwent a measurable reduction in their

T_2 (40% to 60% of decay). Fittingly, these observations are in good agreement with previous reports on the natural nonfluorinated sugars [13]. The obtained data also indicate that the anomeric OH does not seem to be involved in the sugar recognition by the lectin. It is evident that the sensitivity of this NMR protocol is fairly high, and that even very low mM affinities binders (such as Gal (K_D 72 mM), a five-fold weaker binder than Man) is also detected [13]. Indeed, the architecture of the CRD of DC-SIGN allows a rather shallow and solvent exposed sugar-binding site, which is fairly accessible to poor affinity ligands.

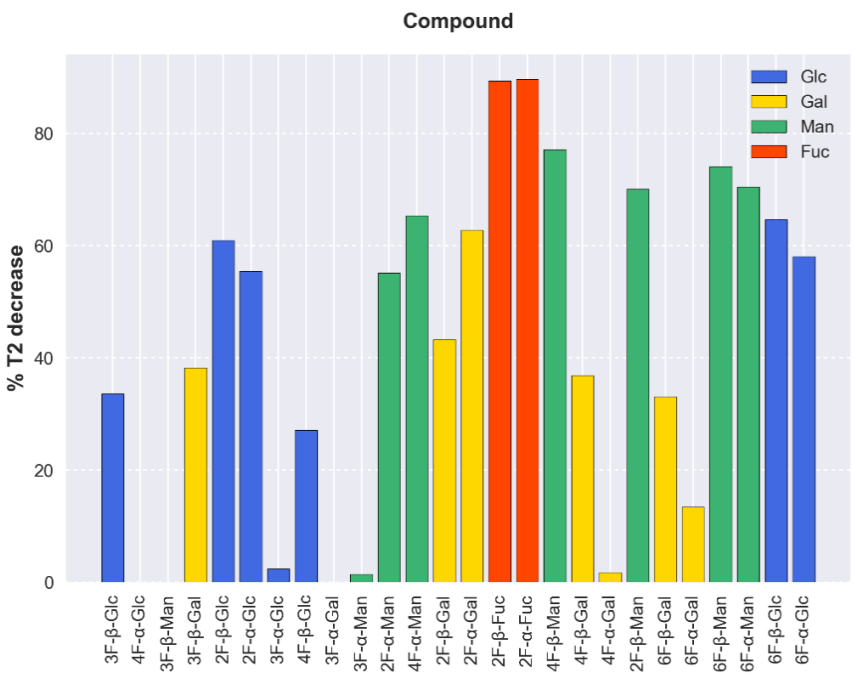


Figure 3. Percentage of decrease of the T_2 values between the free and bound states of the fluorine-containing monosaccharide library obtained from the ^{19}F CPMG experiments. In this particular case, the ligand/DC-SIGN ratio was 47:1. Binders of the lectin display a remarkable decrease in the normalized T_2 values, above 40%. Qualitatively, very weak or non-binders show decreases below 40%. The order in the *x*-axis corresponds to the ^{19}F chemical shift of each molecule, from the lowest to highest shifts.

From the structural perspective, it was shown that Fuc, Man, and even GlcNAc in different oligosaccharides are able to coordinate the calcium ion of DC-SIGN CRD through their OH-3 and OH-4 diols, which also establishes key hydrogen bonds with the neighboring amino acid residues [16,17]. Other C-type lectins also interact with sugars through these hydroxyl groups (i.e., LSECtin and Langerin with Fuc, Man, and GlcNAc) [18,19]. In fact, it has been described that the Man versus Gal specificity in C-lectins is dictated by a short amino acid sequence at the lectin binding site [20].

Therefore, given the subtle interaction network of the coordinating OH, it was expected that replacement of any of these OH-3 and OH-4 hydroxyl groups by a fluorine atom would preclude the sugar-lectin recognition.

As anticipated, the chemical mapping analysis reveals the involvement of Man OH-3, Glc OH-3, and Glc OH-4 as crucial elements in the binding event, but surprisingly not Man OH-4. It also highlights the absence of strong interactions between most Gal moieties and DC-SIGN, with the exception of 2-F-Gal, which could coordinate the calcium ion of the lectin through OH-3 and OH-4. A similar binding mode has been described for Langerin, another C-Type lectin closely related to DC-SIGN [21]. The fact that 6-F-Gal barely interacts with the receptor, despite having the OH-3/OH-4 groups available, suggests some participation of the OH-6 in the very weak binding of the Gal residues [13]. To our knowledge, this is the first evidence of the interaction between Gal and DC-SIGN through OH-3/OH-4-Ca^{2+} contact, disregarding the reported nonphysiological binding detected for Gal with C-type mannose-binding proteins via coordination of Ca^{2+} with OH-1 and OH-2 [13,22]. It is worth mentioning that the residual binding observed for the molecules exhibiting a T_2 decrease below 40% (i.e., for those species considered here as very weak or non-binders), might be related with such a nonphysiological interaction. Interestingly, an anomeric preference for the beta configuration is deduced for those fluorinated-Gal and Glc entities that are scarcely affected by the presence of the lectin, namely 3-, 4-, and 6-F-Gal, and 3- and 4-F-Glc. Since this differentiation does not occur for 3-F-Man, whose hydroxyl at position two is in an axial arrangement, it is possible that the OH-1/OH-2 Ca^{2+} interaction requires a fixed equatorial-equatorial disposition of the contiguous hydroxyl groups in order to take place.

Although the relative decrease in T_2 of the compounds in the mixture by the presence of the lectin depends on several factors, it is related, at least qualitatively, with ligand-receptor affinity as certain conditions are favorable in this particular fluorinated library. Given the small size of the ligands and their similar shape and chemical nature, it is assumed that their T_2 values in the bound state and association rate constants, k_{on} (probably under diffusion control), will be similar. The T_2 values in the free state (1.43 ± 0.73 s), as well as the employed ligand concentrations (0.50 ± 0.15 mM), are also comparable. In addition, it is expected for these molecules sharing the same binding spot on the lectin that the chemical shift differences of the ^{19}F nucleus between the free and bound states, which dictates the exchange contribution to the observed T_2 [5], will also be alike (especially for pairs of anomers when the binding does not involve the hydroxyl at C-1 position). Indeed, the obtained results suggest a general affinity tendency that is in agreement with the current knowledge of DC-SIGN sugar preferences.

Table 1 shows the average decrease in T_2 values by monosaccharide type, discarding those that show a T_2 decrease lower than 40%, herein considered non-binders. Even though only semiquantitative, the fluorine-containing Fuc and Man species experience a more pronounced decrease in their observed T_2 values, followed by Glc and Gal residues.

Table 1. Fluorinated monosaccharides that display effective binding to DC-SIGN and observed averaged decrease in T_2. These values are qualitatively related with the relative binding affinity: Fuc moieties interact better, followed by Man [13]. The Gal moieties displayed a small change in the observed averaged decrease in $T_{2,free}$, highlighting that they are rather poor binders of DC-SIGN [13]. The average $T_{2,free}$ was calculated for all the species depending on the particular sugar. The average T_2 decay % was computed for those species that exhibited a T_2 decay over 40%.

Sugar	$T_{2,free}$ Average	% $T_{2,decrease}$ Average	Binding Molecules
Fuc	2,0	90	(α,β)-2-F-Fuc
Man	1,3	70	(α,β)-2-,4-,6-F-Man
Glc	1,3	60	(α,β)-2-,6-F-Glc
Gal	1,8	51	(α,β)-2-F-Gal

In addition to the identification of 2-F-Gal as a DC-SIGN ligand, these NMR experiments also revealed unexpected findings regarding the binding with Man. As mentioned before, the available X-ray structures show that the interaction of oligosaccharides through Man moieties exclusively takes place by the simultaneous contact of the Ca^{2+} at the DC-SIGN binding site with hydroxyls OH-3/OH-4. The results presented above, however, suggest that the OH-4/Ca^{2+} interaction is actually not essential for binding since 4-F-Man moieties are recognized by the lectin. Indeed, all fluorinated Man molecules displayed a marked decrease in their observed T_2 values upon increasing amounts of tetrameric DC-SIGN in solution, except both 3-F-Man anomers (Figure 4). This fact provides further evidence that the presence of fluorine at position three prevents binding, while the recognition process still takes place when the F substitution occurs at C4. Thus, in the absence of a donor hydroxyl group at position four, other alternative binding modes take place. Based on the sugar recognition requirements observed for DC-SIGN and other C-Type lectins, it can be hypothesized that the alternative binding modes would likely involve the simultaneous contact of hydroxyls OH-2/OH-3 with the Ca^{2+} ion.

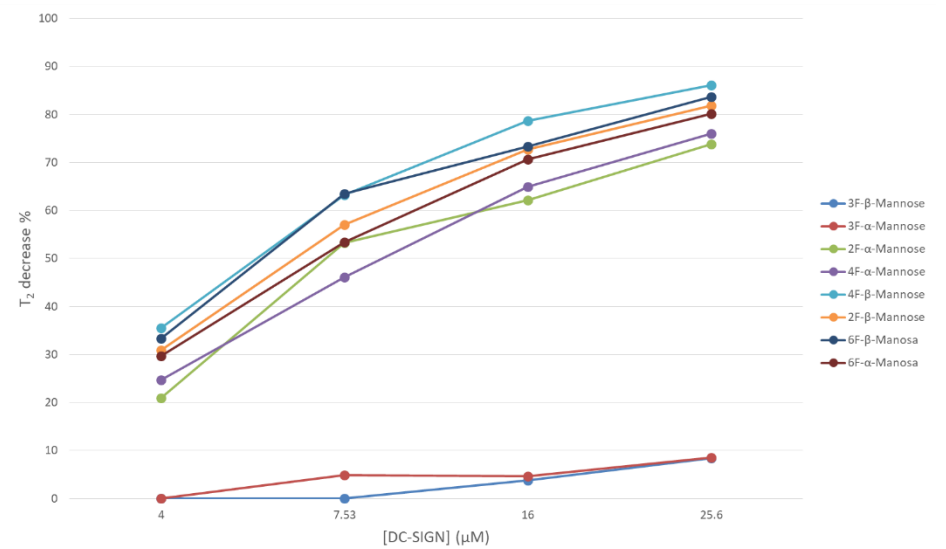
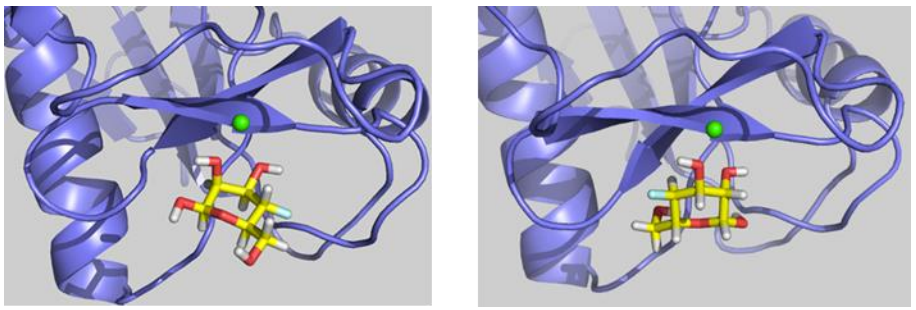


Figure 4. Percentage of decrease of the T_2 values for different ligand/lectin molar ratios deduced for the fluorine-containing Man moieties, as obtained from the ^{19}F-NMR relaxation filter experiments.

The feasibility of these alternative binding modes was explored using molecular modeling procedures. In particular, two putative binding poses were initially built. The first one, binding pose A, was generated using the geometry with PDB code 1SL5 (DC-SIGN complexed to lacto-*N*-fucopentaose), placing the 4-F-Man moiety at the Fuc site by superimposing the OH-2/OH-3 groups of 4-F-Man with OH-4/OH-3 of Fuc (see Figure 5a and Figure 6a, with an equivalent Man molecule in this binding pose). The second one, binding pose B, was generated using the geometry with PDB code 2IT5 (DC-SIGN bound to Man6, a high-mannose *N*-glycan, with the sugar arranged in the major observed orientation), placing the 4-F-Man moiety at the calcium coordinating Man site by superimposing the OH-2/OH-3 groups of 4-F-Man with OH-4/OH-3 of Man (see Figure 5b and also Figure 6b,with an equivalent Man molecule in this binding pose). Both proposed poses are related through a 180° rotation around the line that bisects the pyranose ring across the C3–C2 bond and display two contiguous hydroxyl groups attached to the calcium ion.



(a) **(b)**

Figure 5. Representative structures of β-4-F-Man binding pose A, mimicking Fuc (**a**) and binding pose B, mimicking Man (**b**).

2.2. Molecular Dynamics Simulations

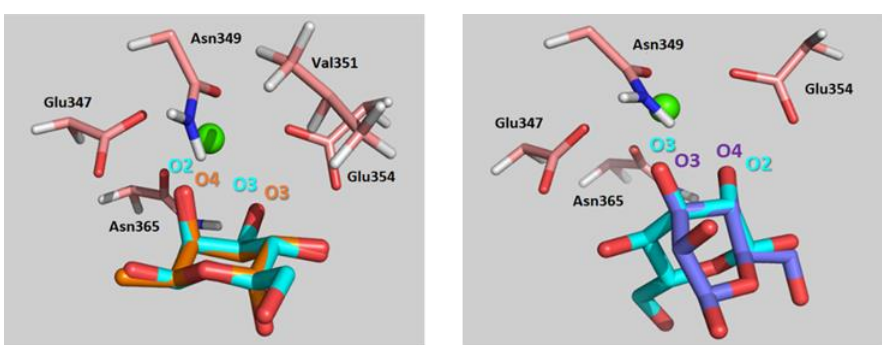
The conformational stability of the complexes was studied by molecular dynamics (MD) simulations in explicit water. To begin with, MD simulations were run for the complexes of DC-SIGN CRD with the natural sugars α-Fuc and α-Man coordinating the principal Ca^{2+}, as found in the 1SL5 and 2IT5 X-ray structures (the other residues of the original oligosaccharides were removed).

The MD simulation showed a stable lectin-Fuc complex throughout the complete trajectory. Apart from the typical coordination of the Ca^{2+} by the equatorial 3- and the axial 4-OH hydroxyl groups of the α-Fuc pyranose ring, essential sugar-lectin hydrogen bonds are also observed with amino acids that coordinate the Ca^{2+} ion during the simulation. Key binding interactions take place with Glu354, Glu347, Asn365, and Asn349. The α-Fuc OH-3 and OH-4 act both as hydrogen bond donors and acceptors, while OH-2 acts only as a donor. Moreover, stabilizing van der Waals contacts of α-Fuc H-2 with the methylene group of Val351 are observed, as previously proposed by Drickamer et al [17]. These major interactions throughout the simulation time are summarized in Table S2.

After Fuc, Man is the monosaccharide that displays higher affinity for DC-SIGN, even though the binding is still weak[13]. It has been suggested that the smaller affinity as compared with that of Fuc could arise from the above-mentioned hydrophobic contact between Fuc H-2 and Val351, which is not present in the complex with Man. In all X-ray diffraction structures, the orientation of Man at the binding site allows its equatorial OH-3 and OH-4 groups to coordinate the Ca^{2+} ion. Accordingly, the MD simulation revealed that these groups provide stabilizing contacts with Glu354, Glu347, Asn365, and Asn349, although in an inverted fashion with respect to α-Fuc

(Table S3). In contrast, no sugar hydroxyls other than the calcium coordinating groups are involved in binding to the lectin. Remarkably, conversion from the X-ray binding mode (involving Ca^{2+} coordination by OH-3/OH-4), to the proposed pose A (involving Ca^{2+} coordination by OH-2/OH-3) is also observed during the MD simulation (beyond 100 ns of simulation time).

The simulation procedure was then tested for the complex of DC-SIGN CRD with the natural α-Man (nonfluorinated) using the two proposed binding poses involving Ca^{2+} coordination by OH-2/OH-3, A and B, as starting geometries. In pose A, the α-Man pyranose ring is entirely superimposable with that of α-Fuc (PDB 1SL5), fulfilling the same geometric features described above for such. More specifically, the oxygen and C3 atoms of the ring coincide in both sugars, while the other carbon atoms are symmetrically presented in one sugar with respect to the other (specifically $C1_{Man}$-$C5_{Fuc}$ and $C2_{Man}$-$C4_{Fuc}$, Figure 6a). In pose B, the O2-C2-C3-O3 segment of the α-Man pyranose ring is superimposed onto the O4-C4-C3-O3 of Man found on the X-ray crystal structure (PDB 2IT5), which translates into a change in the orientation of the pyranose ring plane (Figure 6b). As mentioned above, binding poses B and A are related through a 180° rotation about the C2–C3 bond bisector (Figure 5).



(a) **(b)**

Figure 6. (**a**) Superimposition of Man (blue) in binding pose A and Fuc (orange) as found in PDB code 1SL5 and (**b**) superimposition of Man in binding pose B (light blue) with Man (purple) as found in PDB code 2IT5. Oxygen atoms that coordinate Ca^{2+} are indicated in the same color as their sugar ring. The most relevant residues of the protein stabilizing the sugar at the binding site are shown in each case. All ligand hydrogens have been removed for the sake of clarity.

Interestingly, both simulations displayed rather different behaviors. When binding pose A was used as a starting geometry, the α-Man moiety is able to switch its pose along the MD trajectory (after approximately 100 ns), adopting the binding mode described for Man in the available X-ray crystal

structure of PDB 2IT5 (Figure 7). In fact, as mentioned above, this interchange between the two binding poses (the one of the crystal structure and binding pose A) takes place regardless of the starting geometry used. Thus, α-Man in binding pose A turns around the O3-Ca^{2+} bound axis, to find the usual binding pose that shows the direct interaction of the equatorial OH-3 and OH-4 hydroxyl groups with the Ca^{2+} ion. Nevertheless, during the first 100 ns of simulation where binding pose A takes place, the observed protein–sugar interactions (Table S4) resemble those described above for α-Fuc, with the corresponding geometry changes ($C1_{Man}$–$C5_{Fuc}$ and $C2_{Man}$–$C4_{Fuc}$, Table S2). Stabilizing hydrogen bonds between Man OH-4 (as donor) and Glu354 are predicted, accompanying those involving α-Man OH-2 and OH-3, as well as hydrophobic contacts between α-Man H-4 and Val351. In particular, α-Man OH-4 only acts as a hydrogen bond donor, while α-Man OH-3 acts as both a donor and an acceptor.

In the second part of the simulation, α-Man adopts the presentation described in the X-ray complexes, and O3 and O4 now coordinate the Ca^{2+} ion (Table S5). The analysis of the interactions during the whole simulation is gathered in Table S6 and Figure S1. Obviously, the most populated hydrogen bond contact is that between α-Man OH-4 and Glu354, shared in both binding modes (Figure S2), which takes place for 99% of the time the complex remains bound.

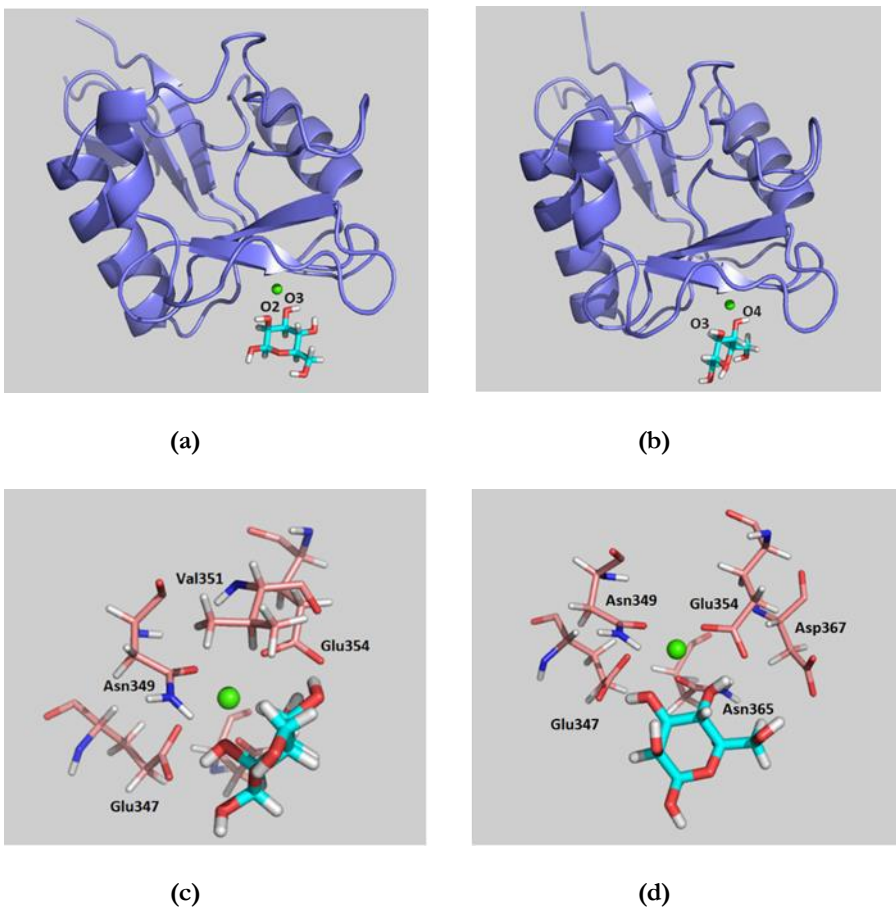


Figure 7. Snapshots from the molecular dynamics (MD) simulation of β-d-Man/DC-SIGN complex using pose A as the starting geometry. The general view is presented at the top panels, while expansions of the binding site are shown at the bottom panels. Panels (**a**) and (**c**) show Man (blue) in binding pose A. Panels (**b**) and (**d**) show Man in the binding pose of the X-ray crystal structures. Both binding modes are found during the MD simulations, regardless of the selected starting pose.

On the other hand, MD simulations for α-Man, using pose B as the starting geometry, showed only this particular binding mode during the whole trajectory (200 ns). The key contacts are described in Table S7. Herein, the participation of α-Man OH-3 in establishing hydrogen bonds is rather minor, in contrast to the observations for the reference X-ray structure. Instead, α-Man OH-4 acts now as a hydrogen bond donor, and there is no hydrophobic stabilization with Val351. Overall, the number of observed interactions for this binding mode is smaller than those for binding mode A.

Once the MD protocol had been satisfactorily evaluated with Fuc and Man, the simulations were conducted for β-4-F-Man bound to DC-SIGN

CRD. The two proposed binding poses described above were employed as starting geometries.

For the complex starting from pose A, the analysis of the trajectory revealed similar interactions to those described above for α- Fuc (Table S8), including the stabilizing van der Waals contact with Val351. The major difference in the recognition pattern for both sugars is found at position four of β-4-F-Man (equivalent to position two of α-Fuc). The presence of the fluorine atom prevents the formation of the hydrogen bond with the carboxylate group of Glu354, which was present throughout the simulation time with α-Fuc. The Man OH-2 mainly acts as hydrogen bond donor, with a smaller role as an acceptor, while Man O3 is involved as both a donor and an acceptor (Figure S3). The MD results for this analogue are also similar to those predicted in the first part of the MD simulation carried out for α-Man in this pose. The MD results also suggest that the anomeric position is not directly involved in the recognition process, in agreement with the experimental observations. As expected, the fluorine atom does not seem to play a role on the sugar stabilization at the binding site.

The analysis of the MD simulations carried out with β-4-F-Man in binding pose B showed that the main stabilizing interactions correspond to the hydrogen bond network formed between the Ca^{2+}-coordinating O2 and O3 atoms of the 4-F-Man moiety with the side chains of Glu347, Glu354, and Asn365 at the binding site. There is no participation of Asn349, contrary to the simulation with the X-ray geometry and the α-Man structure in binding pose B. In addition, Man OH-3 is now strongly involved as a donor in the hydrogen bond interaction with Glu347, taking the role of OH-4 in the regular sugar (Table S9 and Figure S4). Nevertheless, the orientations of β-4-F-Man and α-Man at the binding site are very similar (Figure S5). Again, there is no van der Waals contact between the sugar and Val351, and neither the fluorine atom nor the OH at the anomeric position, which is solvent exposed, are involved in intermolecular interactions with the lectin.

In summary, the interactions established by α-Man and β-4-F-Man in binding pose A are similar to those found for Fuc in PDB 1SL5. Obviously, for binding pose B, the sugars exhibit a totally different binding mode from that described for Man in PDB 2IT5. The detailed analysis of the MD trajectories suggests that binding pose A is more favorable than pose B, according to the number of stabilizing interactions found for both. Furthermore, the simulations suggests that the fluorinated molecule 4-F-Man is recognized by DC-SIGN with smaller affinity than the natural Man. Additionally, the corresponding poses were tested for a putative complex with α-3-F-Man, but all the trials showed the dissociation of the complex prior to the production dynamics stage, in agreement with the experimental results.

2.3. ^{1}H-STD Experiments

^{1}H-STD NMR experiments were then carried out for a mixture of DC-SIGN with α-OMe-4-F-Man to experimentally evaluate the binding epitope. As described above, the ^{19}F-NMR relaxation filter experiments showed that the anomeric configuration does not seem to play any important role in the binding event, which was also corroborated by the MD simulations. Anyhow, additional MD simulations were performed for the corresponding A and B poses of the α-OMe-Man/DC-SIGN system to verify once again that this was still the case and that the obtained trajectories were indeed similar to those described above for the free anomer. Fittingly, the use of the α-OMe derivative notably simplifies the ^{1}H-NMR spectrum, allowing to satisfactorily assign the STD intensities (Figure 8).

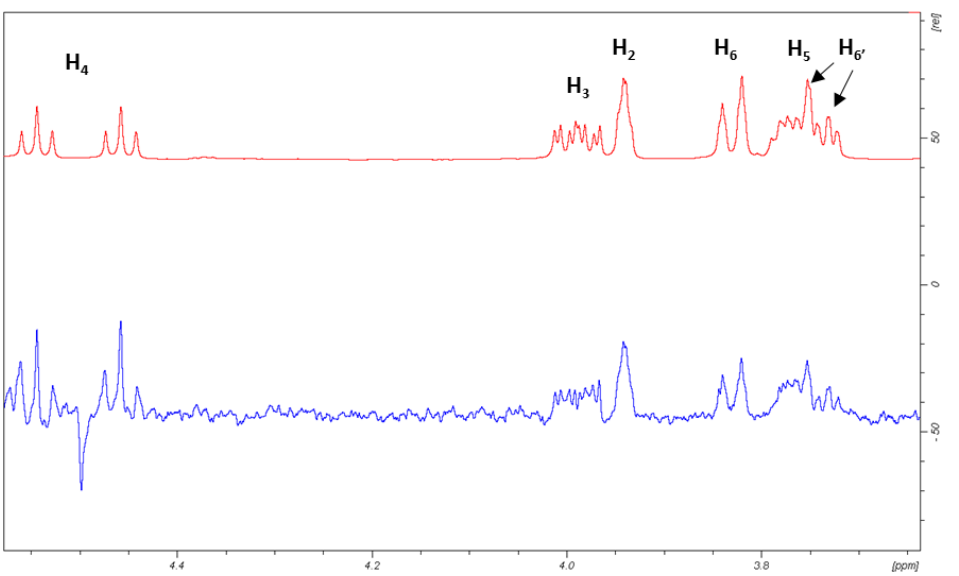


Figure 8. Expansion of the sugar region of the STD-NMR spectrum obtained for α-OMe-4-F-Man (5 mM) in the presence of DC-SIGN (65 uM, ligand/lectin ratio is 77/1). Top: Off-resonance spectrum. Bottom: STD spectrum. H-4 shows the largest STD signal. The spectra were obtained for 1024 scans at 298 K with off-resonance irradiation at 60 ppm. In this particular case, protein irradiation was set at 0.81 ppm with a saturation time of 2 s. Additional experiments were carried out irradiating at the aromatic region and employing different saturation times (0.5, 1.0, and 3.5 s). No STD was obviously observed in the absence of DC-SIGN.

Upon irradiation at the aliphatic region, H-4 showed the largest STD, followed by H-6 (65% with respect to H-4), while H-3 and H-2 showed the lowest STD responses (40%–50%, Figure 9, Table S10 and Figure S6).

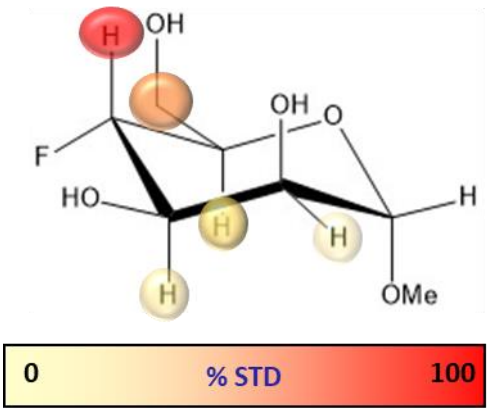


Figure 9. Relative STD profile for α-OMe-4-F-Man (5 mM) in the presence of DC-SIGN (65 uM, ligand/lectin ratio is 77/1), normalized to that of the H-4 proton (100% relative STD).

The observed large STD at H-4 is satisfactorily explained by pose A, since when using this geometry this proton is in direct contact with Val351 (Figure 10a). Furthermore, the observed STD for H-6 also agrees with this binding mode, as these protons are also close to the aliphatic chain of Val351, whereas H-2 and H-3 are far from the hydrophobic site. On the other hand, the opposite STD profile should be expected for binding mode B, in which H-2 and H-3 are facing Val351 while H-4 and H-6 are far from aliphatic hydrogens, exposed to the solvent (Figure 10b). Therefore, the STD results strongly support the presence of a major contribution of binding pose A with respect to pose B, and confirm the presence of a binding mode for 4-F-Man to DC-SIGN, which involves a direct contact of OH-2 and OH-3 with the Ca^{2+} ion. Regarding the influence of the ^{19}F nucleus in the observed STD, it is safely assumed that it is almost negligible. We have extensively explored the ^{1}H to ^{19}F saturation transfer process [6], which was found rather difficult to detect. In fact, the best way to use ^{19}F to detect STDs was to employ the regular ^{1}H to ^{1}H STD, followed by coherence transfer to ^{19}F through the $^{2}JFH/^{3}JFH$ scalar coupling constants [6,23].

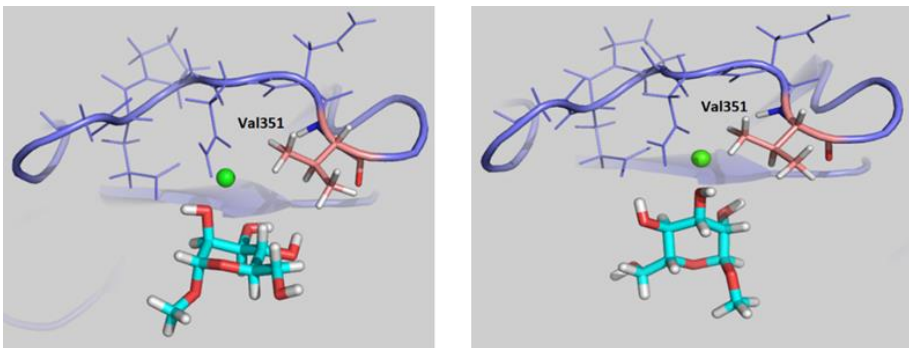


Figure 10. Model frames from the MD simulation performed for α-Me-Man bound to DC-SIGN in poses A (**a**) and B (**b**). For pose A, H-4 and H-6 are close to Val351, whereas H-2 and H-3 points towards the polar region defined by Asn349, Asn365, and Glu354. In contrast, for pose B, the alternative situation takes place, with H-2 and H-3 pointing towards Val351.

3. MATERIALS AND METHODS

3.1. Preparation of F-Monosaccharide Library

The library of mono-fluorinated monosaccharides was rationally prepared using the three basic monosaccharide units: D-glucose (Glc), D-galactose (Gal), and D-mannose (Man). The fluorine-by-hydroxyl substitution took place at every position (2-, 3-, 4- and 6-) with the exception of the anomeric site, therefore, ending up with a mixture of α/β-anomers of each fluoro-sugar. Additionally, a l-fucose (Fuc) derivative, 2-deoxy-2-F-fucose, was also included in the library. Most fluorinated monosaccharides were obtained from a commercial source (Sigma-Aldrich, Madrid, Spain; or Carbosynth, Compton, UK): 2-deoxy-2-F-glucose, 3-deoxy-3-F-glucose, 4-deoxy-4-F-glucose, 6-deoxy-6-F-glucose, 2-deoxy-2-F-galactose, 3-deoxy-3-F-galactose, 4-deoxy-4-F-galactose, 6-deoxy-6-F-galactose, 2-deoxy-2-F-mannose, 3-deoxy-3-F-mannose, 4-deoxy-4-F-mannose and 2-deoxy-2-F-fucose. The 6-deoxy-6-F-mannose and α-Methyl-4-deoxy-4-F-mannose, employed in the ^{1}H-STD experiments, were prepared by the group at Dublin and their synthesis is reported elsewhere. A stock solution of the fluoro-monosaccharide mixture was prepared in a buffer of 25 mM Tris-HCl, 150 mM NaCl, 4 mM $CaCl_2$, pH 8.

3.2. Preparation of DC-SIGN Tetramer

The fragment of DNA encoding the full extracellular domain (ECD, residues 70-404) was inserted in a pET15b plasmid and digested with NdeI and BamHI (Thermo Fischer Scientific, Madrid, Spain). The resulting plasmid was transformed on *E. coli* BL21/DE3 competent cells for subsequent overexpression (Sigma-Aldrich). A single colony holding the expression construct was inoculated into 5 mL LB medium containing 100 ug/mL ampicillin and allowed to grow overnight at 37 °C with shaking. The culture was then added to 1L of LB medium containing ampicillin and grown at 37° C until exponential phase. The cells were induced with 100 mg/L IPTG and growth continued overnight at 20 °C. The induced culture was harvested by centrifugation at 7500 g for 20 min. The pellet was suspended in 10 mL lysis buffer (10 mM Tris-HCl pH 8.0) and the cell suspension was lysed by sonication. Inclusion bodies were isolated via centrifugation at 3000 g for 20 min at 4 °C. The insoluble pellet was further solubilized by gentle rotation for 16 h at 4 °C, with 10 mL of pH 8.0 buffer containing 100 mM Tris-HCl, 6 M urea, and 0.01% v/v 2-mercaptoethanol. The mixture was centrifuged at 100,000 g for 2 hours at 4 °C, and soluble protein was dialyzed against 2 L of 100 mM Tris-HCl, pH 8.0, 0.01% *v/v* 2-mercaptoethanol, 10 mM CaCl2, 4 M urea; subsequently against the same buffer with 2 M urea, and finally with no urea. The last dialysis step was done with 100 mM Tris-HCl, pH 8.0, 10 mM CaCl2. Then, the precipitate was removed by centrifugation at 100,000 g for 30 min at 4 °C. The properly folded DC-SIGN present in the soluble fraction was purified via affinity chromatography using mannose-Sepharose (Sigma-Aldrich), washed in 10 column volumes of loading buffer 20 mM Tris, 150 mM NaCl and eluted in the same buffer containing 10 mM EDTA. Additional purification was performed by size exclusion chromatography using a Superdex 200 column and AKTA liquid chromatography system (GE Healthcare) with a pH 8.0 buffer of 20 mM Tris and 1 mM EDTA. Fractions were concentrated and the buffer was changed to 20 mM Tris, 4 mM $CaCl_2$, and 150 mM NaCl at pH 8.0, then analyzed by 4–12% SDS-PAGE. Protein concentrations were quantified using a NanoDrop UV-Vis spectrophotometer (Thermo Fischer scientific). The identification of the protein was further confirmed by LC-MS. The tetrameric state of the lectin was confirmed by TEM (Jeol JEM-1230, Tokyo, Japan) using negative staining (Figure S8).

3.3. ^{19}F-Based Screening and Chemical Mapping NMR Experiments

All NMR spectra were recorded on either Bruker AV500 or AV600 NMR spectrometers, equipped with a ^{19}F probe (Bruker Española S. A., Madrid, Spain), 5 mm SEF ^{19}F-1H with Z gradient). The ^{19}F-NMR transmitter frequencies were 470.56 MHz and 554.56 MHz, respectively. The experiments were carried out at 298 K in the same buffered solution as that of the lectin (20 mM Tris, 4 mM $CaCl_2$, and 150 mM NaCl at pH 8.0), as described above. For the ^{19}F-NMR signal assignment and sample control, standard coupled and decoupled ^{1}H and ^{19}F monodimensional spectra were acquired. The ^{19}F transmitter frequency offset and spectral width were set to –215 ppm and 50 ppm, respectively, in order to include the range of Larmor frequencies of all species in the mixture. The typical CPMG pulse sequence was used for the relaxation filtered experiments, as follows: [D-90_x-(τ-180_y-τ)$_n$-acquire], with a recovery delay $D = 4\ s$, a free evolution delay $\tau = 2\ ms$, and n spin-echo loops varying from 2 to 4000, depending on the experiment. The overall time employed corresponds to n times the applied spin echo pulse ($n2\tau = n4\ ms$) plus the almost negligible contribution of the refocusing pulse ($\approx n\ (15\mu s)$). The T_2 relaxation times are obtained by fitting the observed data points to the exponential decay curve:

$$I(t) = I_0 e^{-t/T_2} = I_0 e^{-n2\tau/T_2} \quad (2)$$

where, $I(t)$ stands for the intensity of a particular peak at time t, I_0 is its intensity at $t = 0$, and T_2 is the transversal relaxation time.

For the ^{19}F-NMR relaxation filter experiments, an aliquot of the library diluted to 1 mM (i.e, for each sugar type the concentration of each anomer is: $[\alpha] + [\beta] \cong 1\ mM$) was used. After measure transversal relaxation times of the species in the free state, $T_{2,free}$, several additions of DC-SIGN tetramer directly into the NMR tube were carried out and the observed transversal relaxation times, $T_{2,obs}$, computed. The range of ligand/protein ratios covered in the experiments varied from 23 to 235. The employed lectin tetramer (ECD) concentrations varied from 16 to 104 μM, while the ligand concentrations, varied from 600 to 890 μM.

To reduce the likelihood of observing nonspecific binding, low ligand/receptor molar ratios were explored. Moreover, relaxation filter experiments were repeated after addition of a known competitor, Manα1-3Manα1-6-Man (Carbosynth), to the F-sugar/lectin mixtures (Figure S7).

Control experiments varying the 180° pulse repetition frequency using different τ delays between 0.1 and 20 ms for 4F-Man in the presence of DC-

SIGN showed the presence of chemical exchange. This fact hampers the estimation of quantitative binding affinities from the observed T_2 values [24,25]. In any case, in order to minimize this effect short τ delays (2 ms) were used with the monosaccharide mixture. Given the close chemical nature of the ligands, the exchange contribution was expected to be relatively similar for these molecules, allowing us to make justifiable qualitative conclusions regarding binders versus non-binders.

3.4. ^{1}H-STD NMR Experiments

^{1}H-STD NMR experiments[1] were acquired on a Bruker AV600 NMR spectrometer at 298 K for a sample containing α-1-Met-4-F-Man (5 mM) and DC-SIGN CRD (65 uM), corresponding to a 77/1 ligand/protein molar ratio. A deuterated buffer for the STD experiments was used, similar to that described for the CPMG experiments: 20 mM Tris-d_{11}, 4 mM $CaCl_2$, and 150 mM NaCl at pH 8.0 in D_2O. The on-resonance irradiation was set at 0.81 ppm to selectively saturate the aliphatic region of the lectin, while the off-resonance frequency was set at 60 ppm. There were 1024 scans recorded in each experiment, using 6 s of recovery delay, with a spin-lock relaxation filter of 20 ms to reduce the receptor background signals. For selective protein saturation, the standard Bruker gradient shape file was used, SMSQ10.100, which is rectangular in shape with smoothed edges to give a gradient integral of 90% of a square pulse. The STD build-up curve for each proton was computed by conducting the experiment at four saturation times, from 0.5 to 3.5 s. The STD experimental data was fitted to the mono-exponential equation [26]:

$$STD(t) = STD^{max}(1 - e^{(-k_{sat}*t)}) \quad (3)$$

where, $STD(t)$ is the STD signal intensity of a particular proton at saturation time t, STD^{max} stands for the maximum value of the STD intensity curve at long saturation times, and k_{sat} refers to the experimental saturation rate constant. To avoid T_1 bias, the STD contribution of each proton was estimated from the slope of the STD build-up curve at saturation time zero (Table S10, Figure S6), and a relative scale of STD (fit) values referenced to proton H4, which displays the highest STD for epitope mapping discussion. The T_1 relaxation time of each proton was further estimated employing the inversion recovery pulse sequence (Table S10). Control experiments were also performed in the presence of EDTA to sequester the Ca^{2+} ion. As expected, no STD was detected under these conditions.

3.5. Molecular Dynamics Simulations

The lectin structure used in all the simulations was derived from the high-resolution crystal structure of DC-SIGN CRD bound to LNFP III (PDB code 1SL5) [17]. The structure contained three Ca^{2+} ions, one of them directly bound to the ligand. The spatial coordinates of all the sugar atoms employed in the computational studies were generated as follows:

1. For the simulations conducted using the sugar geometries of the crystal:

- The starting structure of α-Fuc from the original 1SL5 ligand was used, and the remaining parts of LNFP III were removed;
- The α-Man starting geometry was built from another DC-SIGN crystal in complex with a high-mannose derivative (PDB code 2IT5, using the geometry of the sugar bound in the major orientation) [27], removing the remaining residues.

2. For the simulations conducted using the geometry of the proposed binding poses A and B:

- The OH-2/OH-3 groups of β-4-F-Man were manually superimposed onto the OH-4/OH-3 pair of L-Fuc from the original crystal structure 1SL5 to create the model for binding pose A. Binding pose B was generated by superimposing the OH-2/OH-3 groups of β-4-F-Man to the OH-4/OH-3 pair of Man from the structure 2IT5;
- The α-Man starting structure was built as in the case of β-4-F-Man.

MD simulations were conducted using version 16 of the Amber molecular dynamics software package[28]. All simulations were run in explicit TIP3P water [29]. The ff14SB [30] and GAFF [31] force fields were employed for the description of the protein and monosaccharides (both natural and fluorinated), respectively. In addition, the MD simulations with the natural sugars (nonfluorinated) were repeated using the GLYCAM [32] force field, as a control for those simulations run using the GAFF parameters for the sugar. The Li/Merz ion parameters for the ions Ca^{2+} and Cl^- (necessary to neutralize the charge of the system) in TIP3P water (12–6 normal usage set) were employed [33].

A general protocol consisting of two minimization stages, heating of the system from 0 to 300 K, density equilibration for a total of 2 ns and production dynamics in the NPT ensemble at 300 K/1 bar was employed in every case. The long-range electrostatic interactions were treated using the particle mesh Ewald summation. In order to maintain a constant temperature, the Langevin thermostat with a collision frequency of 1 ps was used. Simulation times from 20 to 50 ns were used for the study of the 4-F-

Man/DC-SIGN complexes, and 100 to 200 ns for the natural sugar complexes.

4. CONCLUSIONS

The screening and chemical mapping strategy was based on simple ^{19}F-NMR relaxation filter experiments in order to assess that monosaccharides Fuc, Man, Glc, and Gal bind to DC-SIGN with decreasing affinity, the latter showing extremely weak, but still detectable binding. Moreover, a new binding mode for Man moieties was detected and described, which involves the direct coordination of DC-SIGN Ca^{2+} by OH-2 and OH-3 groups at the sugar recognition site, by employing a combination of ^{1}H-STD experiments and MD simulations. Our results suggest that two possible binding modes, via O3/O4-Ca^{2+} and via O2/O3-Ca^{2+}, may coexist for the natural sugar. In fact, the detailed analysis of the interaction network formed in both cases indicates that more favorable contacts are present in the new pose, resembling those found on Fuc units. The biological relevance of these findings is still to be demonstrated, but could open new avenues for the design and synthesis of novel DC-SIGN inhibitors.

Funding Sources

We thank Agencia Estatal de Investigación (Spain) for grants CTQ2015-64597-C2-1-P and 2-P, CTQ2015-68756-R, and for FPI and FPU fellowships to JDM and PV, respectively, and for the Severo Ochoa Excellence Accreditation (SEV-2016-0644). JJB also thanks the European Research Council (RECGLYCANMR, Advanced Grant no. 788143).

5. REFERENCES

1. Mayer, M.; Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117.

2. Gimeno, A.; Delgado, S.; Valverde, P.; Bertuzzi, S.; Berbís, M.A.; Echavarren, J.; Lacetera, A.; Martín-Santamaría, S.; Surolia, A.; Cañada, F.J.; et al. Minimizing the Entropy Penalty for Ligand Binding: Lessons from the Molecular Recognition of the Histo Blood-Group Antigens by Human Galectin-3. *Angew. Chem. Int. Ed.* **2019**, *58*, 7268–7272.

3. Rahkila, J.; Ekholm, F.S.; Ardá, A.; Delgado, S.; Savolainen, J.; Jiménez-Barbero, J.; Leino, R. Novel Dextran-Supported Biological Probes Decorated with Disaccharide Entities for Investigating the Carbohydrate-Protein Interactions of Gal-3. *ChemBioChem* **2019**, *20*, 203–209.

4. Ardá, A.; Jiménez-Barbero, J. The recognition of glycans by protein receptors. Insights from NMR spectroscopy. *Chem. Commun.* **2018**, *54*, 4761–4769.

5. Dalvit, C. Ligand- and substrate-based^{19}F NMR screening: Principles and applications to drug discovery. *Prog. Nucl. Magn. Reson. Spectrosc.* **2007**, *51*, 243–271.

6. Diercks, T.; Ribeiro, J.P.; Cañada, F.J.; Andre, S.; Jiménez-Barbero, J.; Gabius, H.J. Fluorinated carbohydrates as lectin ligands: Versatile sensors in ^{19}F-detected saturation transfer difference NMR spectroscopy. *Chem. A Eur. J.* **2009**, *15*, 5666–5668.

7. Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62*, 2218–2244.

8. Denavit V.; Lainé D.; Bouzriba C.; Shanina E.; Gillon É.; Fortin S.; Rademacher C.; Imberty A.; Giguère D. Stereoselective Synthesis of Fluorinated Galactopyranosides as Potential Molecular Probes for Galactophilic Proteins: Assessment of Monofluorogalactoside-LecA Interactions. *Chem. A Eur. J.* **2019**, *25*, 4478–4490.

9. Geijtenbeek, T.B.H.; Krooshoop, D.J.E.B.; Bleijs, D.A.; Van Vliet, S.J.; Van Duijnhoven, G.C.F.; Grabovsky, V.; Alon, R.; Figdor, C.G.; Van Kooyk, Y. DC-SIGN-1CAM-2 interaction mediates dendritic cell trafficking. *Nat. Immunol.* **2000**, *1*, 353–357.

10. Geijtenbeek, T.B.H.; Torensma, R.; Van Vliet, S.J.; Van Duijnhoven, G.C.F.; Adema, G.J.; Van Kooyk, Y.; Figdor, C.G. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* **2000**, *100*, 575–585.

11. Geijtenbeek, T.B.H.; Kwon, D.S.; Torensma, R.; Van Vliet, S.J.; Van Duijnhoven, G.C.F.; Middel, J.; Cornelissen, I.L.M.H.A.; Nottet, H.S.L.M.; KewalRamani, V.N.; Littman, D.R.; et al. DC-SIGN, a dendritic cell-specific HIV-1-binding protein that enhances trans-infection of T cells. *Cell* **2000**, *100*, 587–597.

12. Baribaud, F.; Doms, R.W.; Pöhlmann, S. The role of DC-SIGN and DC-SIGNR in HIV and Ebola virus infection: Can potential therapeutics block virus transmission and dissemination? *Expert Opin. Ther. Targets* **2002**, *6*, 423–431.

13. Mitchell, D.A.; Fadden, A.J.; Drickamer, K. A Novel Mechanism of Carbohydrate Recognition by the C-type Lectins DC-SIGN and DC-SIGNR. *J. Biol. Chem.* **2001**, *276*, 28939–28945.

14. Soilleux, E.J. DC-SIGN (dendritic cell-specific ICAM-grabbing non-integrin) and DC-SIGN-related (DC-SIGNR): Friend or foe? *Clin. Sci. (Lond.)* **2003**, *104*, 437–446.

15. Meiboom, S.; Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instrum.* **1958**, *29*, 688–691.

16. Feinberg, H.; Mitchell, D.A.; Drickamer, K.; Weis, W.I. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001**, *294*, 2163–2166.

17. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D.A.; Alvarez, R.; Blixt, O.; Taylor, M.E.; Weis, W.I.; Drickamer, K. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **2004**, *11*, 591–598.

18. Stambach, N.S.; Taylor, M.E. Characterization of carbohydrate recognition by langerin, a C-type lectin of Langerhans cell. *Glycobiology* **2003**, *13*, 401–410.

19. Tang, L.; Cui, Y.; Liu, W.; Kang, G.; Zhu, Y.; Jiang, D.; Wei, H.; He, F.; Zhang, G.; Gou, Z.; et al. Characterization of a Novel C-type Lectin-like Gene, LSECtin. *J. Biol. Chem.* **2004**, *279*, 18748–18758.

20. Drickamer, K. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **1992**, *360*, 183–186.

21. Feinberg, H.; Taylor, M.E.; Razi, N.; McBride, R.; Knirel, Y.A.; Graham, S.A.; Drickamer, K.; Weis, W.I. Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol.* **2011**, *405*, 1027–1039.

22. Ng, K.K.S.; Drickamer, K.; Weis, W.I. Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.* **1996**, *271*, 663–674.

23. Diercks, T.; Infantino, A.S.; Unione, L.; Jiménez-Barbero, J.; Oscarson, S.; Gabius, H.-J. Fluorinated Carbohydrates as Lectin

Ligands: Synthesis of OH/F-Substituted N-Glycan Core Trimannoside and Epitope Mapping by 2D STD-TOCSYreF NMR spectroscopy. *Chem. Eur. J.* **2018**, *24*, 15761–15765.

24. Dalvit, C.; Parent, A.; Vallée, F.; Mathieu, M.; Rak, A. Fast NMR methods for measuring in the direct and/or competition mode the dissociation constants of chemical fragments interacting with a receptor. *ChemMedChem* **2019,** *14*, 1115–1127.

25. Dalvit, C.; Piotto, M. F-^{19} NMR transverse and longitudinal relaxation filter experiments for screening: A theoretical and experimental analysis. *Magn. Reson. Chem.* **2017**, *55*, 106–114.

26. Mayer, M.; James, T.L. NMR-Based Characterization of Phenothiazines as a RNA Binding Scaffoldt. *J. Am. Chem. Soc.* **2004**, *126*, 4453–4460.

27. Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P.H.; Weis, W.I. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.* **2007**, *282*, 4202–4209.

28. Case, D.A.; JTB, R.; Betz, D.S.; Cerutti III, T.E.; Cheatham III, T.A.; Darden, R.E.; Duke, T.J.; Giese, H.; Gohlke, A.W.; Goetz, N.; et al. AMBER 16 **2016**, *Univ. California, San Fr.*.

29. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

30. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C.;Simmerling, C. ff 14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff 99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

31. Wang, J.M.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

32. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.

33. Li, P.; Roberts, B.P.; Chakravorty, D.K.; Merz, K.M. Rational design of particle mesh ewald compatible lennard-jones parameters for +2

metal cations in explicit solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.

6. SUPPORTING INFORMATION

Table S1. Transverse relaxation times deduced for the fluorine-containing monosaccharide library using the CPMG spin echo pulse sequence.

Peak	δ (ppm)	Assigment	T_2 free (s)
1	-195.041	3F-β-Glc	1.46
2	-198.140	4F-α-Glc	1.59
3	-198.675	3F-β-Man	1.58
4	-199.151	3F-β-Gal	1.74
5	-199.273	2F-β-Glc	1.61
6	-199.412	2F-α-Glc	1.90
7	-199.939	3F-α-Glc	1.42
8	-200.153	4F-β-Glc	1.56
9	-203.250	3F-α-Gal	1.77
10	-203.870	3F-α-Man	1.30
11	-204.747	2F-α-Man	1.68
12	-204.802	4F-α-Man	0.70
13	-207.385	2F-β-Gal	1.66
14	-207.542	2F-α-Gal	1.94
15	-207.730	2F-β-Fuc	1.89
16	-207.974	2F-α-Fuc	2.16
17	-208.374	4F-β-Man	1.54
18	-217.479	4F-β-Gal	1.57
19	-220.024	4F-α-Gal	1.73
20	-223.158	2F-β-Man	1.80
21	-229.641	6F-β-Gal	1.40
22	-229.779	6F-α-Gal	1.40
23	-233.625	6F-β-Man	0.99
24	-234.449	6F-α-Man	0.83
25	-234.773	6F-β-Glc	0.92
26	-235.497	6F-α-Glc	0.90

Table S2. Main interactions found on the MD simulations for L-Fuc bound to DC-SIGN in the X-Ray crystal structure. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 100 % of the total simulation time (100 ns).

	Donor	Aceptor	Avg Distance (A)	Fraction
Hydrogen Bond	α-Fuc HO4	Glu347 O	2.6	1
	α-Fuc HO2	Glu354 O	2.6	1
	α-Fuc HO3	Glu354 O	2.7	0.89
	Asn365 NH	α-Fuc O3	3.2	0.81
	Asn349 NH	α-Fuc O4	3	0.70

Charge-dipole	α-Fuc O3	Ca2+	2.65	0.99
	α-Fuc O4	Ca2+	2.6	1
	Atom1	**Atom2**	**Avg Distance (A)**	**Fraction**
Van der Waals	Val351 H	α-Fuc H2	2.4	0.90

Table S3. Main interactions found on the MD simulations performed for D-Man bound to DC-SIGN in the X-Ray crystal structure. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 100 % of the total simulation time (100 ns).

	Donor	**Aceptor**	**Avg Distance (A)**	**Fraction**
Hydrogen Bond	β-Man HO4	Glu354 O	2.65	0.99
	Asn365 NH	β-Man O4	3.1	0.96
	β-Man HO3	Glu347 O	2.65	0.88
	Asn349 NH	β-Man O3	3.1	0.50
Charge-dipole	β-Man O3	Ca2+	2.6	1
	β-Man O4	Ca2+	2.65	0.96

Table S4. Major stabilizing interactions found on the first part (ca. 112 ns) of the MD simulation performed for D-Man in binding pose A and DC-SIGN. Fraction is referred to the number of frames in which the complex is fully associated before the switch in the binding pose occurs. In this case, ca. 56 % of the simulation time (200 ns).

	Donor	**Aceptor**	**Avg Distance (A)**	**Fraction**
Hydrogen Bond	β-Man HO4	Glu354 O	2.7	0.95
	β-Man HO2	Glu347 O	2.6	0.94
	β-Man HO3	Glu354 O	2.7	0.91
	Asn365 NH	β-Man O3	3.1	0.87
	Asn349 NH	β-Man O2	3.1	0.51
Charge-dipole	β-Man O2	Ca2+	2.6	1
	β-Man O3	Ca2+	2.6	0.96
	Atom1	**Atom2**	**Avg Distance (A)**	**Fraction**
Van der Waals	Val351 H	β-Man H4	2.8	0.86

Table S5. Major stabilizing interactions found on the second part (ca. 113-185 ns) of the MD simulation performed for DC-SIGN and D-Man in binding pose A as the starting configuration. These interactions are similar to those found on the X-Ray structure of D-Man at DC-SIGN's binding site (Table S3). Fraction is referred to the number of frames in which the complex is fully associated after the switch in the binding pose (113-185 ns). In this case, ca. 36 % of the simulation time (200 ns).

	Donor	**Aceptor**	**Avg Distance (A)**	**Fraction**
Hydrogen Bond	β-Man HO4	Glu354 O	2.7	1
	Asn365 NH	β-Man O4	3.0	0.98

	β-Man HO3	Glu347 O	2.7	0.91
	Asn349 NH	β-Man O3	3.1	0.17
Charge-dipole	β-Man O3	Ca2+	2.6	1
	β-Man O4	Ca2+	2.7	1

Table S6. Main interactions found on the complete MD simulations of the complex between DC-SIGN/D-Man starting from binding pose A. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 93.5 % of the total simulation time (187 ns). The relative populations of both binding modes in this MD simulation can be extracted from the relative fractions of the oxygen-calcium interactions.

Type of interaction	Donor	Aceptor	Avg Distance (A)	Fraction
Hydrogen Bond	β-Man HO2	Glu354 O	2.7	1
	β-Man HO2	Glu347 O	2.7	0.59
	β-Man HO3	Glu354 O	2.7	0.57
	Asn365 NH	β-Man O3	3.1	0.54
	Asn365 NH	β-Man O4	3	0.40
	β-Man HO3	Glu347 O	2.7	0.39
Charge-dipole	β-Man O2	Ca2+	2.6	0.62
	β-Man O3	Ca2+	2.6	0.99
	β-Man O4	Ca2+	2.7	0.4
	Atom1	**Atom2**	**Avg Distance (A)**	**Fraction**
Van der Waals	Val351 H	β-Man H4	2.8	0.57

Table S7. Key interactions in the MD simulations performed for the complex between D-Man and DC-SIGN using binding pose B as starting geometry. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 100 % of the total simulation time (200 ns).

	Donor	Aceptor	Avg Distance (A)	Fraction
Hydrogen Bond	Man HO4	Glu347 O	2.7	0.98
	Asn365 NH	Man O2	3.1	0.95
	Man HO2	Glu354 O	2.7	0.91
	Asn349 NH	Man O3	3.2	0.36
	Man HO3	Glu347 O	2.8	0.28
Charge-dipole	β-Man O2	Ca2+	2.6	1
	β-Man O3	Ca2+	2.7	1

Table S8. Main interactions in the MD simulations performed for the complex between 4-F-Man and DC-SIGN using binding pose A as starting geometry. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 15 % of the total simulation time (100 ns).

	Donor	Aceptor	Avg Distance (A)	Fraction
Hydrogen Bond	β-4F-Man HO2	Glu347 O	2.6	1
	β-4F-Man HO3	Glu354 O	2.7	0.98
	Asn365 NH	β-4F-Man O3	3.2	0.71
	Asn349 NH	β-4F-Man O2	3	0.55
Charge-dipole	β-4F-Man O2	Ca2+	2.6	1
	β-4F-Man O3	Ca2+	2.6	1
	Atom1	**Atom2**	**Avg Distance (A)**	**Fraction**
Van der Waals	Val351 H	β-4F-Man H4	2.7	0.94

Table S9. Main interactions In the MD simulations performed for the complex between 4-F-Man and DC-SIGN using binding pose B as starting geometry. Fraction is referred to the number of frames in which the complex is fully associated. In this case, ca. 40 % of the total simulation time (100 ns).

	Donor	Aceptor	Avg Distance (A)	Fraction
Hydrogen Bond	Asn365 NH	β-4F-Man O2	3.05	0.99
	β-4F-Man HO3	Glu347 O	2.7	0.94
	β-4F-Man HO2	Glu354 O	2.8	0.75
	Ser360 OH	β-4F-Man O6	3.0	0.23
Charge-dipole	β-4F-Man O2	Ca2+	2.63	0.98
	β-4F-Man O3	Ca2+	2.58	1

Table S10. Initial STD slope (t=0) for each proton of the 4F-Man□ OMe derivative calculated from the fitted exponential equation. % STD (fit) values are normalized with respect to H4, set to 100%. As expected, T_1 relaxation values for H6s are considerable lower than those exhibited by the other protons, highlighting the importance of using initial slopes (at t_{sat}=0) to avoid misinterpretation of ligand-receptor proton distances. H5 and H6 build-up curves practically converge at 3.5 of protein saturation, whereas H2 and H4 are still growing at that point.

	STDmax*Ksat*10^{-3}	% STD (fit)	T1 (s)
H4	4.323	100.00	2.24
H3	1.987	52.07	2.26
H2	1.662	40.25	2.86
H6	2.808	64.94	0.94
H5	2.590	59.90	-

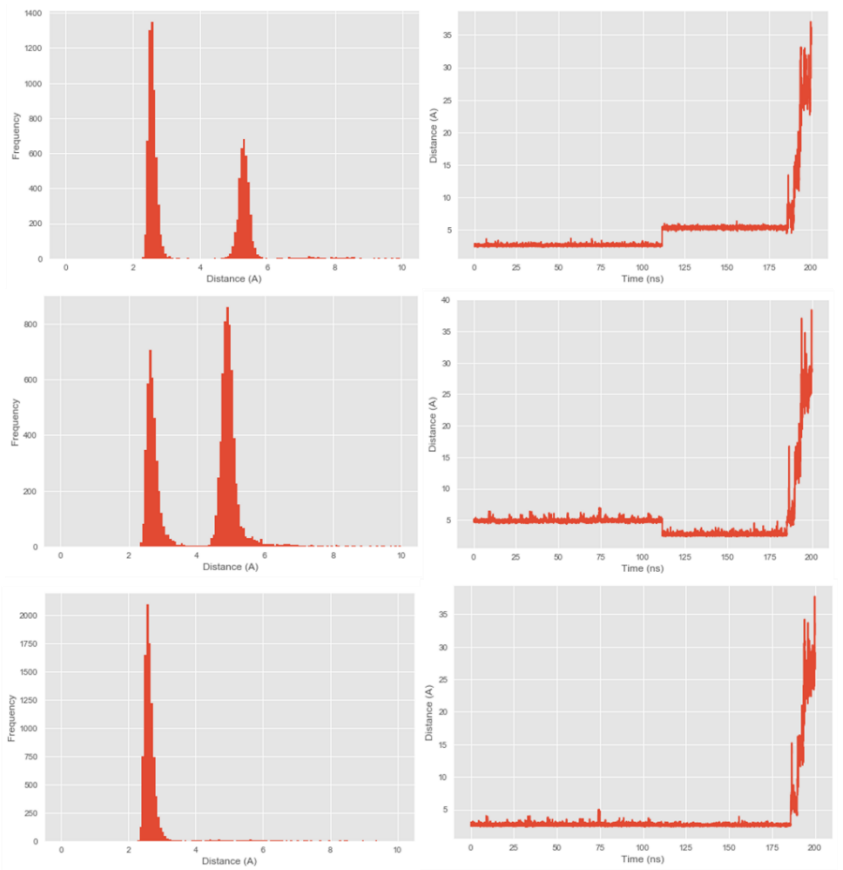


Figure S1. The complete MD simulation of the complex between DC-SIGN/D-Man starting from binding pose A shows the sugar moiety switching its position from the starting geometry. The distance between Man O2, O3 and O4 groups and the Ca2+ ion is shown. Top) frequency and trajectory of the O2-Ca2+ distance. Middle) frequency and trajectory of the O4-Ca2+ distance. Bottom) frequency and trajectory of the O3-Ca2+ distance. The positional switch between binding pose A (O2/O3 attached to the Ca2+ ion) and the X-Ray binding pose (O3/O4 attached to the Ca2+) ion takes place at ca. 113 ns.

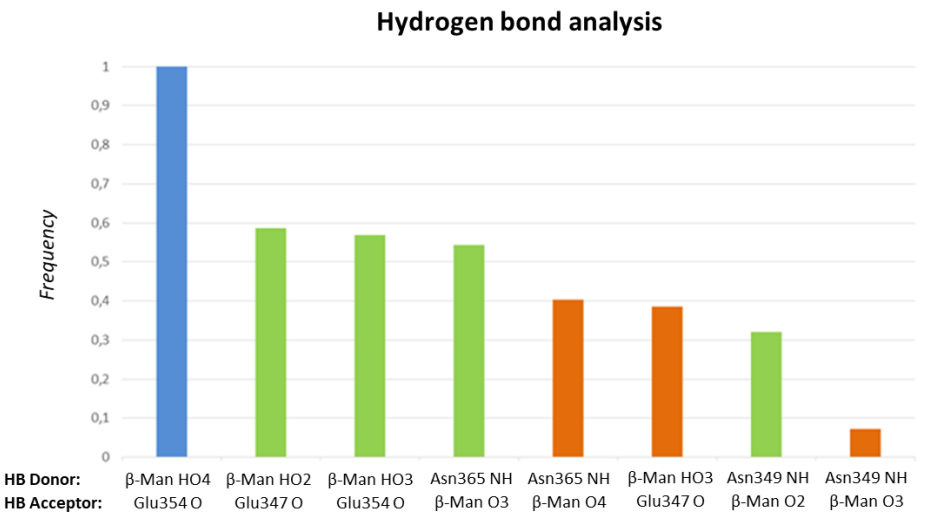


Figure S2. Populations of the different HB found along the MD simulation (187 ns) of the D-Man/DC-SIGN complex using binding pose A as starting geometry, which switches to the X-Ray crystallographic pose. The blue bar refers to the HB shared in both binding modes. The green and orange are characteristic for Man in binding pose A and in the X-Ray structure, respectively.

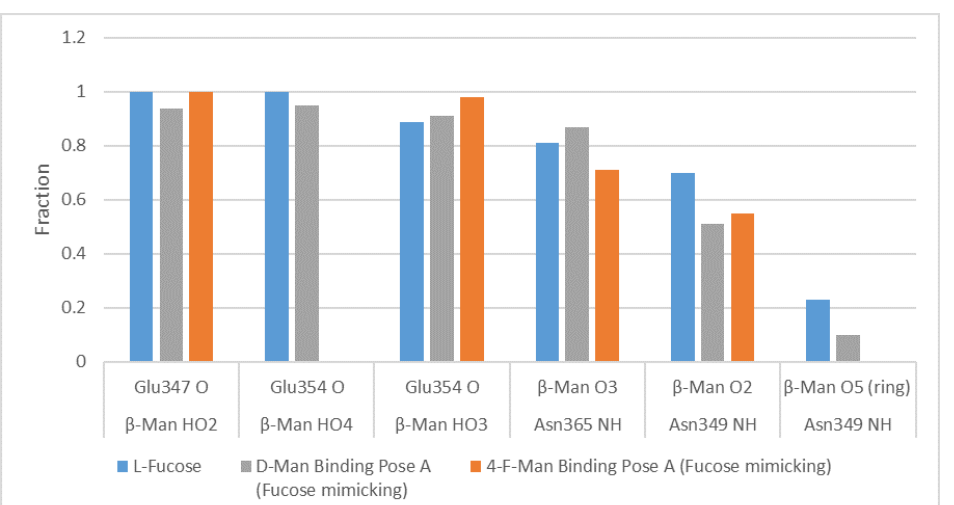


Figure S3. Populations of the different HB found along the MD simulation of the D-Man/DC-SIGN, 4-F-Man/DC-SIGN, and L-Fuc/DC-SIGN complexes using the corresponding binding poses A as starting geometries. For D-Man and L-Fuc, the data refer to the complete 100 ns simulation time, while for 4-F-Man, it is referred to the time the complex remain fully associated, 15 ns.

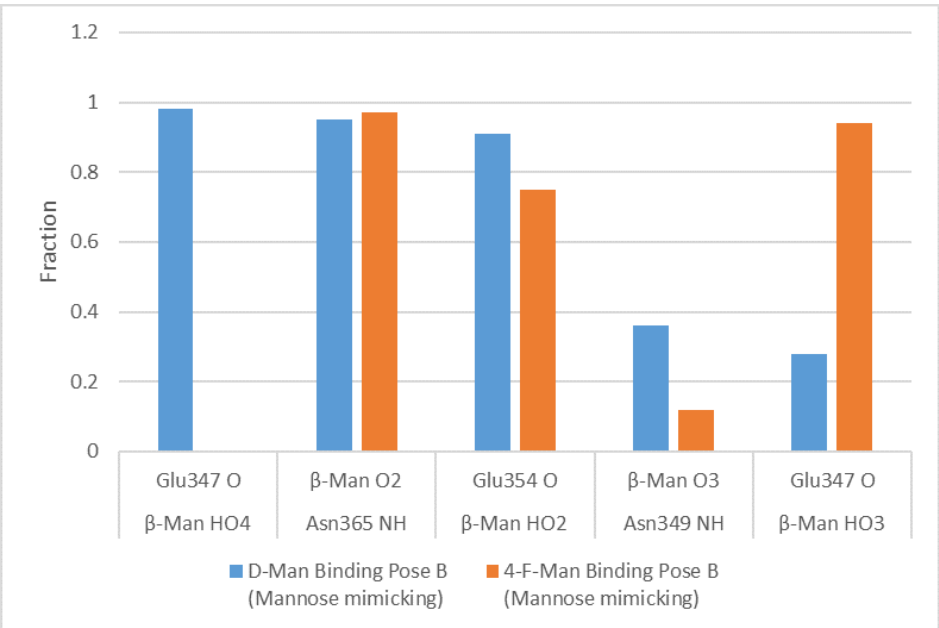


Figure S4. Populations of the different HBs found along the MD simulation of the D-Man/DC-SIGN and 4-F-Man/DC-SIGN complexes using binding pose B as starting geometry. The role of Man HO3 as donor is drastically enhanced in the 4-F-Man analogue with respect to Man, obviously due to the lack of HO4 in the fluorinated compound. For D-Man, the data refer to the complete 200 ns MD simulation time, while for 4-F-Man, it is referred to the time the complex remain fully associated, 50 ns.

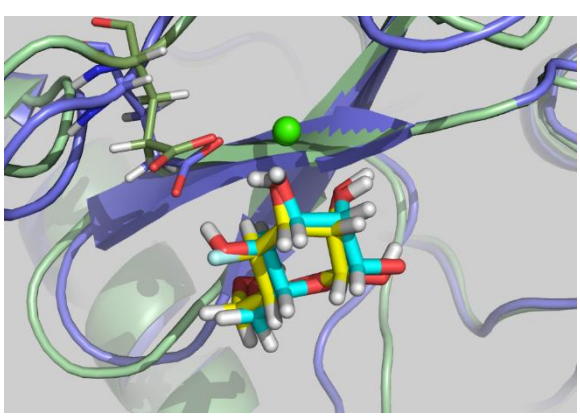


Figure S5. Superimposition of two snapshots taken from the MD simulations carried out for D-Man and 4-F-Man (yellow) bound to DC-SIGN using pose B. The green backbone corresponds to the complex with 4-F-Man. The key difference involves the role of Glu347, which establishes HB interactions with OH4 (D-Man) and with OH3 (4-F-Man).

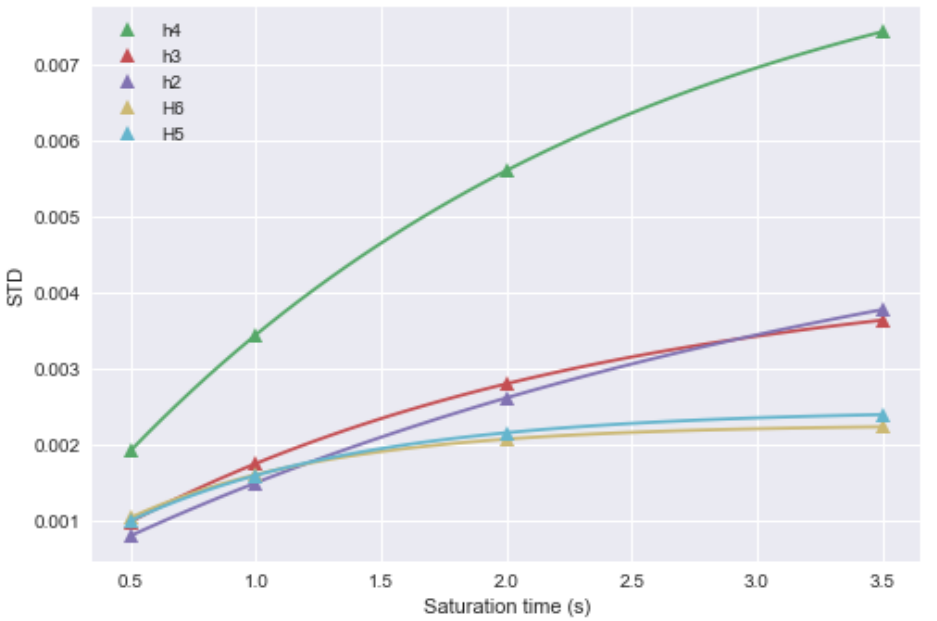


Figure S6. STD build-up curves for 4F-ManαOMe as a function of the saturation time. The STD^{max} value and the saturation rate constant k_{sat} were derived by least-squares fitting of the experimental data (triangle) to the monoexponential function $STD = STD^{max}(1 - e^{(-k_{sat}*t)})$.

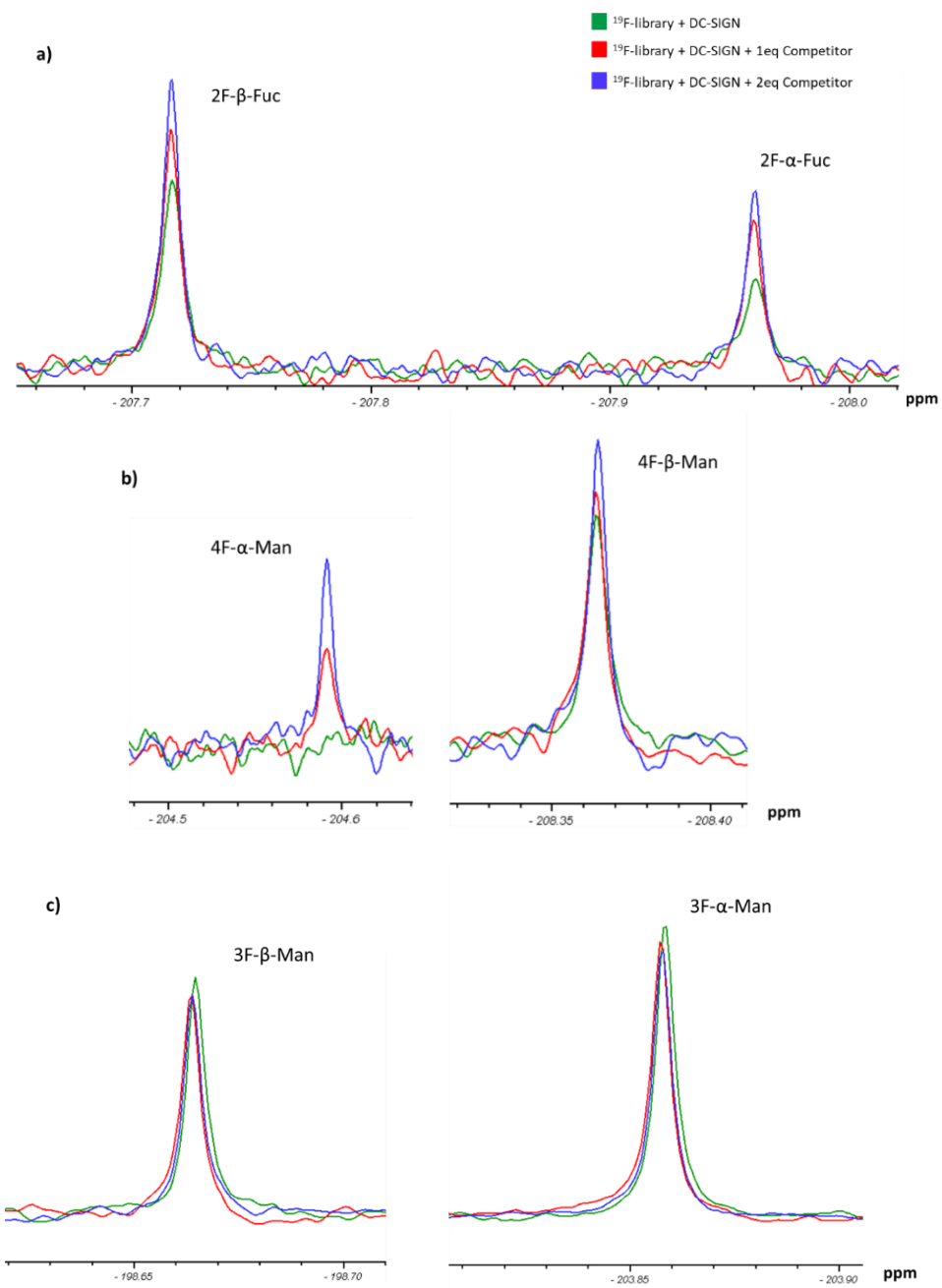


Figure S7. Superimposition of ^{19}F-NMR relaxation filter spectra in absence (green) and presence (red and blue) of a competing molecule (Manα1-3Manα1-6-Man) to assess the specific binding of some fluorinated monosaccharides to DC-SIGN. The T_2 relaxation filter duration was 241 ms. The equivalents of the competitor are with respect to the concentration of the fluorinated monosaccharide mixture (for each sugar type the concentration of each anomer is: $[\alpha] + [\beta] \cong 0.55\ mM$). The close-ups corresponds to the binders α- and β-2-F-Fuc (a)), α- and β-4-F-Man (b)), and the non-binders α- and β-3-F-Man.

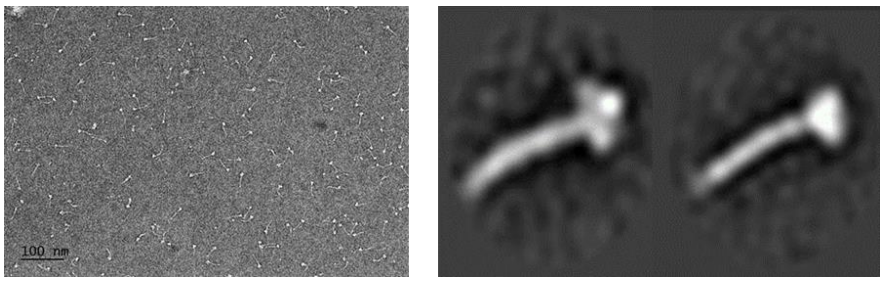


Figure S8. Electron microscopy pictures using negative staining. Representative areas of digital micrographs are shown at the left. Selected class averages resulting from 2D reference-free alignment at the right. Model exhibit a tail with a length of 245–250 Å and a diameter of 25–30 Å, whereas the head has a length of 70 –75 Å and transverse dimensions in the range of 50 to 80 Å.

CHAPTER 5

The interaction of fluorinated glycomimetics with DC-SIGN: multiple binding modes disentangled by the combination of NMR methods and MD simulations

THE INTERACTION OF FLUORINATED GLYCOMIMETICS WITH DC-SIGN: MULTIPLE BINDING MODES DISENTANGLED BY THE COMBINATION OF NMR METHODS AND MD SIMULATIONS

Reproduced with permission from the authors: Angela S. Infantino, Pablo Valverde, Tammo Diercks, Sandra Delgado, Niels-Christian Reichardt, Ana Ardá, Francisco Javier Cañada, Stefan Oscarson and Jesús Jiménez-Barbero. 'The Interaction of Fluorinated Glycomimetics with DC-SIGN: Multiple Binding Modes Disentangled by the Combination of NMR Methods and MD Simulations', *Pharmaceuticals* **2020**, *13*(8), 179.

Abstract: Fluorinated glycomimetics are frequently employed to study and eventually modulate protein–glycan interactions. However, complex glycans and their glycomimetics may display multiple binding epitopes that enormously complicate the access to a complete picture of the protein–ligand complexes. We herein present a new methodology based on the synergic combination of experimental ^{19}F-based saturation transfer difference (STD) NMR data with computational protocols, applied to analyze the interaction between DC-SIGN, a key lectin involved in inflammation and infection events with the trifluorinated glycomimetic of the trimannoside core, ubiquitous in human glycoproteins. A novel 2D-STD-TOCSYreF NMR experiment was employed to obtain the experimental STD-NMR intensities, while the COmplete Relaxation MAtrix Analysis (CORCEMA-ST) was used to predict that expected for an ensemble of geometries extracted from extensive MD simulations. Then, an in-house built computer program was devised to find the ensemble of structures that provide the best fit between the theoretical and the observed STD data. Remarkably, the experimental STD profiles obtained for the ligand/DC-SIGN complex could not be satisfactorily explained by a single binding mode, but rather with a combination of different modes coexisting in solution. Therefore, the method provides a precise view of those ligand–receptor complexes present in solution.

1. INTRODUCTION

Molecular recognition processes are essential for communication of biological entities. In this context, the three-dimensional structures of the partners that are involved in the interaction are strongly related to the corresponding biological response. Within biomolecules in nature, the chemical diversity of carbohydrates (oligosaccharides, glycans) endows them with a strong capacity for coding messages and translating them. [1–5].

The last years have enjoyed large advances in the Glycoscience field [6–11], providing new tools and perspectives to expand the understanding on the structure-function relationship of the events mediated by sugar–lectin interactions. These developments have also led to important progress in the use of glycomimetics [12,13], sugar analogues devised to bind lectins, antibodies, or enzymes [14–16]. These synthetic molecules are structurally related to sugars although granted with altered physical chemical properties through structural modifications [17]. One of the most common alterations uses fluorine [18], the less abundant halogen in metabolites in nature, but highly present in drugs and pharmaceuticals [19,20], where it may boost adsorption and distribution properties [21].

In the NMR context, ^{19}F spectroscopy is especially suitable for monitoring molecular recognition events. Besides, the intrinsic sensitivity of the ^{19}F nucleus, ^{19}F-NMR parameters provide additional chemical and structural information complementary to that achieved by using standard ^{1}H and ^{13}C-NMR [22]. Particularly in glycans, the tremendous overlap of ^{1}H NMR signals can be overcome by introducing specific fluorine atoms in the molecule. Thus, glycomimetics decorated with fluorine atoms may be used in medicinal and pharmaceutical chemistry as well as in chemical biology and their interactions can be monitored by ^{19}F-NMR experiments [23–26].

Among glycan binding proteins, C-type lectins mediate efficient immune responses. They feature broad recognition profiles that nevertheless can be precisely tuned with the chemical complexity of the saccharide. The Dendritic Cell-Specific ICAM-3-Grabbing Non-integrin, DC-SIGN, is one of the most widely studied C-type lectins. This lectin is highly abundant in dermal dendritic cells (DCs) [27–29], and is integrated in the asialoglycoprotein receptor-(ASGPR)-related and endocytic DC membrane receptors [30,31], which play key roles in innate immunity [29,32]. It is directly involved in HIV infectivity [33–36] and it is also crucial for the development of other pathogen-involved diseases, such as Ebola [37] and dengue [38–40], which has boost its attraction as a therapeutic target.

With respect to its glycan binding selectivity, DC-SIGN represents the paradigm of binding promiscuity [41–43]. This fact is strongly related to its

3D architecture. The primary Ca^{2+} site is located at a flat surface, exposed to the solvent and displays very few projecting amino acid sidechains. Thus, this construction enables the adaptation of a large variety of sugars. DC-SIGN preferentially binds D-mannose (Man) and L-fucose (Fuc), but also D-glucose (Glc), N-acetyl-D-glucosamine (GlcNAc), and N-acetyl-D-mannosamine (ManNAc) [42]. Despite this fact, its biological and biomedical interest have encouraged the design, synthesis, and applications of diverse glycomimetics.

The possibility of employing fluorinated glycomimetics to address their recognition features in comparison to those of their natural counterparts was explored in this study. In particular, we explored the binding features of a trifluorinated glycomimetic (**1**) of the ubiquitous trimannoside present in the inner core of N-glycans (**1b**), and its two constituent difluoro mannoside disaccharides **2** and **3** (Figure 1), to DC-SIGN by using an integrated approach of ^{19}F-based STD-NMR spectroscopy methods and computational techniques. The novel 2D-STD-TOCSYreF NMR experiment [44] was employed, in combination with COmplete Relaxation and Conformational Exchange MAtrix Analysis of Saturation Transfer (CORCEMA-ST) and extensive Molecular Dynamics (MD) simulations, to deduce the binding epitope of each ligand. For that purpose, we designed an in-house computer program that is able to find the ensemble of structures that provide the best fit to the experimental NMR data. Remarkably, none of the experimental STD profiles obtained for each ligand bound to DC-SIGN can be satisfactorily explained by a single binding mode, but rather with a combination of different modes coexisting in solution.

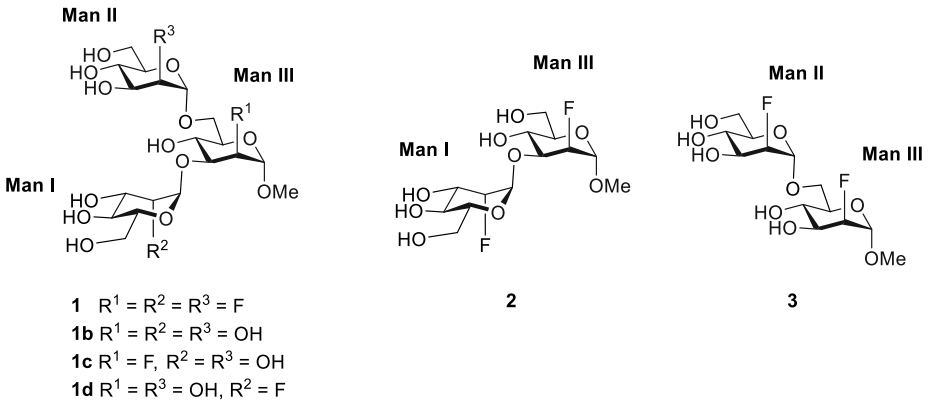


Figure 1. The trifluorinated trisaccharide glycomimetic **1** and its disaccharide components (**2,3**). The natural trimannoside core (**1b**) and related monofluorine-containing glycomimetics (**1c,1d**) are also shown. The non-reducing terminal residues are labelled as Man I (1→3 linkage) and Man II (1→6 linkage). The reducing (branched) residue in the trisaccharide is Man III.

2. RESULTS AND DISCUSSION

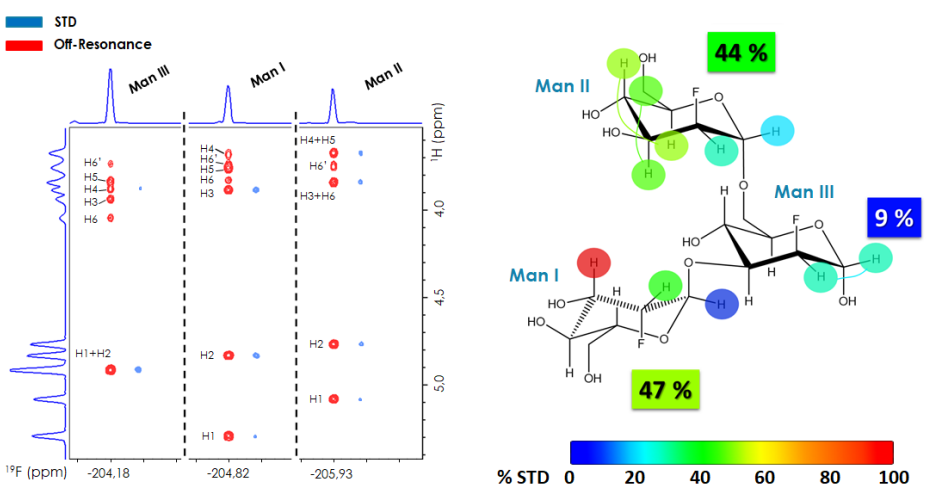
The interaction of DC-SIGN with Man-containing oligosaccharides has been widely studied at the molecular level [43,45–47]. Man moieties interact with DC-SIGN by anchoring the Ca^{2+} at the primary binding site through O3-O4 (or its alternative O4-O3 pose) [45] or O2-O3 diols [48]. For high-Man ligands with the core trimannoside (**1b**) motif, the Manα1-3Man branch coordinates the Ca^{2+} ion of the lectin, whereas the Manα1-6Man branch is typically accommodated in a secondary site facing Phe313 (PDB codes: 1SL4, 1K9I) [43,47]. A different binding mode to DC-SIGN has been described (PDB: 2XR5) [49] for a pseudo-1,2-mannobioside ligand consisting of a Man moiety attached to a cyclohexane ring. In this case, the Man unit binds to the Ca^{2+} ion through the O3-O4 diol, while the cyclohexane ring folds towards the Val351 side chain, establishing van der Waals contacts.

In order to determine the bound conformations of each fluorinated glycomimetic to DC-SIGN, a protocol that combines the versatility of STD-NMR experiments with extensive computational data was adopted. First, 2D-STD-TOCSYreF NMR spectra [44] were recorded to take advantage of the increased resolution of these ^{19}F-NMR experiments over the regular ^{1}H-NMR based methods. In this manner, the experimental STD data, which contain the binding epitope information, were measured. In a parallel manner, the possible binding poses for each ligand, based on previously reported X-ray [43,45,47,49] and NMR data [46,48,50], were employed as starting geometries for extensive MD simulations. For every proposed binding mode, a large number of frames from the computed MD trajectories

were submitted to CORCEMA-ST calculations [51,52] to predict their expected STD profiles. Finally, a computer program was developed to iteratively search for the best combination of MD frames that are able to quantitatively fit the experimental STD-NMR data. As a result, the contribution of every binding mode present in solution was assessed.

2.1. 2D-STD-TOCSYreF experiments and affinity measurements

The interaction between DC-SIGN and the three Man-based 2-deoxy-fluoro oligosaccharides (Figure 1) was first analyzed using 2D-STD-TOCSYreF experiments [44] (Figure 2). A long TOCSY mixing time (128 ms) was used to extend the polarization to all protons within the same spin system for each Man ring. The enhanced resolution of the 2D-TOCSYreF allowed the assignment of most of the STD NMR cross peaks with high reliability for the three fluorinated compounds (Figure 2 and Table 1). Generally speaking, the inherent loss of sensitivity due to the heteronuclear reF transfer in the 2D-STD-TOCSYreF experiments [44] is overcompensated by the gained spectral dispersion. Nevertheless, overlapping of some signals still occurred in **3**, as indicated in Figure 2.



1

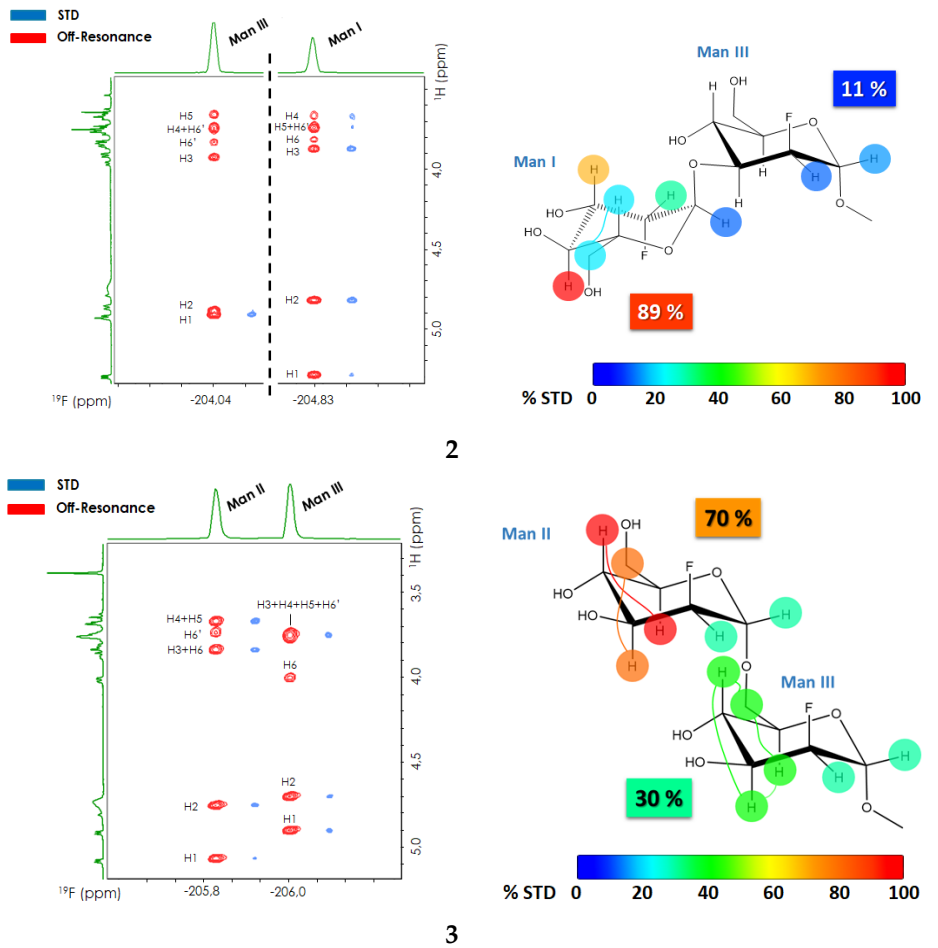


Figure 2. 2D TOCSYreF off-resonance spectrum (red) and 2D STD-TOCSYreF peaks (blue) obtained for **1** (top panel), **2** (central panel) and **3** (bottom panel) in the presence of DC-SIGN. The ^{19}F-NMR frequencies are given in the abscissa axis and the ^{1}H-NMR frequencies in the ordinate axis. The STD cross peaks are shifted to high field in the ^{19}F frequency to facilitate visual inspection. Every blue cross peak corresponds to the STD-NMR intensity of the particular ^{1}H-NMR nucleus in a given monosaccharide, which is encoded by the ^{19}F frequency. The ligand concentration was 1.4 mM in each case, with 9.14 μM of DC-SIGN tetramer, thus containing four carbohydrate recognition domains (CRD) (ligand/protein-CRD ratio is 38:1). The on-resonance irradiation was set to 0.85 ppm, with a saturation time of 2 s. The relative STD values, shown by colored circles, are obtained from the ratio between the corresponding peaks in the 2D STD-TOCSY-reF and 2D-TOCSYreF off-resonance spectra, and normalized to the largest STD in each molecule. The circles connected by a colored thin line correspond to overlapped signals in the spectrum, so that the STD intensity is shown as 'spread' between the corresponding ^{1}H nuclei (meaning each of these protons may contribute or not to the total STD observed). In this case, the added STD intensity is given. The total relative STD per Man ring for each molecule is also specified with a colored tag, using the same color code as the circles.

The 2D STD-TOCSYreF analysis of each molecular system revealed the parts of the ligand receiving efficient saturation from the protein (Table 1). For **1**, the terminal 1-3 and 1-6 branches exhibited the highest relative STD intensities (47% and 44%, respectively, normalized to the global observed STD), while the STD effects for the central Man III were notoriously weaker (9%). For disaccharide **2**, a large STD was observed at Man I (89% relative STD) versus Man III (11%), whereas for **3,** the observed STDs were 70% at Man II and 30% at Man III. Fittingly, these results agree to some extent with those observed for the trisaccharide **1**, whose STD profile is roughly the addition of the observed STDs for the two disaccharides. Noteworthy, the STD displayed by the central Man III is much lower than that of the terminal Man I and Man II units. Indeed, this behavior is also noticeable on the dimannosides **2** and **3**, for which the reducing Man III unit always displays the lowest STD effects. In all cases, H3 or H4 of either Man I (in **1** and **2**) or Man II (in **3**) always showed the highest STD-NMR intensity when on-resonance saturation was set at the aliphatic region (0.85 ppm).

Table 1. Normalized STD NMR data from 2D-STD-TOCSYreF experiments for the fluorinated glycomimetics **1**-**3**.

Ligand	H	% STD
1	ManI-1	16
	ManI-2	42
	ManI-3	100
	ManIII-1/2	31
	ManII-1	19
	ManII-2	31
	ManII-3/6	47
	ManII-4/5	52
2	ManI-1	15
	ManI-2	33
	ManI-3	74
	ManI-4	100
	ManI-5	22
	ManIII-1	17
	ManIII-2	14
3	ManIII-1	33
	ManIII-2	31
	ManIII-3/4/5/6	47
	ManII-1	31
	ManII-2	45
	ManII-3/6	86
	ManII-4/5	100

Although the experimental information obtained is reliable, it is not straightforward to translate it into a particular binding pose. As mentioned above, multiple binding modes have been widely observed for Man-containing ligands when interacting with DC-SIGN. In particular, Man moieties may interact with DC-SIGN through direct coordination of the Ca^{2+} via O3-O4 or the 'inverse' O4-O3 [45], and also through the O2-O3 diol [48]. Although the presence of a fluorine atom at position C-2 in **1**-**3** precludes this binding mode herein, the possible binding poses for these ligands in solution are still numerous.

In addition to the binding epitope, the binding affinities of **1-3** to DC-SIGN were also deduced by competitive ^{19}F-R_2 filtered NMR experiments, using 6-F-ManαOMe (K_D 1.3 mM) as a spy molecule (see experimental section and Figure S1 in SI). In order to gain insights into the effects that the different fluorination patterns may have in the binding event, the affinities of **1b**, the natural analogue of **1**, as well as **1c** and **1d**, two partially fluorinated trisaccharides, were also estimated. (Table 2, Figure S1).

Table 2. Affinities determined by competitive ^{19}F-R_2 filtered NMR experiments.

Structure	[a] K_I (mM)	SD (mM)	[b] r^2
1	2.5	0.7	0.91
1b	1.1	0.2	0.97
1c	1.0	0.1	0.99
1d	1.1	0.2	0.98
2	9.6	3.5	0.77
3	4.0	0.6	0.98

[a] K_I is determined by non-linear least squares fitting (see SI). [b] Coefficient of determination for the fitting.

The results indicate that the best binders are the natural molecule **1b** and the two monofluorinated analogues, **1c** and **1d**, which essentially display the same affinity. The trifluorinated trisaccharide **1** is two-fold weaker, while the difluorinated disaccharides **2** and **3** are nine- and four- fold weaker ligands, respectively.

Interestingly, mono-fluorination at C-2 of **1b** at the branched Man III (to give **1c**) or at the terminal Man I (to give **1d**) do not appreciably influence affinity. However, when all the three Man moieties are fluorinated (**1**), the affinity drops by half. In a qualitative manner, these data suggest that, individually, the contribution of Man I and Man III's OH-2 to the interaction with the lectin is negligible, while that of Man II is relevant. Whether the loss in affinity observed for **1** is caused by the fluorination of Man II or rather by a combined effect of the simultaneous fluorination of the three Man rings, cannot be directly inferred solely from these affinity measurements.

2.2. Molecular Dynamics Simulations

Molecular Dynamics simulations were then employed to provide possible 3D models of the complexes between DC-SIGN and each fluorinated glycomimetic. As mentioned above, X-Ray crystallography and NMR studies have demonstrated that Man-containing saccharides may bind this lectin in different poses, given the diverse possibilities of the Man O3 and O4 atoms to coordinate the key Ca^{2+}. Therefore, several starting geometries were generated for the complexes of glycomimetics **1**-**3** with DC-SIGN, by considering the following restraints:

- The vicinal O3 and O4 atoms of any Man residue coordinate the key Ca^{2+}.
- The corresponding Man moiety was superimposed with the equivalent Man unit deposited on PDB 2IT5 [45]. This crystallographic structure contains two alternative O3-O4 (major) and O4-O3 (minor) poses, respectively.

Thus, there are four potential binding modes for **1** (O3-O4 and O4-O3 alternatives at Man I and the equivalent ones at Man II). Similarly, there are four potential poses for **3** (through Man II and Man III) while there are only two (through Man I) for analogue **2**, since Man III O3 is now involved in the glycosidic linkage. Overall, 10 different starting geometries of the DC-SIGN-glycomimetics complexes were tested and analyzed by long MD simulations in explicit water.

In order to classify the different binding poses, a nomenclature for each ligand was defined as follows: "name of the Man unit interacting with Ca^{2+} + identifier of the coordinating oxygen closer to the viewer + identifier of the other coordinating oxygen" (Figure 3). For instance, the two possible binding modes for **2** are dubbed "ManI_O3-O4" and "ManI_O4-O3". The modelled complexes are gathered in Table 3, while their corresponding 3D structures are given in Figure S3.

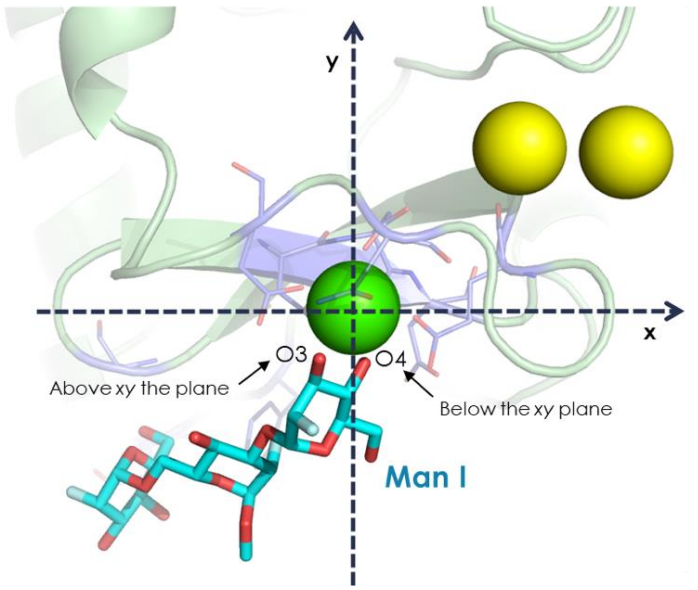


Figure 3. Schematic view of one possible complex between **1** and DC-SIGN describing the employed Cartesian system to describe the coordinating oxygen atoms. The geometric center of the three Ca^{2+} ions of DC-SIGN is on the *xy* plane, the key Ca^{2+} (green sphere) is set at the origin, and the other two ions (yellow spheres) are placed in the first quadrant. The coordinating oxygen closer to the viewer is first specified (O3 in this case), and the other one (O4) is named later. Therefore, this binding pose is "ManI_O3-O4".

The stability of the different complexes during the MD simulations was evaluated. In general, the major stabilizing contacts with the lectin involve the Man unit that coordinates the Ca^{2+} ion either via O3-O4 or O4-O3. A well conserved hydrogen bond (HB) network is observed for the three glycomimetics in all binding poses. The OH groups at positions 3 and 4 establish HB contacts with Glu354, Glu347, Asn365 (>80% of simulated time in which the complex is associated), and Asn349 (50–80%). More specifically, for the O3-O4 binding modes, OH-3 contacts the side chains of Glu347 and Asn349, whereas OH-4 interacts with Glu354 and Asn365. These interactions are swapped for the O4-O3 poses (see Figure S5 in the SI for a detailed summary of all interactions).

Table 3. Modelled structures for MD simulations. The stability of the different complexes is also given as the association time averaged for the different MD replicas (6–12 replicas of 400 ns each). The ManI_O3-O4 pose for **1**, ManI_O3-O4 for **2** and ManII_O3-O4 for **3** are the most stable ones.

Ligand	Binding Poses	Complex stability (ns) [1]
1	ManI_O3-O4	321
	ManI_O4-O3	9
	ManII_O3-O4	27
	ManII_O4-O3	22
2	ManI_O3-O4	77
	ManI_O4-O3	18
3	ManII_O3-O4	400
	ManII_O4-O3	22
	ManIII_O3-O4	19
	ManIII_O4-O3	1

The most stable binding mode found for **1** was ManI_O3-O4, where the ligand adopts an extended conformation which resembles those present in PDB structures 1SL4 and 1K9I [43,47]. In the alternative O4-O3 pose, the ligand is folded over Val351, as found for the pseudo-1,2-mannobioside in 2XR5. Several van der Waals contacts are detected along the simulations, mainly involving H-3, H-4, and H-5 of any of the three Man moieties, depending on the particular binding mode (Figure S5). Finally, for the ManII_O3-O4 pose, the ligand is mostly solvent exposed, since the large mobility of the 1-6 branch hampers the existence of stable interactions of the rest of the ligand with the lectin side chains.

Regarding the fluorinated dimannoside **2**, the binding mode ManI_O3-O4 is the most stable and somehow similar to the equivalent in **1**, but since it lacks the Man II unit, the ligand displays more mobility over the lectin surface. In the alternative pose, the less stable ManI_O4-O3 geometry, occasional van der Waals contacts of Man II H-4 and Man I H3 with Val351 are detected (10% of simulated time), and the ligand is more solvent exposed (Figure S5).

Finally, **3** shows a rather variable stability for its different binding modes. Remarkably, the ManII_O3-O4 geometry binds very tightly in the simulations (400 ns of association time on average) and displays HBs involving OH-3 and OH-4 of Man II and Glu358. A CH-π between Phe313 and H-3 Man III is also present for most of the simulated time (Figure S5).

2.3. CORCEMA-ST and Best-Model STD Fitting

As mentioned above, the interpretation of the entire set of experimental STD NMR intensities in terms of ligand/lectin molecular complexes is not straightforward, even considering the MD simulations described above. Thus, a quantitative approach was applied. First, the computed MD trajectories were submitted to a COmplete Relaxation and Conformational Exchange MAtrix Analysis of Saturation Transfer (CORCEMA-ST) analysis [51,52] to estimate the expected STD-NMR intensities. Interestingly, CORCEMA-ST has been previously employed to explore multiple binding events in sugar-lectin complexes. As leading examples, the dual binding mode of Man-containing saccharides to DC-SIGN [46,50] has been shown, as well as different interaction possibilities of Tn-bearing glycopeptides to bind MGL [53].

Then, an in-house computer program dubbed *BM-Mixer* was developed and applied to find the relative proportion of each binding mode that best fit the theoretical to the experimental STD NMR values. This protocol, which might be of general application, allowed the contribution of the different binding modes of each ligand to be dissected when bound to DC-SIGN. In particular, the procedure employs the complete set of calculated STDs from the MD trajectories of each modelled binding mode and search for the combination of poses that better explain the experimental STD data in a fully automated manner (see experimental section for a detailed view of the protocol).

BM-Mixer calculations were performed using ensembles of 400–800 frames per simulated binding mode. Only reliable experimental STDs were used in the calculation, to avoid the introduction of noise. In cases of isochronous NMR frequencies, the observed experimental STD was considered as the sum of the individual contributions from each signal (see 3.5 section in "Material and Methods" part). The fractions of each binding mode of **1**-**3** to DC-SIGN that better fitted the experimental STD data are shown in Figure 4 (see also Table S1). Globally speaking, the predicted contributions for each binding mode agree with the relative stability observed in the MD simulations of each ligand (Table 3).

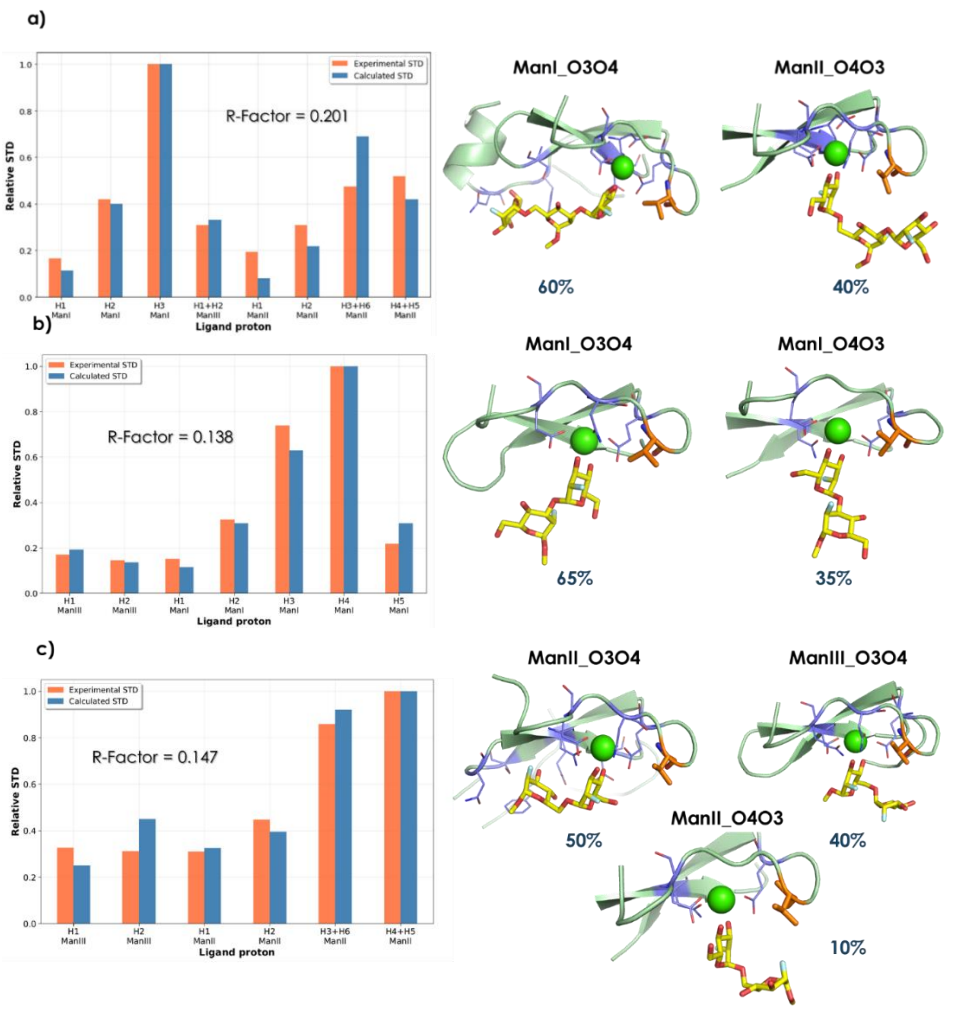


Figure 4. Contributions of the individual binding modes that provide the best fit to the experimental STD NMR data. Representative structures along with the associated NOE R-Factorrel of the best binding mode combination found for compounds 1a, 2b, and 3c are shown. Val351, which is the only aliphatic residue at the lectin binding site, is colored in orange.

According to the computational procedure for the complex of trimannoside **1** with the lectin**,** the best fit was obtained with a combination of 60% ManI_O3-O4, and 40% ManII_O4-O3 (Figure 4a). Thus, both Man I (major) and Man II (minor) directly interact with the calcium at the binding site. For the ManII_O4-O3 geometry, stabilizing intermolecular van der Waals contacts are predicted with Val351. Qualitatively, the presence of this geometry in the distribution allows an explanation for the strong STD observed for Man I H-3. In contrast, the observed STDs at Man III (H-1/H-2) are probably accounted for by its vicinity to Phe313 in the major

ManI_O3-O4 geometry. Remarkably, the most populated binding modes found by *BM-Mixer* are similar to those of the deposited X-ray structures 1K9I and 1SL4 (ManI_O3-O4, 60%) and 2XR5 (ManII_O4-O3, 40%). Fittingly, *BM-Mixer* is able to find the best solution taking from the complete MD ensemble calculated for each binding mode instead of using just a limited number of representative structures.

Interestingly, the predicted major pose, ManI_O3-O4, is the only one in which several potential interactions involving the substituent at C-2 with the lectin side chains could take place. Thus, the corresponding geometry was built for the natural analogue **1b** bound to the lectin, and submitted to MD simulations. The analysis of the trajectory revealed the existence of intermolecular HB interactions involving OH-2 groups, obviously absent in the trifluorinated glycomimetic **1**. For the major ManI_O3-O4 geometry, OH-2 of the terminal Man II acts as a HB donor to Asn362 during 40% of the simulation time. Furthermore, HB contacts between OH-3 of the same Man II with Asn344 and Asn362 are more populated in **1b** than in **1** (70% versus 10%). Additionally, the MD simulations carried out for **1b** predict that OH-2 of the central Man III may be involved in HB interactions with Ser360 (as a donor) and with Glu358 (as an acceptor) around 70% of the simulation time (Figure 5). In fact, the averaged complex stability of trifluorinated **1** in the MD simulations is significantly smaller (lower association time) than that of **1b** (>400 ns in 5 MD replicas). The mobility of the ligand at the binding site is also much higher for **1** than for **1b** (see Figure S6 for a detailed summary of the interactions and Figure S7 for a summary of the computed torsion angles and root-mean-square Deviation (RMSD) plots).

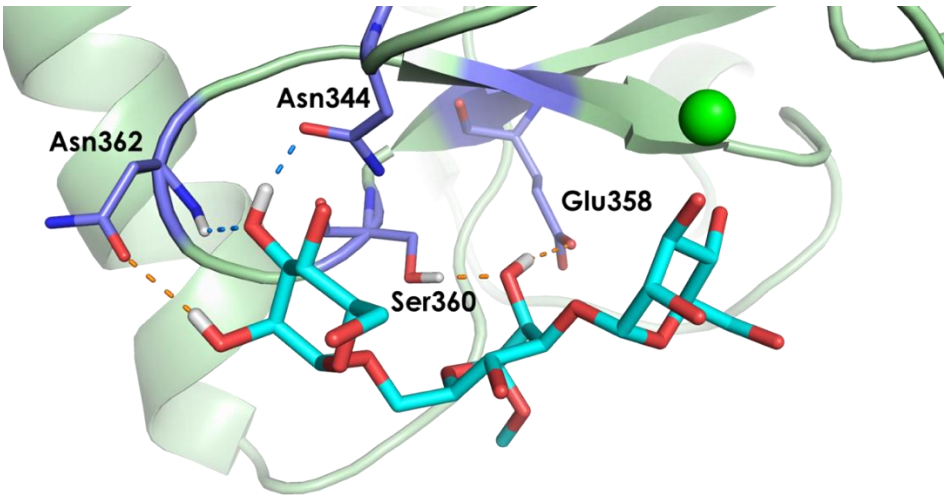


Figure 5. Possible intermolecular hydrogen bond interactions predicted by the MD simulations for the complex between DC-SIGN and the core trimannoside 1b. These hydrogen bonds (HBs) are absent in the corresponding complex with the trifluorinated glycomimetic 1. Hydrogen bonds involving OH-2 groups of Man II and Man III are colored orange, and the ones involving OH-3 are in blue. The most stable binding pose, ManI_O3-O4, was used as starting geometry.

For the complex of DC-SIGN with **2**, with O3 of Man III involved in the 1-3 linkage, only two poses are possible. The best fit was found for a population distribution of 65% ManI_O3-O4 and 35% of the alternative ManI_O4-O3 pose. The obtained major geometry is in agreement with that described above for the trisaccharide **1**, while the intermolecular contacts are also analogous, with the obvious exception of those that involve Man II. Indeed, the experimental STD profile for Man I is analogous to that observed for **1**, with the added contribution of H-4 arising from the alternative ManI_O4-O3 pose.

In the complex of compound **3**, Man III can also participate in the coordination of the Ca^{2+}. In fact, the best fit is provided by a large contribution of the ManII_O3-O4 (50%) and ManIII_O3-O4 (40%) poses, with a small participation of the ManII_O4-O3 geometry (10%). The latter was predicted to a larger extent for the complex of trisaccharide **1**. However, in that case, Man I provided contacts with Val351 further stabilizing the pose. Interestingly, the major ManII_O3-O4 geometry for **3** cannot take place in **1**, given the presence of Man I, which would collapse with the protein surface.

Moreover, the analysis of the geometries permits an explanation of the observed relative affinities. As described in Table 2, the trifluorinated ligand **1** displays about half the affinity of its natural counterpart **1b**, and its monofluorinated analogues **1c** and **1d** versus DC-SIGN. Therefore, it is

likely that the predicted hydrogen bonds involving OH-2 and OH-3 at Man II are at the heart of the observed loss of binding energy. In any case, the combined effect caused by the simultaneous fluorination of the three Man units is rather low in terms of binding energy (2.5 fold variation in binding affinity amounts to ca. 1 kcal/mol in binding free energy). The lack of Man II in **2** and of Man I in **3** precludes a number of stabilizing HBs and van der Waals contacts that lead to the observed loss in affinities in the fluorinated disaccharides versus the trisaccharide **1** (ca. four-fold **1** versus **2** and ca. two-fold **1** versus **3**).

3. MATERIALS AND METHODS

3.1. Man-Based Ligands

The synthesis of compounds **1**, **1b, 2**, and **3** has already been described [44]. The synthesis of **1c** and **1d** is detailed in SI and Scheme S1. $^{1}H/^{19}F$-NMR assignments for **1, 2,** and **3** are compiled in Table S2, and where previously described in [44].

3.2. DC-SIGN ECD Preparation

Plasmids pET15b holding the DC-SIGN full extracellular domain (ECD, residues 70-404, Thermo Fischer Scientific) were amplified in E. coli DH5α and subsequently transformed on BL21/DE3 competent cells (Sigma-Aldrich). A single colony was added to a preinocule of 100 mL, in the presence of 100 mg/L ampicillin, and grown overnight at 37 °C under continuous shaking. Then, each preinocule was diluted up to 1 L of Luria-Bertani (LB) broth containing 100 mg/L ampicillin, and cell growth was maintained at 37 °C with gentle shaking until OD = 0.60–0.65. Cultures were finally induced with 1 mM isopropyl-1-thio-b-D-galactopyranoside (IPTG) and allowed to grow overnight at 20 °C. The resulting pellets were harvested by centrifugation (4 °C, 4500 rpm, 20 min) and resuspended in the minimal amount of Tris-HCl buffer 10 mM (pH = 8.0).Cells were lysed by sonication and the unfolded ECD was separated in the insoluble fraction by ultracentrifugation (4 °C, 30 k rpm, 60 min). The isolated inclusion bodies were further solubilized in 6 M urea and 2 mM β-mercaptoethanol (0.01% v/v). By ultracentrifugation (4 °C, 40 k rpm, 180 min), the remaining insoluble cell debris were sedimented and the supernatant containing the unfolded ECD was carefully decanted. Protein refolding was performed by stepwise dialyzing the supernatant against 2 L of 100 mM Tris-HCl buffer (NaCl 150 mM, $CaCl_2$ 10 mM, pH = 8.0, 24 h each step), containing 4 M urea, 2 M urea, and no urea, respectively. The ECD was first purified by

affinity chromatography in a Mannose–Sheparose column, using as loading buffer Tris-HCl 20 mM, NaCl 150 mM, $CaCl_2$ 10 mM, pH = 8.0; and as elution buffer, 20 mM Tris-HCl, 150 mM NaCl, 10 mM EDTA, pH = 8.0. Eluted fractions containing significant amounts of the ECD ($A(\lambda_{280}) > 0.10$) were mixed and repurified by size exclusion chromatography (AKTA sys., GE Healthcare) in a HiLoad 26/600 Superdex 200 column (UV_{max} at V_{elut} = 140 mL, flow rate 2.5 mL/min) using as elution buffer Tris-HCl 20 mM, NaCl 150 mM, EDTA 1 mM, pH = 8.0. Conditions of the collected pure ECD fractions were changed to Tris-d_{11} 20 mM, $CaCl_2$ 4 mM, NaCl 150 mM, pH = 8.0 (D_2O), using Vivaspin membrane filters of 100k MWCO. The tetrameric state of the lectin was previously confirmed by TEM [48]. The presence of the protein was monitored throughout the entire protocol by 4–12% SDS-PAGE, and ECD concentrations were determined by UV-Vis spectroscopy ($\varepsilon_{280,tetramer}$ = 280,600 M^{-1} cm^{-1}, estimated from ProtParam).

3.3. NMR Experiments

All spectra were acquired at 298 K on either a Bruker AV500 or AV600 NMR spectrometers, equipped with a ^{19}F probe (5 mm SEF ^{19}F-^{1}H with Z gradient). All the samples were prepared using the same protein buffer (20 mM Tris-d_{11}, 4 mM $CaCl_2$, 150 mM NaCl, pH = 8.0, in D_2O) was employed. For molecular recognition experiments, controls were performed in the presence of either EDTA to sequester the Ca^{2+} or a known DC-SIGN ligand, to assess specific binding at the primary binding site.

Samples for the 2D STD-TOCSY-reF contained 9.14 μM ECD (tetramer) and 153 eq. of ligand (1.4 mM, the ratio ligand/protein (CRD) is 38:1) in each case. The 2D STD-TOCSY-reF setup consists on a proton saturation transfer scheme ($^{1}H_{protein} \rightarrow {}^{1}H_{ligand}$-STD) with subsequent evolution through J_{HH} couplings (TOCSY), to finally refocus the STD-edited information in the ^{19}F dimension (reF). As STD parameters, selective saturation was applied for 2 s using 90º PC9 pulses of 10 ms length, with on-resonance saturation set at 0.85 ppm and off-resonance at 60 ppm. The TOCSY mixing time used was 128 ms, and the corresponding J_{HF} evolution and refocusing delays in the reF module were Δ^H = 10.4 ms and Δ^F = 6 ms, respectively. The relative saturation values (%) were calculated from the ratio between peaks in the 2D STD-TOCSY-reF spectrum and the corresponding ones in the off-resonance TOCSY-reF spectrum. The values were normalized by assigning a 100% value to the peak that displayed the highest STD ratio.

^{19}F-NMR R_2/T_2 experiments for affinity measurements were carried out employing a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence as follows: $[D - 90_x - (\tau - 180_y - \tau)_n - acquire]$, with a recovery delay $D = 4\ s$, and a list of 18 values for the spin-echo loop n to yield T_2 by fitting to $I(t) =$

$I_0 e^{-n2\tau/T_2}$ [48]. For K_D estimation of 6-deoxy-6-F-Man (the spy-molecule), the free evolution delay was set to an optimized value $\tau = 1\ ms$, after using relaxation dispersion experiments to assure a negligible chemical exchange contribution to $R_{2,obs}$ (Figure S1, a)). The K_I of **1**, **1b-d**, **2**, and **3** with DC-SIGN (ECD) was determined by titration experiments in a competitive fashion [54,55]. Increasing amounts of the competitor molecules (compounds **1**, **1b-d**, **2**, and **3**) were added to a solution mixture of the spy molecule and the protein in each case, calculating the $R_{2,obs}$ of the spy molecule at each competitor concentration while the ratio [Spy-molecule]/[Protein] is fixed (40:1) (Figure S1, c)).

3.4. Molecular Dynamics Simulations

All compounds were initially built using Glycam-web tool [56]. The fluorinated glycomimetics **1**, **2,** and **3** were generated by modifying the Glycam structures with Discovery Studio or PyMol according to the particular molecule. Then, the fluorinated structures were submitted to energy minimization, charge and atom type assignment using Antechamber [57].

The deposited protein structure on PDB 2IT5 [30] was used as the lectin geometry for all MD simulations, after removing all but one CRD with its corresponding structural Ca^{2+} ions. The two models of the crystalized Man ligands coordinating the Ca^{2+} at the primary binding site served as templates for superimposing the starting geometries for the O3-O4 (most populated ligand orientation on 2IT5) and O4-O3 (less populated one) 3D models. Then, the crystalized ligands on the original 2IT5 structure were removed as well.

Overall, 10 models of DC-SIGN in complex with the ligands **1** (4 binding modes), **2** (2 binding modes), and **3** (4 binding modes) were built according to the criteria specified in the MD simulations section on *Results and Discussion* (also see SI). Additionally, a model of DC-SIGN with the natural analogue **1b** in the binding pose ManI_O3-O4 (see *Results and Discussion*) was also prepared.

All MD simulations were conducted using AMBER molecular dynamics simulations software (version 16). The employed preparation, minimization and production protocol was the same for all the simulations. First, solutes (either ligand or protein-ligand complexes) were solvated using a pre-equilibrated TIP3P rectangular water box, spanning 10 Å from the solute in each direction. Cl^- ions were added to neutralize the system, using Li/Merz parameters [58]. The AMBER's ff14SB force field [59] was used to parametrize the protein, while either GAFF2 [60] (for analogues **1**, **1b–d**, **2,** and **3**) or GLYCAM06 [56] (for **1b**) parameters were employed for the

ligands. The system was minimized in two stages, starting with the solvent and counterions while keeping the solute fixed, and following with a general minimization of the whole system. Then, gradual heating from 0 to 300 K in the NVT ensemble for 100 ps was performed. Further equilibration was run at 300 K in NPT during 1 ns, employing a Langevin thermostat with collision frequency 1 ps. The production dynamics were run in the NPT ensemble at 300 K, 1 bar, using the Langevin thermostat as in the previous stage. SHAKE was employed to constrain bond lengths on hydrogen atoms during production dynamics, and the time step was set to 2 ps. The cutoff for non-bonded interactions was set to 10 Å, and the Particle Mesh Ewald Method was used to introduce long-range electrostatic effects. The simulation time of the complexes in water was set to 400 ns, and 6 to 12 replicas were run in each case.

Additional MD simulations of 500 ns of the ligands **1** and **1b** in explicit water were also conducted, to assess the average conformational dispositions of the free molecules in solution (Table 2). The MD protocol employed was the same as beforementioned.

3.5. CORCEMA-ST and Best-Model STD Fitting

CORCEMA-ST allows prediction of the theoretical STD-NMR intensities of the ligand protons in a given receptor-ligand complex [51,61,62]. The STD profile depends on the geometry of the complex, rotational motion correlation times of the system, irradiation conditions, and dissociation constant, among others.

The following protocol was employed. For every ligand, a subset of 400–800 frames extracted from the MD trajectories computed for each binding mode (BM) was randomly selected. Only the BMs whose averaged association time in the MD simulation was larger than 5 ns were analyzed. The predicted STDs for each individual frame according to CORCEMA-ST were compiled and submitted to a Python based program, dubbed *BM-Mixer*, which aims to find the minimum ensemble of conformations of the bound ligand that best fits the experimental STD data, in an iterative manner. This fitting is guided by the minimization of the relative NOE R-factor (agreement factor), which is defined as follows for the ensemble of binding mode geometries [52]:

$$NOE\ R-factor_{rel} = \sqrt{\frac{\sum_{i=1}^{k}(STD_{exp,rel,i} - \overline{STD}_{teor,rel,i})^2}{\sum_{i=1}^{k}(STD_{exp,rel,i})^2}} \quad (1)$$

where $STD_{exp,rel,i}$ is the experimental STD value of proton i normalized to the most intense experimental peak, and $\overline{STD}_{teor,rel,i}$ is the calculated averaged STD value of proton i, normalized to the most intense calculated peak for the particular combination of binding mode frames. It was observed that, in general, *BM-Mixer* calculations converged when a minimum number of about 400-800 frames are used in the calculations, depending on the ligand mobility at the binding site.

The results of the search for the best combination of binding modes is now exemplified for **2.** Conveniently, only two binding modes are possible for this ligand. Figure 6 shows the evolution of the 20 iterations required to screen the space of possible BM combinations, expressed as the percentage of each BM in the *x*-axis. As it can be observed, the agreement between the calculated and the experimental STD data is rather poor when either of the two ensembles of structures representing each binding mode (100% ManI_O4O3 or 100% ManI_O3O4) are separately considered. However, the NOE R-factor$_{rel}$ converges to a minimum when 65% of the ManI_O3O4 binding pose is considered.

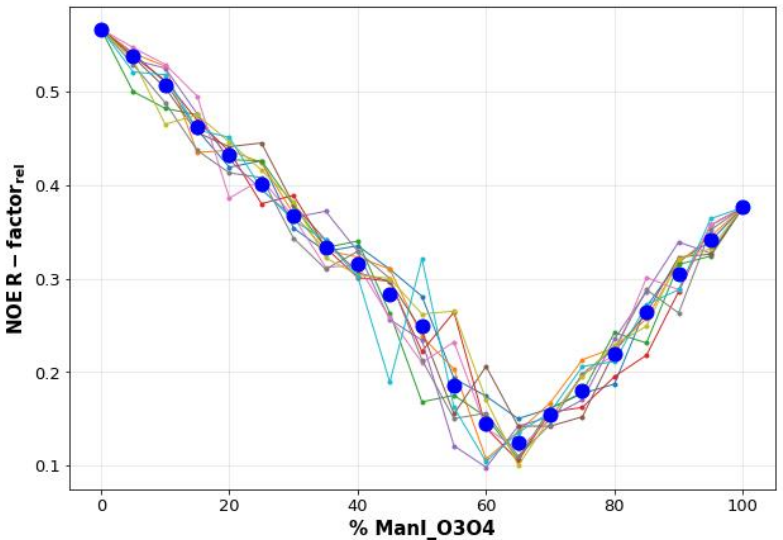


Figure 6. The fit of the NOE R-factor$_{rel}$ value computed by *BM_Mixer* to the relative population of ManI_O3O4 frames. Obviously, the percentage of the alternative binding mode ManI_O4O3 is $100 - \%ManI\_O3O4$. The quantitative STD data predicted by CORCEMA-ST for 800 frames from the MD simulations were employed. Each colored line represents a different iteration of the main loop of the program (only 10 out of 20 are shown for clarity). Blue circles represent the mean value among all iterations for that particular combination of binding modes. The minimum NOE R-factor$_{rel}$ is reached for an ensemble of structures formed by 65% of ManI_O3O4 and 35% of ManI_O4O3. All the assigned experimental STD intensities were used in the calculations. In this case, 400 frames from the ManI_O4O3 trajectory and 400 frames from the ManI_O3O4 trajectory were employed.

The list of experimental STDs provided to *BM-Mixer* is crucial to obtain reliable results. It is essential to use a large enough number of experimental STD peaks to minimize possible ambiguities. It was observed that different BM combinations may provide a similar NOE R-factor$_{rel}$. Since errors for very weak experimental STD signals (close to zero STD) may be rather high, it is advisable to discard these signals. Thus, the list of experimental STDs used for the calculations should only contain the reliable assigned STD peaks. Notably, the very weak peaks can be introduced in the computation a posteriori as internal controls for detecting false positives.

Obviously, it is convenient to examine whether each individual MD ensemble provides STD profiles that are different enough. Otherwise, *BM-Mixer* will provide an ambiguous fractional contribution of the possible binding poses. In these cases, additional experimental information may be required to differentiate between them.

For those peaks that are virtually isochronous in the NMR spectrum, the experimental STD intensity observed can be considered as the sum of the individual contributions of each individual proton (see Figure 5). *BM-Mixer*

handles this 'aggregated' experimental STD file and automatically performs the corresponding sum of contributions of the CORCEMA-ST predicted STD values for the overlapped protons. In this manner, both calculated and experimental STD aggregate can be evaluated using the agreement factor. It was assumed that the shapes of the NMR signals are similar, and the protons are isochronous. Nevertheless, if the overlapped protons display rather different T_2 relaxation times or their resonance frequencies are not exactly the same, considerable errors could be introduced in the calculations. For instance, in the case of compound **3**, the experimental STD cross peak observed for Man III at 3.7 ppm could correspond to H-3, H-4, H-5 or H-6'. Since considering the measured STD intensity as the sum of the individual contribution from four H atoms would introduce noise in the search, the corresponding STD peak was not considered in the best-model fitting shown in Figure 5.

Since CORCEMA-ST is designed to calculate the absolute intensity of 1D ^{1}H-STD-NMR spectra, some considerations were taken to assess its application to the analysis of the 2D-STD-TOCSYreF experiments. In particular, the experiment employed herein includes a homonuclear ^{1}H-TOCSY block and a heteronuclear ^{1}H,^{19}F reF transfer. Thus, the absolute STD intensities are not directly comparable to those of the regular 1D-STD-NMR experiment. However, the relative STD profiles (i.e, the STD intensities normalized to the most intense STD peak) are expected to remain similar. Regarding the employment of the ^{19}F nucleus, since the ^{1}H→^{19}F heteronuclear cross-relaxation is largely less efficient than the ^{1}H→^{1}H homonuclear one, it is not expected that its effect is noticeable. In any case, as control, CORCEMA-ST calculations were also performed using molecular models the ^{19}F atoms were replaced by ^{1}H. Fittingly, these CORCEMA-ST controls demonstrated that the predicted contribution of each binding mode that provided the best NOE-R-factor$_{rel}$ was rather insensitive to the presence of the F atom (Table S3).

4. CONCLUSIONS

A new protocol for unraveling multiple binding modes is presented and employed to study the interaction of the trifluorinated mimetic of the trimannoside core with DC SIGN. Although applied to fluorine containing sugars, it might be used for all types of molecules. The protocol is based on the synergic combination of the novel 2D STD-TOCSY-reF experiment to extract reliable STD values, with extensive MD simulations and CORCEMA calculations. Finally, a new computer program, *BM-Mixer*, has been able to find the combination of geometries from the complete MD ensembles of the

ligand/protein complexes that provided the best match to the experimental data. The results indicate that no single solution can account for the experimental results, and strongly suggest the existence of more than one binding mode.

5. REFERENCES

1. Reily, C.; Stewart, T.J.; Renfrow, M.B.; Novak, J. Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019**, *15*, 346–366.
2. André, S.; Kaltner, H.; Manning, J.C.; Murphy, P.V.; Gabius, H.J. Lectins: Getting familiar with translators of the sugar code. *Molecules* **2015**, *20*, 1788–1823.
3. Rodríguez, E.; Schetters, S.T.T.; Van Kooyk, Y. The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nat. Rev. Immunol.* **2018**, *18*, 204–211.
4. van Kooyk, Y.; Rabinovich, G.A. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat. Immunol.* **2008**, *9*, 593–601.
5. Kleene, R.; Schachner, M. Glycans and neural cell interactions. *Nat. Rev. Neurosci.* **2004**, *5*, 195–208.
6. Rahfeld, P.; Sim, L.; Moon, H.; Constantinescu, I.; Morgan-Lang, C.; Hallam, S.J.; Kizhakkedathu, J.N.; Withers, S.G. An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **2019**, *4*, 1475–1485.
7. Naresh, K.; Schumacher, F.; Hahm, H.S.; Seeberger, P.H. Pushing the limits of automated glycan assembly: Synthesis of a 50mer polymannoside. *Chem. Commun. (Camb).* **2017**, *53*, 9085–9088.
8. Schönemann, W.; Cramer, J.; Mühlethaler, T.; Fiege, B.; Silbermann, M.; Rabbani, S.; Dätwyler, P.; Zihlmann, P.; Jakob, R.P.; Sager, C.P.; et al. Improvement of Aglycone π-Stacking Yields Nanomolar to Sub-nanomolar FimH Antagonists. *ChemMedChem* **2019**, *14*, 749–757.
9. Fernández de Toro, B.; Peng, W.; Thompson, A.J.; Domínguez, G.; Cañada, F.J.; Pérez-Castells, J.; Paulson, J.C.; Jiménez-Barbero, J.; Canales, A. Avenues to Characterize the Interactions of Extended N-Glycans with Proteins by NMR Spectroscopy: The Influenza

Hemagglutinin Case. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 15051–15055.

10. Parsons, T.B.; Struwe, W.B.; Gault, J.; Yamamoto, K.; Taylor, T.A.; Raj, R.; Wals, K.; Mohammed, S.; Robinson, C.V.; Benesch, J.L.P.; et al. Optimal Synthetic Glycosylation of a Therapeutic Antibody. *Angew. Chem., Int. Ed.* **2016**, *55*, 2361–2367.
11. Klukowski, P.; Schubert, M. Chemical shift-based identification of monosaccharide spin-systems with NMR spectroscopy to complement untargeted glycomics. *Bioinformatics* **2019**, *35*, 293–300.
12. Tamburrini, A.; Colombo, C.; Bernardi, A. Design and synthesis of glycomimetics: Recent advances. *Med. Res. Rev.* **2020**, *40*, 495–531.
13. Franconetti, A.; López, O.; Fernandez-Bolanos, J.G. Carbohydrates: Potential Sweet Tools against Cancer. *Curr. Med. Chem.* **2020**, *27,* 1206–1242.
14. Egger, J.; Weckerle, C.; Cutting, B.; Schwardt, O.; Rabbani, S.; Lemme, K.; Ernst, B. Nanomolar E-Selectin Antagonists with Prolonged Half-Lives by a Fragment-Based Approach. *J. Am. Chem. Soc.* **2013**, *135*, 9820–9828.
15. Peng, W.; Paulson, J.C. CD22 Ligands on a Natural N-Glycan Scaffold Efficiently Deliver Toxins to B-Lymphoma Cells. *J. Am. Chem. Soc.* **2017**, *139*, 12450–12458.
16. Peterson, K.; Kumar, R.; Stenström, O.; Verma, P.; Verma, P.R.; Hakansson, M.; Kahl-Knutsson, B.; Zetterberg, F.; Leffler, H.; Akke, M.; et al. Systematic Tuning of Fluoro-galectin-3 Interactions Provides Thiodigalactoside Derivatives With Single-Digit nM Affinity and High Selectivity. *J. Med. Chem.* **2018**, *61*, 1164–1175.
17. Hevey, R. Strategies for the development of glycomimetic drug candidates. *Pharmaceuticals* **2019**, *12*, 55.
18. Rillahan, C.D.; Antonopoulos, A.; Lefort, C.T.; Sonon, R.; Azadi, P.; Ley, K.; Dell, A.; Haslam, S.M.; Paulson, J.C. Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat. Chem. Biol.* **2012**, *8*, 661–668.
19. Mei, H.; Han, J.; Fustero, S.; Medio-Simon, M.; Sedgwick, D.M.; Santi, C.; Ruzziconi, R.; Soloshonok, V.A. Fluorine-containing drugs approved by the FDA in 2018. *Chemistry* **2019**, *25*, 11797–11819.

20. Mei, H.; Remete, A.M.; Zou, Y.; Moriwaki, H.; Fustero, S.; Kiss, L.; Soloshonok, V.A.; Han, J. Fluorine-containing drugs approved by the FDA in 2019. *Chin. Chem. Lett.* **2020**, doi:10.1016/j.cclet.2020.03.050.

21. Purser, S.; Moore, P.R.; Swallow, S.; Gouverneur, V. Fluorine in Medicinal Chemistry. *Chem. Soc. Rev.* **2008**, *37*, 320–330.

22. Linclau, B.; Ardá, A.; Reichardt, N.C.; Sollogoub, M.; Unione, L.; Vincent, S.P.; Jiménez-Barbero, J. Fluorinated carbohydrates as chemical probes for molecular recognition studies. Current status and perspectives. *Chem. Soc. Rev.* **2020**, doi:10.1039/c9cs00099b.

23. Schulze, J.; Baukmann, H.; Wawrzinek, R.; Fuchsberger, F.F.; Specker, E.; Aretz, J.; Nazaré, M.; Rademacher, C. CellFy: A Cell-Based Fragment Screen against C-Type Lectins *ACS Chem. Biol.* **2018**, *13*, 3229–3235.

24. Wamhoff, E.C.; Hanske, J.; Schnirch, L.; Aretz, J.; Grube, M.; Silva, D.V.; Rademacher, C. 19F NMR-Guided Design of Glycomimetic Langerin Ligands. *ACS Chem. Biol.* **2016,** *11,* 2407–2413.

25. Dalvit, C.; Vulpetti, A. Ligand-Based Fluorine NMR Screening: Principles and Applications in Drug Discovery Projects. *J. Med. Chem.* **2019**, *62*, 2218–2244.

26. Valverde, P.; Delgado, S.; Martínez, J.D.; Vendeville, J.-B.; Malassis, J.; Linclau, B.; Reichardt, N.-C.; Cañada, F.J.; Jiménez-Barbero, J.; Ardá, A. Molecular insights into DC-SIGN binding to self-antigens: The interaction with the blood group A/B antigens. *ACS Chem. Biol.* **2019**, *14*, 1660–1671.

27. Figdor, C.G.; van Kooyk, Y.; Adema, G.J. C-type Lectin Receptors on Dendritic Cells and Langerhans Cells. *Nat. Rev. Immunol.* **2002**, *2*, 77–84.

28. van Vliet, S.J.; García-Vallejo, J.J.; van Kooyk, Y. Dendritic cells and C-type lectin receptors: Coupling innate to adaptive immune responses. *Cell Biol.* **2008**, *86*, 580–587.

29. Valverde, P., Martinez, J.D.; Cañada, F.J.; Arda, A.; Jimenez-Barbero, J. Molecular recognition in C-type lectins: The cases of DC-SIGN, Langerin, MGL, and L-sectin. *ChemBioChem* **2020**, doi:10.1002/cbic.202000238.

30. Nagae, M.; Yamaguchi, Y. Structural Aspects of Carbohydrate Recognition Mechanisms of C-Type Lectins. *Curr. Top. Microbiol. Immunol.* **2019**, 1–30, doi:10.1007/*82_2019_181*.

31. Cummings, R.D.; McEver, R.P. *Essentials of Glycobiology*, 3rd ed.; Varki, A., Cummings, R.D., Esko, J.D., Stanley, P., Hart, G.W., Aebi, M., Darvill, A.G., Kinoshita, T., Packer, N.H., Prestegard, J.H., et al., Eds.; Cold Spring Harbor (NY), USA, 2015-17; Chapter 34.

32. Geijtenbeek, T.B.; Gringhuis, S.I. C-type lectin receptors in the control of T helper cell differentiation. *Nat. Rev. Immunol.* **2016**, *16*, 433–448.

33. Mason, C.P.; Tarr, A.W. Human Lectins and Their Roles in Viral Infections. *Molecules* **2015**, *20*, 2229–2271.

34. Kwon, D.S.; Gregorio, G.; Bitton, N.; Hendrickson, W.A.; Littman, D.R. DC-SIGN-Mediated Internalization of HIV Is Required for Trans-Enhancement of T Cell Infection. *Immunity* **2002**, *16*, 135–144.

35. Hodges, A.; Sharrocks, K.; Edelmann, M.; Baban, D.; Moris, A.; Schwartz, O.; Drakesmith, H.; Davies, K.; Kessler, B.; McMichael, A.; et al. Activation of the lectin DC-SIGN induces an immature dendritic cell phenotype triggering Rho-GTPase activity required for HIV-1 replication. *Nat. Immunol.* **2007**, *8*, 569–577.

36. Sarkar, R.; Mitra, D.; Chakrabarti, S. HIV-1 gp120 protein downregulates Nef induced IL-6 release in immature dentritic cells through interplay of DC-SIGN. *PLoS ONE* **2013**, *8*, e59073.

37. Simmons, G.; Reeves, J.D.; Grogan, C.C.; Vandenberghe, L.H.; Baribaud, F.; Whitbeck, J.C.; Burke, E.; Buchmeier, M.J.; Soilleux, E.J.; Riley, J.L.; et al. DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* **2003**, *305,* 115–123.

38. Tassaneetrithep, B.; Burgess, T.H.; Granelli-Piperno, A.; Trumpfheller, C.; Finke, J.; Sun, W.; Eller, M.A.; Pattanapanyasat, K.; Sarasombath, S.; Birx, D.L.; et al. DC-SIGN (CD209) mediates dengue virus infection of human dendritic cells. *J. Exp. Med.* **2003**, *197*, 823–829.

39. Alen, M.M.; Dallmeier, K.; Balzarini, J.; Neyts, J.; Schols, D. Crucial role of the N-glycans on the viral E-envelope glycoprotein in DC-

SIGN-mediated dengue virus infection. *Antiviral Res.* **2012**, *96*, 280–287.

40. Liu, P.; Ridilla, M.; Patel, P.; Betts, L.; Gallichotte, E.; Shahidi, L.; Thompson, N.L.; Jacobson, K. Beyond attachment: Roles of DC-SIGN in dengue virus infection. *Traffic* **2017**, *18*, 218–231.

41. van Liempt, E.; Bank, C.M.C.; Mehta, P.; García-Vallejo, J.J.; Kawar, Z.S.; Geyer, R.; Alvarez, R.A.; Cummings, R.D.; van Kooyk, Y.; van Die, I. Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS Lett.* **2006**, *580*, 6123−6131.

42. Mitchell, D.A.; Fadden, A.J.; Drickamer, K. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR subunit organization and binding to multivalent ligands. *J. Biol. Chem.* **2001**, *276*, 28939–28945.

43. Guo, Y.; Feinberg, H.; Conroy, E.; Mitchell, D.A.; Alvarez, R.; Blixt, O.; Taylor, M.E.; Weis, W.I.; Drickamer, K. Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* **2004**, *11*, 591−598.

44. Diercks, T.; Infantino, A.S.; Unione, L.; Jiménez-Barbero, J.; Oscarson, S.; Gabius. H.-J. Fluorinated Carbohydrates as Lectin Ligands: Synthesis of OH/F-Substituted N-Glycan Core Trimannoside and Epitope Mapping by 2D STD-TOCSYreF NMR spectroscopy. *Chem. Eur. J.* **2018**, *24*, 15761–15765.

45. Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P.H.; Weis, W.I. Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.* **2007**, *282*, 4202–4209.

46. Angulo, J.; Diaz, I.; Reina, J.J.; Tabarani, G.; Fieschi, F.; Rojo, J.; Nieto, P.M. Saturation Transfer Difference (STD) NMR Spectroscopy Characterization of Dual Binding Mode of a Mannose Disaccharide to DC-SIGN. *ChemBioChem* **2008**, *9*, 2225–2227.

47. Feinberg, H.; Mitchell, D.A.; Drickamer, K.; Weis, W.I. Structural Basis for Selective Recognition of Oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001**, *294*, 2163−2166.

48. Martínez, J.D.; Valverde, P.; Delgado, S.; Romanò, C.; Linclau, B.; Reichardt, N.C.; Oscarson, S.; Ardá, A.; Jiménez-Barbero, J.;

Cañada, F.J. Unraveling Sugar Binding Modes to DC-SIGN by Employing Fluorinated Carbohydrates. *Molecules* **2019**, *24*, 2337.

49. Thépaut, M.; Guzzi, C.; Sutkevičiūtė, I.; Sattin, S.; Ribeiro-Viana, R.; Varga, N.; Chabrol, E.; Rojo, J.; Bernardi, A.; Angulo, J.; et al. Structure of a glycomimetic ligand in the carbohydrate recognition domain of C-type lectin DC-SIGN. Structural requirements for selectivity and ligand design. *J. Am. Chem. Soc.* **2013**, *135*, 2518−2529.

50. Guzzi, C.; Alfarano, P.; Sutkevičiūtė, I.; Sattin, S.; Ribeiro-Viana, R.; Fieschi, F.; Bernardi, A.; Weiser, J.; Rojo, J.; Angulo, J.; et al. Detection and quantitative analysis of two independent binding modes of a small ligand responsible for DC-SIGN clustering. *Org. Biomol. Chem.* **2016**, *14*, 335–344.

51. Jayalakshmi, V.; Krishna. N.R. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* **2002**, *155*, 106–118.

52. Jayalakshmi, V.; Krishna. N.R. CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. *J. Magn. Reson.* **2004**, *168*, 36–45.

53. Marcelo, F.; Garcia-Martin, F.; Matsushita, T.; Sardinha, J.; Coelho, H.; Oude-Vrielink, A.; Koller, C.; André, S.; Cabrita, E.J.; Gabius, H.J.; et al. Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin. *Chemistry* **2014**, *20*, 16147–16155.

54. Dalvit, C. Theoretical analysis of the competition ligand-based NMR experiments and selected applications to fragment screening and binding constant measurements. *Concepts Magn. Reson.* **2008**, *32A*, 341–372.

55. Doerr, A.J.; Case, M.A.; Pelczer, I.; McLendon, G.L. Design of a functional protein for molecular recognition: Specificity of ligand binding in a metal-assembled protein cavity probed by 19F NMR. *J. Am. Chem. Soc.* **2004**, *126,* 4192–4198.

56. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06:

A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655.

57. Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *J. Comput. Chem.* **2005**, *25*, 1157–1174.

58. Li, P.; Song, L.F.; Merz, K.M., Jr. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**, *11*, 1645–1657.

59. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99S. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

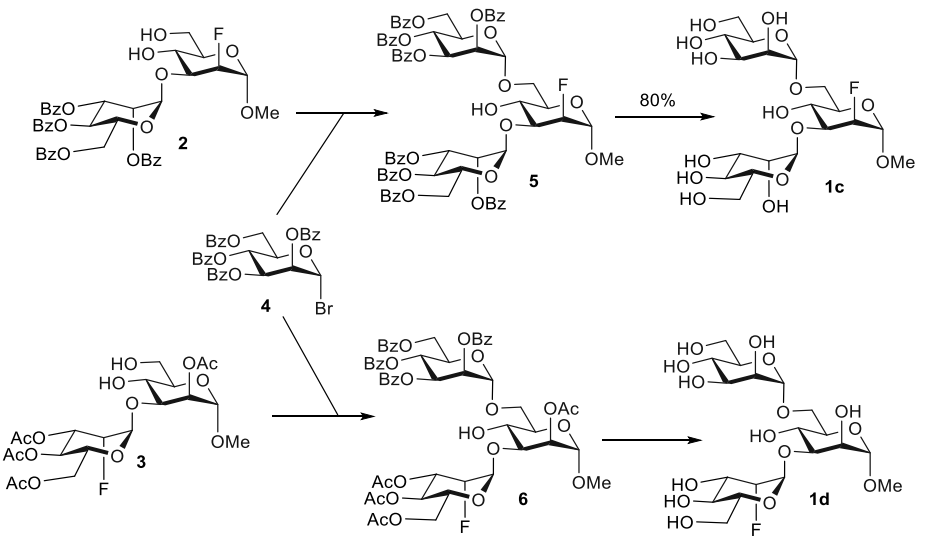
60. Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *Inc. J. Comput. Chem.* **2004**, *25*, 1157–1174.

61. Jayalakshmi, V.; Krishna. N.R. Determination of the conformation of trimethoprim in the binding pocket of bovine dihydrofolate reductase from a STD-NMR intensity-restrained CORCEMA-ST optimization. *J. Am. Chem. Soc.* **2005**, *127*, 14080–14084.

62. Gimeno, A.; Reichardt, N.C.; Cañada, F.J.; Perkams, L.; Unverzagt, C.; Jiménez-Barbero, J.; Ardá, A. NMR and molecular recognition of N-Glycans: Remote modifications of the saccharide chain modulate binding features. *ACS Chem. Biol.* **2017**, *12*, 1104–1112.

6. SUPPORTING INFORMATION

Synthesis of 1c and 1d



Scheme S1. Synthesis of compounds **1c** and **1d**.

Methyl 2,3,4,6-tetra-O-benzoyl-α-D-mannopyranosyl-(1→3)-[2,3,4,6-tetra-O-benzoyl-α-D-mannopyranosyl-(1→6)-]-2-deoxy-2-fluoro-α-D-mannopyranoside (5)

4,6-diol acceptor **2** and donor **4** were coupled to give trisaccharide **5** (55%). R_f 0.38 (toluene/EtOAc, 7:1). $[\alpha]_D^{20}$ -32.1 (*c* 1.0; $CHCl_3$). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.16 – 8.00 (m, 10H; H_{Ar}), 7.99 – 7.94 (m, 2H; H_{Ar}), 7.84 (m, 4H; H_{Ar}), 7.58 (m, 4H; H_{Ar}), 7.53 – 7.33 (m, 16H; H_{Ar}), 7.31 – 7.22 (m, 4H; H_{Ar}), 6.15 – 6.05 (m, 2H; H-4', H-4''), 6.01 – 5.93 (m, 2H; H-3', H-3''), 5.85 (dd, $J_{2',3'}$ = 3.3 Hz, $J_{2',1'}$ = 1.8 Hz, 1H; H-2'), 5.79 (dd, $J_{2'',3''}$ = 3.4 Hz, $J_{2'',1''}$ = 1.8 Hz, 1H; H-2''), 5.38 (d, $J_{1',2'}$ = 1.8 Hz, 1H; H-1'), 5.23 (d, $J_{1'',2''}$ = 1.8 Hz, 1H; H-1''), 5.05 – 4.89 (m, 1H; H-2), 4.81 – 4.66 (m, 4H; H-1, H-5', H-6'a, H-6''a), 4.63 (ddd, $J_{5'',4''}$ = 10.1 Hz, $J_{5'',6''a}$ = 4.5 Hz, $J_{5'',6''b}$ = 2.5 Hz, 1H; H-5''), 4.53 (m, 2H; H-6'b, H-6''b), 4.12 (m, 2H; H-4, H-6a), 4.04 – 3.85 (m, 3H; H-6b, H-3, H-5), 3.45 – 3.36 (m, 3H; OCH_3), 3.32 ppm (d, $J_{OH,4}$ = 3.5 Hz, 1H; O*H*-4). ^{13}C NMR (126 MHz, $CDCl_3$) δ 166.24, 166.05, 165.62, 165.58, 165.51, 165.44, 165.38, 165.29 (8 *C*OPh), 133.52, 133.46, 133.38, 133.33, 133.24, 133.07, 133.01, 129.96-128.25 (48 C_{Ar}), 99.95 (C-1'), 98.28 (d, *J* = 29.1 Hz; C-1), 97.63 (C-1''), 88.13 (d, *J* = 177.6 Hz; C-2), 81.42 (d, *J* = 17.1 Hz; C-3), 71.19 (C-5), 70.49 (C-2''), 70.33 (C-2'), 70.21, 70.07 (C-3'', C-

3'), 69.62 (C-5'), 68.85 (C-5''), 66.86 (C-6), 67.12, 66.69 (C-4', C-4''), 65.90 (C-4), 63.04, 62.97 (C-6'', C-6'), 55.22 ppm (O*C*H$_3$). ^{19}F NMR (376 MHz, $CDCl_3$) δ -203.00 ppm (ddd, $J_{F,2}$ = 49.2 Hz, $J_{F,3}$ = 29.4 Hz, $J_{F,1}$ = 7.5 Hz; 1F, F-2). HR-MS (ESI) $[M+Na]^+$ m/z calcd for $C_{75}H_{65}O_{23}FNa$ 1375.3798; found 1375.3844.

Methyl α-D-mannopyranosyl-(1→3)-[α-D-mannopyranosyl-(1→6)-]-2-deoxy-2-fluoro-α-D-mannopyranoside (1c)

Compound **5** was deacylated to give **1c** (91%). R_f 0.11 (EtOAc/MeOH/water, 7:2:1). $[\alpha]_D^{20}$ + 79.6 (*c* 0.25; water). ^{1}H NMR (400 MHz, D_2O) δ 5.00 (as, 1H; H-1'), 4.92 – 4.74 (m, 3H; H-1, H-2, H-1''), 3.98 – 3.90 (m, 2H; H-2', 1H-6), 3.90 – 3.50 (m, 15H), 3.32 ppm (s, 3H; OC*H*$_3$). ^{13}C NMR (101 MHz, D_2O) δ 102.45 (C-1'), 99.41 (C-1''), 98.00 (d, J = 29.3 Hz; C-1), 88.78 (d, J = 172.9 Hz; C-2), 77.92 (d, J = 16.8 Hz; C-3), 73.27, 72.65, 70.61, 70.53, 70.30, 69.87, 69.82, 66.61 (overlapping), 65.37, 64.82, 60.88, 60.86, 55.11 ppm (O*C*H$_3$). ^{19}F NMR (376 MHz, D_2O) δ -204.15 ppm (ddd, $J_{F,2}$ = 49.1 Hz, $J_{F,3}$ = 32.9 Hz, $J_{F,1}$ = 7.3 Hz; 1F, F-2). HR-MS (ESI) $[M+Na]^+$ m/z calcd for $C_{19}H_{33}O_{15}FNa$ 543.1701; found 543.1687.

Methyl 3,4,6-tri-O-acetyl-2-deoxy-2-fluoro-α-D-mannopyranosyl-(1→3)-[2,3,4,6-tetra-O-benzoyl-α-D-mannopyranosyl-(1→6)-]-2-O-acetyl-α-D-mannopyranoside (6)

4,6-diol acceptor **3** and donor **4** were coupled to yield **6** (64%). R_f 0.31 (toluene/EtOAc, 2:1). $[\alpha]_D^{20}$ + 10.9 (*c* 1.0; $CHCl_3$). ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.12 – 8.01 (m, 4H; H_{Ar}), 7.97 – 7.92 (m, 2H; H_{Ar}), 7.85 – 7.79 (m, 2H; H_{Ar}), 7.62-7.54 (m, 2H; H_{Ar}), 7.53-7.47 (m, 1H; H_{Ar}), 7.46-7.33 (m, 7H; H_{Ar}), 7.29 – 7.22 (m, 2H; H_{Ar}), 6.12 (at, $J_{4'',3''}$ = $J_{4'',5''}$ = 10.0 Hz, 1H; H-4''), 5.94 (dd, $J_{3'',4''}$ = 10.0 Hz, $J_{3'',2''}$ = 3.2 Hz, 1H; H-3''), 5.78 (dd, $J_{2'',3''}$ = 3.2 Hz, $J_{2'',1''}$ = 1.8 Hz, 1H; H-2''), 5.40 (dd, $J_{1',F}$ = 7.3 Hz, $J_{1',2'}$ = 1.3 Hz, 1H; H-1'), 5.34 (at, $J_{4',3'}$ = $J_{4',5'}$ = 10.1 Hz, 1H; H-4'), 5.27 (d, $J_{1''-2''}$ = 1.8 Hz, 1H; H-1''), 5.26 – 5.16 (m, 2H; H-2, H-3'), 4.88 (m, 1H; H-2'), 4.77 – 4.70 (m, 2H; H-6''a, H-1), 4.58 – 4.48 (m, 2H; H-5'', H-6''b), 4.27 (dd, $J_{6'a,6'b}$ = 12.3 Hz, $J_{6'a,5'}$ = 5.7 Hz, 1H; H-6'a), 4.20 – 4.07 (m, 5H; H-5', H-6'b, H-6a, H-4, H-3), 3.95 (dd, $J_{6b,6a}$ = 11.6 Hz, $J_{6b,5}$ = 1.9 Hz, 1H; H-6b), 3.81-3.75 (m, 1H; H-5), 3.41 (s, 3H; OC*H*$_3$), 2.83 (d, $J_{OH,4}$ = 4.6 Hz, 1H; O*H*-4), 2.16, 2.11, 2.09, 2.07 ppm (4s, 12 H; 4 OCOC*H*$_3$). ^{13}C NMR (125 MHz, $CDCl_3$) δ 170.70, 170.56, 170.14, 169.56, (4 O*C*OCH$_3$), 166.30, 165.49, 165.46, 165.20 (4 *C*OPh), 133.43, 133.41, 133.18, 133.08, 129.82 -128.29 (24 C_{Ar}), 98.86 (*J* = 29.9 Hz; C-1'), 98.51 (C-1), 98.07 (C-1''), 86.79 (*J* = 180.0 Hz; C-2'), 77.13, 71.65 (C-

5), 71.25 (C-2), 70.30 (C-2"), 70.00 (C-3"), 69.85 (J = 16.8 Hz; C-3'), 69.43, 68.88 (C-5"), 67.12 (C-4"), 66.96, 66.39 (C-6), 65.56 (C-4'), 62.95 (C-6"), 62.18 (C-6'), 55.13 (O*C*H$_3$), 20.75, 20.73, 20.67, 20.59 ppm (4 OCO*C*H$_3$). ^{19}F NMR (376 MHz, $CDCl_3$) δ -203.61 ppm (ddd, $J_{F\text{-}2'}$ = 49.4 Hz, $J_{F\text{-}3'}$ = 28.2 Hz, $J_{F\text{-}1'}$ = 7.3 Hz; 1F, 2'-F). HR-MS (ESI) $[M+Na]^+$ m/z calcd for $C_{55}H_{57}O_{23}F$ 1127.3172; found 1127.3116.

Methyl 2-deoxy-2-fluoro-α-D-mannopyranosyl-(1→3)-[α-D-mannopyranosyl-(1→6)-]-α-D-mannopyranoside (1d)

Compound **6** was deacylated to give trisaccharide **1d** (80%). R_f 0.15 (EtOAc/MeOH/H_2O, 7:2:1). $\alpha]_D^{20}$ + 99.0 (*c* 0.5; water). ^{1}H NMR (500 MHz, D_2O) δ 5.16 (dd, $J_{1',F}$ = 8.0 Hz, $J_{1',2'}$ = 1.8 Hz, 1H; H-1'), 4.80 – 4.64 (m, 2H; H-1", H-2'), 4.57 (as, 1H; H-1), 3.94 (as, 1H; H-2) 3.89 – 3.71 (m, 7H), 3.70 – 3.47 (m, 9H), 3.25 ppm (s, 1H; OC*H*$_3$). ^{13}C NMR (126 MHz, D_2O) δ 100.96 (C-1), 99.44 (C-1"), 99.32 (d, J = 30.4 Hz; C-1'), 89.55 (d, J = 172.3 Hz; C-2'), 79.06, 73.27, 72.71, 70.81, 70.60, 69.96, 69.60 (d, J = 17.5 Hz; C-3'), 69.49, 66.73, 66.66, 65.52, 65.21, 60.95, 60.50, 54.85 ppm (O*C*H$_3$). ^{19}F NMR (376 MHz, D_2O) -204.73 ppm (ddd, $J_{F,2'}$ = 49.3 Hz, $J_{F,3'}$ = 31.6 Hz, $J_{F,1'}$ = 8.0 Hz; 1F; F-2'). HR-MS (ESI) $[M+Na]^+$ m/z calcd for $C_{19}H_{33}O_{15}FNa$ 543.1701; found 543.1709.

NMR experiments

Table S2: ^{1}H and ^{19}F NMR assignment of compounds 1, 2 and 3

Table S2a. 2-F-Man$_3$, compound **1**.

Position	***ManI***		***ManII***		***ManIII***	
	^{1}H	^{19}F	^{1}H	^{19}F	^{1}H	^{19}F
1	5.27		5.06		4.90	
2	4.81	-204.86	4.75	-205.97	4.89	-204.21
3	3.87		3.83		3.92	
4	3.66		3.66		3.86	
5	3.75		3.66		3.82	
6, 6'	3.82, 3.73		3.82, 3.73		4.03, 3.72	
Me					3.37	

Table S2b. 2-F-Man$_{2,\alpha1\text{-}3}$, compound **2.**

Position	***ManI***		***ManIII***	
	^{1}H	^{19}F	^{1}H	^{19}F
1	5.28		4.91	
2	4.81	-204.82	4.87	-204.03
3	3.87		3.92	
4	3.65		3.75	
5	3.75		3.65	
6, 6'	3.81, 3.73		3.83, 3.73	
Me			3.37	

Table S2c. 2-F-Man$_{2,\alpha1\text{-}6}$, compound **3.**

Position	***ManI***		***ManIII***	
	^{1}H	^{19}F	^{1}H	^{19}F
1	5.07		4.90	
2	4.75	-205.86	4.70	-206.00
3	3.83		3.76	
4	3.66		3.73	
5	3.66		3.75	
6, 6'	3.83, 3.73		4.00, 3.74	
Me			3.37	

^{19}F-R$_2$ filtered experiments

The K_D of compounds **1**, **1b-d**, **2** and **3** was estimated applying a ^{19}F-R$_2$ filtered approach. 6-F-ManαOMe, which weakly binds to DC-SIGN, was selected as the spy molecule. Relaxation rates R_2 were determined employing a CPMG pulse sequence, by fitting the observed ^{19}F signal intensity to the exponential decay curve:

$$I(t) = I_0 e^{-tR_2} = I_0 e^{-n2\tau R_2} \quad (1)$$

where $I(t)$ refers to intensity at time t, I_0 is intensity at $t = 0$, and R_2 is the transversal relaxation rate ($R_2 = 1/T_2$).

In the limit of fast exchange where the exchange contribution to the observed transversal relaxation rate $R_{2,obs}$ is insignificant (Figure S1, a)), the following equations apply:

$$R_{2,obs} = R_{2,f} + (R_{2,b} - R_{2,f})p_b \quad (2)$$

$$p_b = \frac{[P]_T + [L]_T + K_D - \sqrt{([P]_T + [L]_T + K_D)^2 - 4[P]_T[L]_T}}{2[L]_T} \quad (3)$$

where $[P]_T$ and $[L]_T$ are the total protein and ligand concentrarion respectively, $[R]_{2,f}$ and $[R]_{2,b}$ are the relaxation rates in the free and bound states, p_b is the fraction of bound ligand and K_D the dissociation constant of the protein-ligand complex. $[R]_{2,f}$ of 6-F-ManαOMe was measured in absence of the lectin, and Equation 2 was used to estimate $K_D$and $[R]_{2,b}$ for the complex (Figure S1, b)).

Then, K_I of compounds **1**, **1b-d**, **2** and **3** was measured in a competitive manner. $R_{2,obs}$ of 6-F-ManαOMe (spy molecule) in solution with DC-SIGN ECD was monitored at 5 different competitor concentrations ($[I]$) in each case with a fixed $[P]_T/[L]_T$ ratio (Figure S1, c)), to derive K_I by fitting to Equation 2 with p_b as defined in Equations (4) and (5) (Table 2):

$$p_b = \frac{2\cos(\theta/3)\sqrt{a^2 - 3b} - a}{3K_D + 2\cos(\theta/3)\sqrt{a^2 - 3b} - a} \quad (4)$$

$$\theta = \cos^{-1}\left(\frac{-2a^3+9ab-27c}{2\sqrt{(a^2-3b)^3}}\right), a = K_D + K_I + [L]_T + [I]_T - [P]_T, \quad (5)$$
$$b = ([I]_T - [P]_T)K_D + ([L]_T - [P]_T)K_D + K_I K_D, c = -K_I K_D [P]_T$$

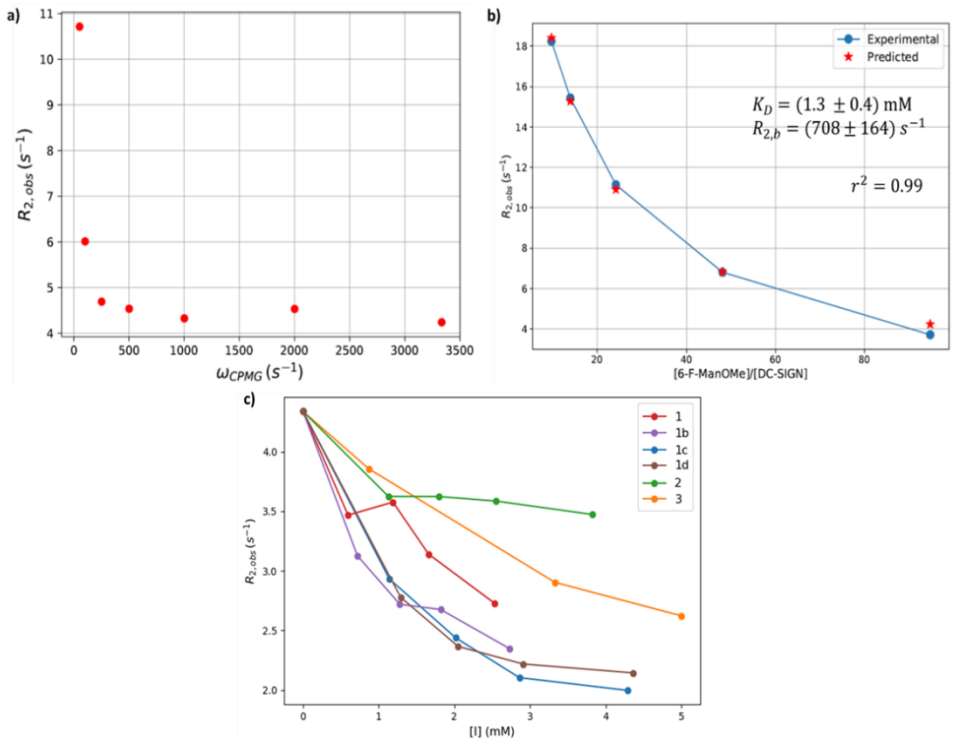


Figure S1. **^{19}F-R_2 filtered experiments**. a) Relaxation dispersion experiment for 6-F-ManαOMe (the spy molecule). $R_{2,obs}$ of the ^{19}F nucleus is measured for different values of ω_{CPMG}. Ligand and protein sample concentrations were: [6-F-ManαOMe] = 400 μM, [DC-SIGN (CRDs)] = 10 μM (counting concentration of CRDs, i.e, 4 CRDs per DC-SIGN ECD tetramer). For $\omega_{CPMG} > 1000\ s^{-1}$, there is virtually negligible exchange contribution to R_2. Therefore, all the subsequent R_2 filtered experiments were carried out with $\tau_{CPMG} = 1/\omega_{CPMG} = 1$ ms. b) K_D determination of 6-F-ManαOMe with DC-SIGN. $R_{2,obs}$ was measured for increasing amounts of [6-F-ManαOMe]/[DC-SIGN (CRDs)] (blue dots). K_D and $R_{2,b}$ were obtained from fitting to Equation 2, which is valid in the fast-exchange regime ($R_{ex} = 0$) [39a]. The predicted values at each [6-F-ManαOMe]/[DC-SIGN (CRDs)] are shown as red stars for comparison c) Titration curves showing the variation in ^{19}F-$R_{2,obs}$ of the spy molecule 6-F-ManαOMe, when increasing amounts of the competitors (I) **1**, **1b-d**, **2** and **3** are added to a mixture [6-F-ManαOMe] = 400 μM, [DC-SIGN (CRDs)] = 10 μM.

MD simulations

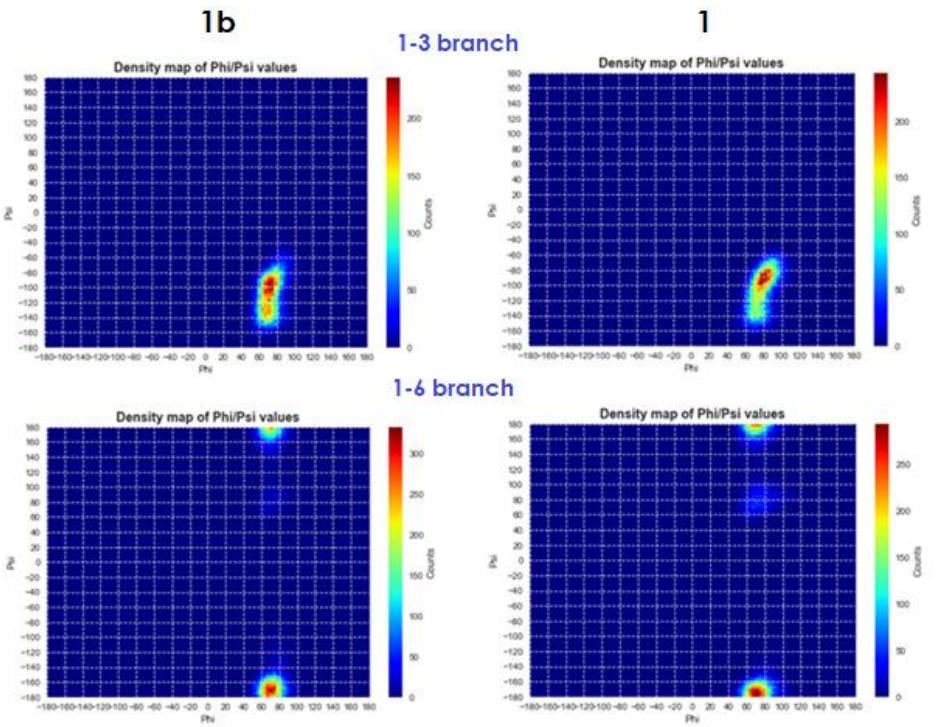


Figure S2. **Conformational maps.** Density of conformers populations around φ/ψ torsion angles computed for **1** and **1b** during a 500 ns MD simulation in explicit TIP3P water. φ and ψ torsion angles are defined as O5(*i*)-C1(*i*)-O*n*(*i*-1)-C*n*(*i*-1) and C1(*i*)-O*n*(*i*-1)-C*n*(*i*-1)-C(*n*-1)(*i*-1) respectively, where *n* indicates ring position and *i* a given residue. For **1b**, the GLYCAM 06-j [40] forcefield was employed, whereas GAFF2 [44] was used for **1**. The MD protocol in both simulations is described in the Materials and Methods section. The maps are fairly similar, independently of the employed force field.

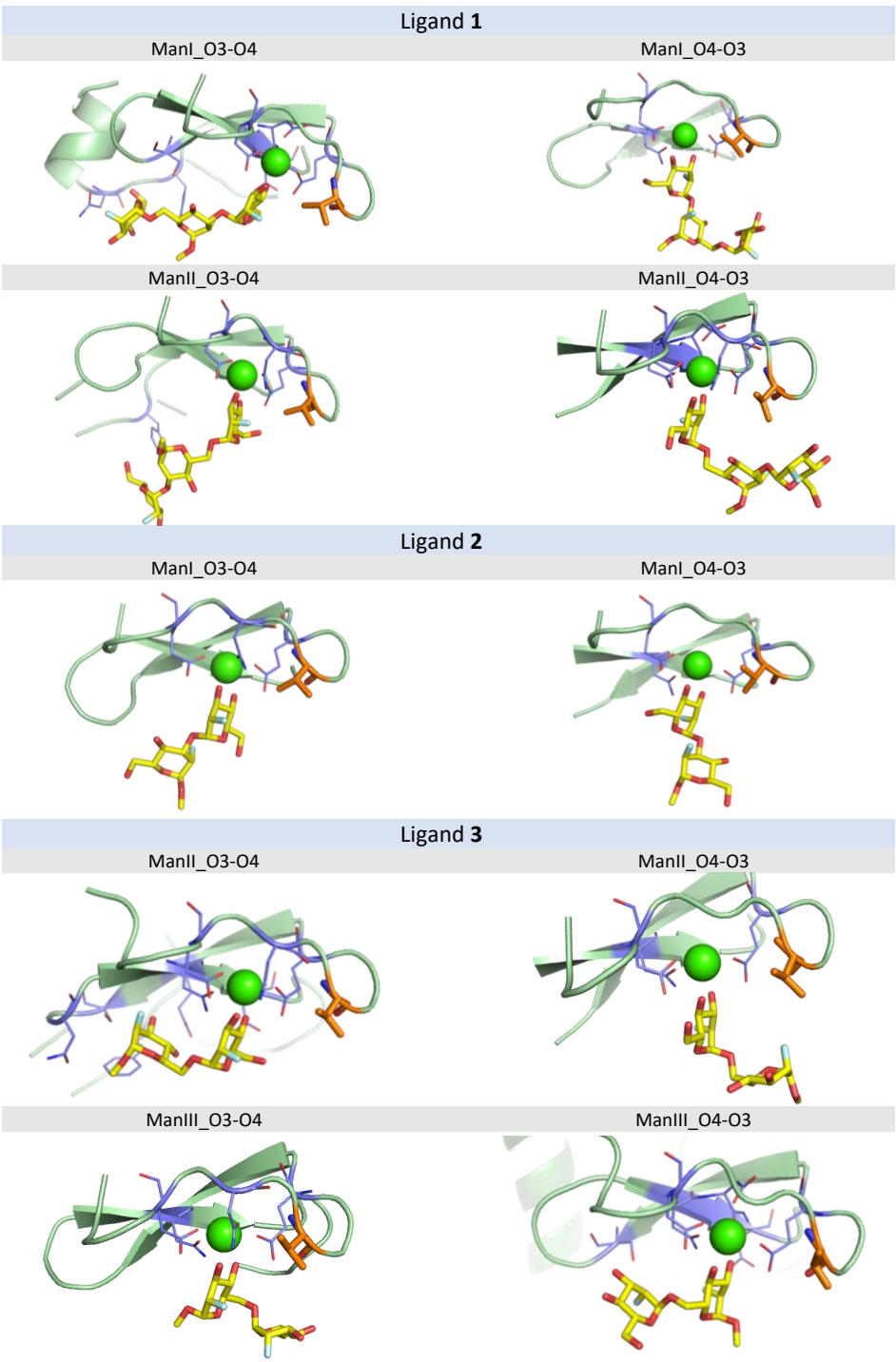


Figure S3. Selected MD frames: Representative optimized structures of each proposed binding mode for **1**, **2** and **3** in complex with DC-SIGN after system minimization of the first MD replica.

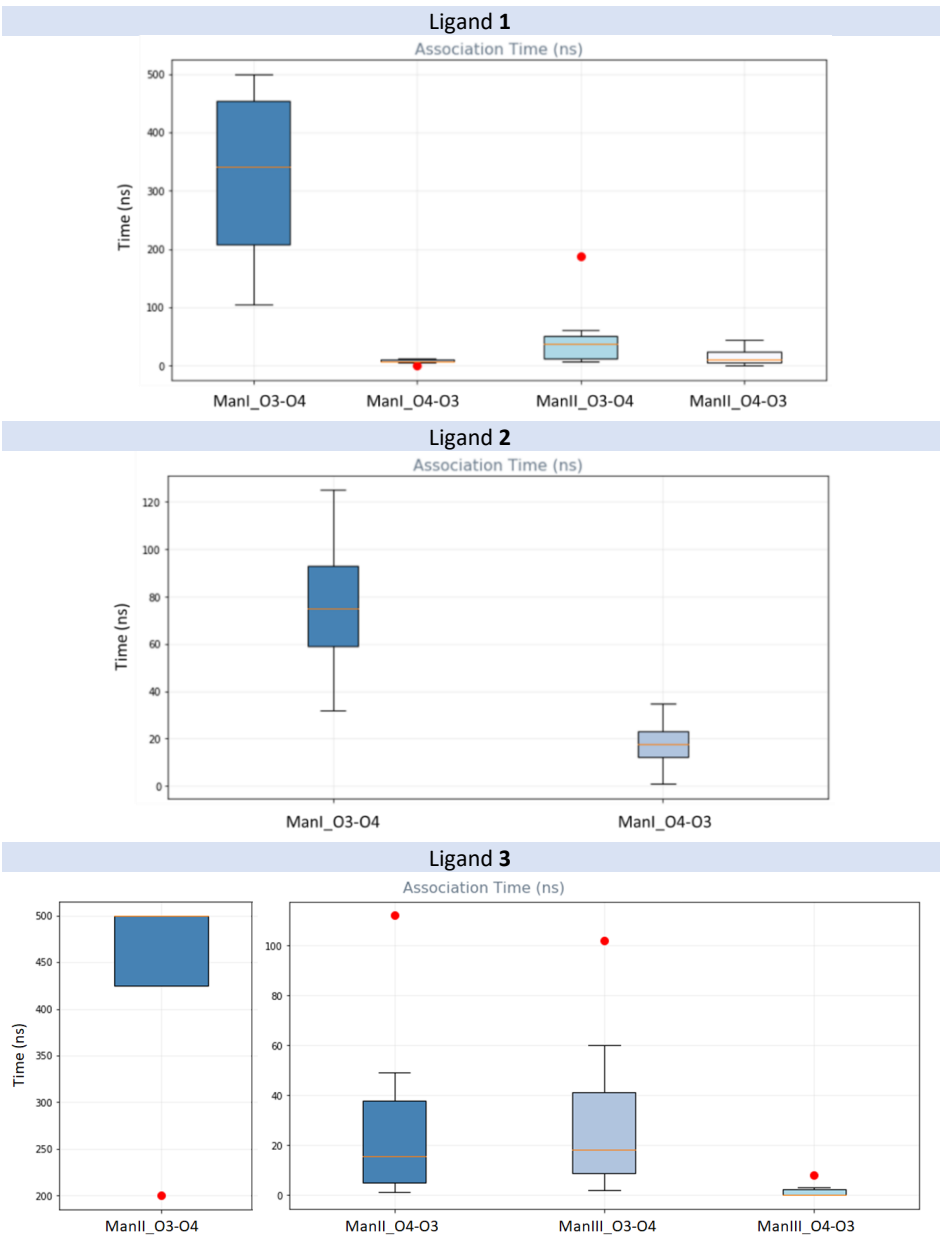
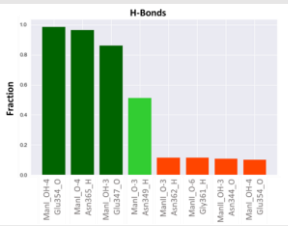


Figure S4. MD derived complexes association times: Box plot representation of association times observed in the MD simulations of the different ligand-protein complexes. The number of MD replicas ran in each case varies from 6 to 12, depending on the variability observed. Outliers are represented as red dots.

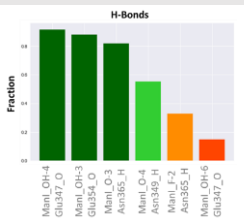
Ligand **1**

ManI_O3-O4



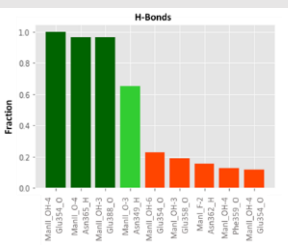
Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
CH-Pi	ManIII_H6/H6'	Phe313	8

ManI_O4-O3



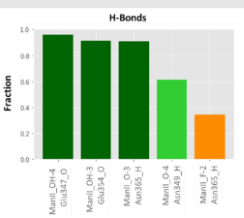
Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
VdW	ManI_H-3	Val351	21
VdW	ManII_H-4	Val351	29
VdW	ManIII_H-3	Val351	27
VdW	ManIII_H-5	Val351	23

ManII_O3-O4



Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
VdW	ManIII_H-4	Val351	20

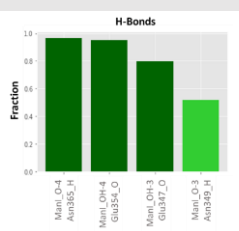
ManII_O4-O3



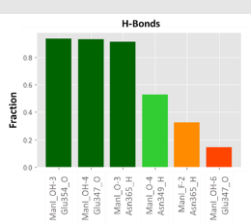
Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
VdW	ManI_H-3	Val351	50
VdW	ManI_H-5	Val351	15
VdW	ManII_H-3	Val351	40
VdW	ManIII_H-3	Val351	10
VdW	ManIII_H-4	Val351	13

Ligand **2**

ManI_O3-O4

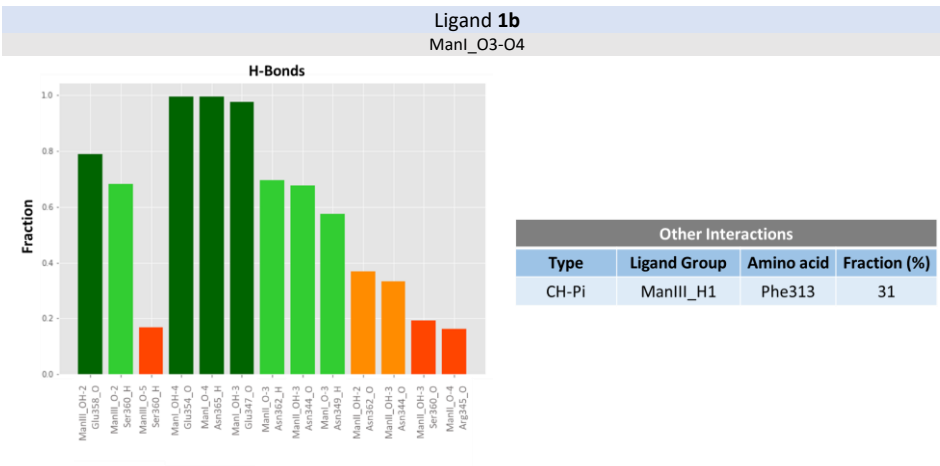


ManI_O4-O3



Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
VdW	ManI_H-3	Val351	12
VdW	ManII_H-4	Val351	14

Figure S5. **Ligand-protein interactions**: Significant ligand-protein interactions computed during the MD replicas. The fraction axis shows the percentage of the simulation time that the interaction is found. Hydrogen-bonds are accounted from the MD trajectories based on distance and angle criteria: d(A-H-D) < 3 Å and (A-H-D) < 130°, where H refers to the coordinates of the hydrogen atom, D and A the hydrogen bond donor and acceptor, respectively. Similarly, CH-Pi interactions are accounted by the distance of the aromatic ring-center to the hydrogen atom involved in the interaction according to d(Ring-H) < 3 Å, as well as the C/H/Ring-center angle (C-H-Ring) > 120°. Van der Waals interactions are considered when the interatomic distance of the atoms involved is lower than 1.2 times the sum of the VdW radii of the atoms.



Other Interactions			
Type	**Ligand Group**	**Amino acid**	**Fraction (%)**
CH-Pi	ManIII_H1	Phe313	31

Figure S6. Ligand-protein interactions for ligand 1b: Significant ligand-protein interactions computed during the MD of DC-SIGN bound to **1b** *via* ManI_O3-O4. All interactions are accounted in the same way as described in Figure S5

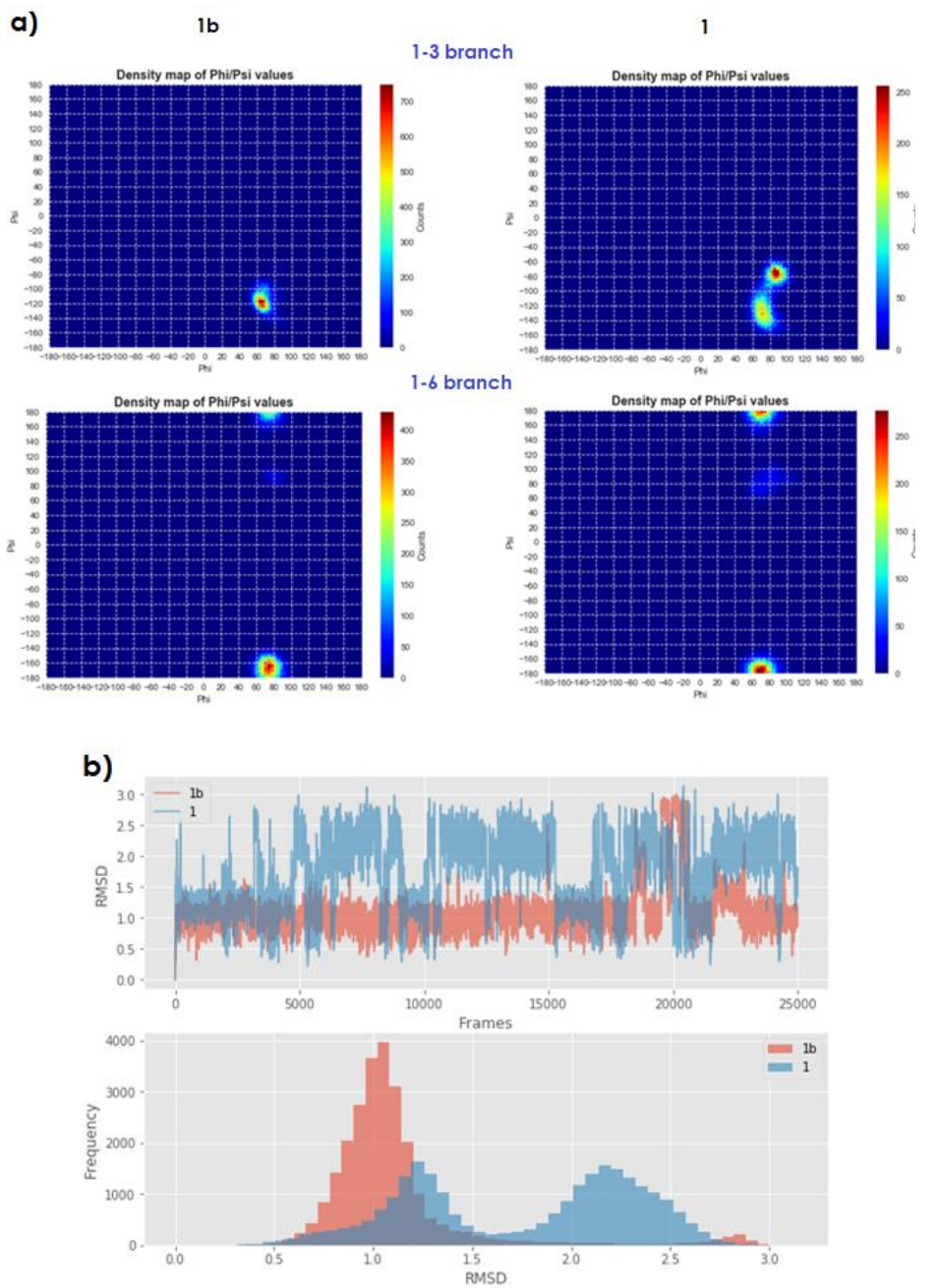


Figure S7. Conformational population comparisons: Comparison the populations around φ/ψ torsion angles (*a*)) and the RMSD (*b*)) of the ligands computed in a 500 ns MD simulation of **1** and **1b** bound *via* ManI_O3-O4 to DC-SIGN. It can be observed the larger mobility of the glycomimetic **1** with respect to the natural trimannose **1b** at the binding site when bound through the same pose.

CORCEMA-ST

The CORCEMA-ST script was ran sequentially for 400-800 frames extracted from each MD simulation trajectory. The same experimental parameters employed in the STD-NMR experiments were used in the calculations: [DC-SIGN] = 9.14 uM, [Ligand] = 1.4 mM, 2 s saturation time. Different k_{on} values in the range of 10^5-10^8 M^{-1} $s^{-1}$and K_D 0.5-3 mM were tested with all the complexes, giving rise to very similar normalized calculated STD profiles. Thus, a k_{on} of 10^6 M^{-1} s^{-1}, of the same order of other sugar-lectin systems [1,2], was used for all the models. K_D was set to 1 mM, similar to the observed K_D of other Man derivatives in complex with DC-SIGN [3-5]. An instantaneous irradiation of the aliphatic receptor residues Ile, Leu, and Val methylgroups to account for the selective on-resonance irradiation of the STD-NMR experiment, 0.85 ppm, was used. The size of the relaxation matrix was adjusted using a distance cutoff, *d*, of 10 Å away from any ligand atom, since virtually the same STD profiles were obtained for larger values, while some differences appeared when $d < 10$ Å. The value of the order parameter S^2 and the methyl group internal correlation time τ_m were set to 0.85 and 10 ps respectively, as previously described [6]. A typical value for the free ligand correlation time τ_L = 0.5 ns was used, whereas for the bound ligand a correlation time assuming a tetrameric protein of globular shape was estimated as τ_b = 85 ns.

Since CORCEMA-ST does not recognize ^{19}F as an active nucleus in the relaxation matrix, the effect of the presence of an active nuclei at position C-2 in the 2-F compounds **1**, **2** and **3**, was assessed by substituting all fluorine atoms by hydrogens in each analogue, while keeping the original C-F distance. Then, CORCEMA-ST calculations were ran with the same parameters described in the previous paragraph. This way, the ^1H nucleus is used as a probe to simulate the most pronounced expectable effect (since it can give rise to homonuclear cross-relaxation) on the observed STD signals. Remarkably, it was found that the predicted best fitting models with *BM-Mixer* are in general unaffected by the presence of the active nucleus at C-2 for the three ligands (see Table S3).

Best-model STD fitting

BM-Mixer is able to find the best combination of frames (in %) from different MD trajectories explaining the experimental STD-NMR data. For the program to work properly, it is important to use a list of experimental STDs only containing reliable assigned peaks. In this work, we used the list provided in Table 1 in the main text, with the exception of compound **3**. For compound **3**, the experimental STD cross peak observed for Man III at 3.7 ppm could correspond to H-3, H-4, H-5 or H-6'. As accounting the

measured STD intensity as the sum of the individual contribution from four H atom would potentially introduce noise in the search (see *CORCEMA-ST and best-model STD fitting* heading in the experimental section), the corresponding STD peak was not taken into account in the search for best-model fitting showed in Figure 4 and Table S2.

There are two main parameters that must be set in a *BM-Mixer* run: *mix_leap* and *search_iterator*. *mix_leap* defines the minimum percentage of frames to be used from each trajectory to explore the different combinations. For example, setting *mix_leap* = 5 allows the program to combine frames from different trajectories using a minimum of 5 % of the frames. Although it depends on the number of frames and binding modes (trajectories) to work with, typically a value of *mix_leap* = 10 is sufficient to get accurate enough results (according to NOE R-Factor$_{Rel}$) in a decent amount of time. *search_iterator* specify how many times the program is run before computing the final NOE R-Factor$_{Rel}$ averages. Every time a new iteration start (when *search_iterator* > 1), the frames used in each combination are randomly selected, so that the larger the value of *search_iterator*, the better sampling of the trajectory-space is done. In general, we have found that for the studied systems, when using 400-800 frames of each binding mode trajectory, best NOE R-Factor$_{Rel}$ averages are similar when setting *search_iterator* > 15.

Table S1. Best-model STD fitting by *BM-Mixer*. Top 3 best-model STD fitting results for each ligand, found by *BM-Mixer*. For ligands **1** and **3**, 800 frames from each simulated binding mode were used in the calculations, while 400 frames were employed for **2**. *mix_leap* was set to 10 for ligands **1** and **3**, and to 5 for **2**; a *search_iterator* of 30 was used in all cases.

Ligand **1**				
ManI_O3-O4 (%)	ManI_O4-O3 (%)	ManII_O3-O4 (%)	ManII_O4-O3 (%)	NOE R-Factor$_{Rel}$
60	0	0	40	0.2010
50	0	0	50	0.2201
40	0	10	50	0.2238
Ligand **2**				
ManI_O3-O4 (%)		ManI_O4-O3 (%)		NOE R-Factor$_{Rel}$
65		35		0.1380
60		40		0.1396
70		30		0.1575
Ligand **3**				
ManII_O3-O4 (%)	ManII_O4-O3 (%)	ManIII_O3-O4 (%)		NOE R-Factor$_{Rel}$
50	10	40		0.1470
50	0	50		0.1482
40	10	50		0.1496

Table S3. Best-model STD fitting by *BM-Mixer* with non-fluorinated control
Top 3 best-model STD fitting results found by *BM-Mixer* for each ligand-control CORCEMA-ST calculated STD. Ligand-controls were built by substituting all fluorine atoms in the MD trajectories by hydrogens, and then computing CORCEMA-ST on those. The same *BM-Mixer* set up described in Table S2 was used.

Ligand **1**				
ManI_O3-O4 (%)	ManI_O4-O3 (%)	ManII_O3-O4 (%)	ManII_O4-O3 (%)	NOE R-Factor$_{Rel}$
60	0	0	40	0.2143
40	0	10	50	0.2212
40	0	10	50	0.2368
Ligand 2				
ManI_O3-O4 (%)		ManI_O4-O3 (%)		NOE R-Factor$_{Rel}$
70		30		0.1467
65		35		0.1538
75		25		0.1681
Ligand 3				
ManII_O3-O4 (%)	ManII_O4-O3 (%)	ManIII_O3-O4 (%)		NOE R-Factor$_{Rel}$
50	0	50		0.1555
40	10	50		0.1645
50	10	40		0.1654

References

1. Scharenberg, M.; Jiang, X.; Pang, L.; Navarra, G.; Rabbani, S.; Binder, F.; Schwardt, O.; Ernst, B. Kinetic properties of carbohydrate-lectin interactions: FimH antagonists. *ChemMedChem* **2014**, *9*, 78–83.

2. Milton, J.D.; Fernig, D.G.; Rhodes, J.M. Use of a biosensor to determine the binding kinetics of five lectins for Galactosyl-N-acetylgalactosamine. *Glycoconj. J.* **2001**, *18*, 565–569.

3. Holla, A.; Skerra, A. Comparative analysis reveals selective recognition of glycans by the dendritic cell receptors DC-SIGN and Langerin. *Protein Eng. Des. Sel.* **2011**, *24*, 659–669.

4. Bordoni, V.; Porkolab, V.; Sattin, S.; Thépaut, M.; Frau, I.; Favero, L.; Crotti, P.; Bernardi, A.; Fieschi, F.; Di Bussolo, V. Stereoselective innovative synthesis and biological evaluation of new real carba analogues of minimal epitope Manα(1,2)Man as DC-SIGN inhibitors. *RSC Adv.* **2016**, *6,* 89578–89584.

5. Reina, J.J.; Sattin, S.; Invernizzi, D.; Mari, S.; Martinez-Prats, L.; Tabarani, G.; Fieschi, F.; Delgado, R.; Nieto, P.M.; Rojo, J.; et al. 1,2-mannobioside mimic: Synthesis, DC-SIGN interaction by NMR and docking, and antiviral activity. *ChemMedChem* **2007**, *2*, 1030–1036.

6. Jayalakshmi, V.; Krishna. N.R. Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. *J. Magn. Reson.* **2002**, *155*, 106–118.

CHAPTER 6

Conclusions

CONCLUSIONS

From the work described on this essay, the following foremost conclusions can be drawn:

- ^{19}F-NMR provides a myriad of possibilities to address the study of carbohydrate-lectin interactions from the point of view of the ligand. Additionally, it presents a number of important advantages from a practical perspective too. To name a few, the fluorine detection in NMR enormously facilitates the spectra interpretation -especially in cases where several molecules are present in solution-, circumvents the need of using deuterated buffers (that are expensive, might harm the receptor, and require tedious and time consuming procedures of solvent exchange), and ^{19}F-NMR-based experiments can be typically acquired in a reasonable period of time without requiring complicated experimental setups (in general).
- The use of a rationally designed library of mono-fluorinated monosaccharides -in which each OH group is systematically substituted by F- in combination with ^{19}F-NMR-T_2 experiments, have demonstrated to be a simple and versatile approach to perform screening and chemical-mapping analysis, applicable to lectins from various sources. In particular, it allowed deducing that:
  - MGL recognizes Gal moieties, pointing out the importance of the sugar hydroxyls 3 and 4 for binding.
  - PSA binds Glc and Man analogues and tolerates OH-by-F modifications at OH 2 and 3, whereas substitution at OH 6 seems to reduce binding, which is abolished when the substitution takes place at position 4.
  - GGBP interacts with Glc and Gal monosaccharides, and only OH 2 and 3 turns out to be crucial for binding (the pattern is completely opposite to PSA).
- The study of the interaction of DC-SIGN with a library of mono-fluorinated monosaccharides by means of the abovementioned screening and chemical mapping strategy, revealed the existence of an alternative binding mode (with respect to the natural sugar) for the 4F-Man/DC-SIGN complex. ^{1}H-STD experiments and MD simulations confirmed that the binding pose mimics that of Fuc/DC-SIGN complexes. These results suggest that both the crystal-observed and the herein-described binding poses of Man at the binding site coexist in solution.

- The 2D-STD-TOCSYreF experiment provides a powerful tool to help determine the binding epitope of fluorinated ligands in complex molecules. The H-F heteronuclear correlation dramatically improves the resolution of the STD-NMR spectra. Although the overall the sensitivity of the experiment is lower (which translates into increased acquisition times), the quality of the obtained STD information significantly outperforms the sensitivity loss in the study of complex systems.
- A computer program, named *BM-Mixer*, which allows disentangling the individual contribution from different binding modes to the experimental STD of ligands that weakly bind to receptors in a multi-modal manner, has been written. BM-Mixer analyzes the ensembles of theoretical STD data from different binding modes of the ligand-receptor complex, obtained by means of CORCEMA-ST calculations performed on different MD trajectories, and compares these with the experimentally obtained STDs. By performing sequential mixing stages that exhaustively cover different proportions of the modeled binding modes to create new combined theoretical STD data sets, which are evaluated with a NOE R-factor-like function, the program is capable of finding the contribution from each binding mode that best fit the experimental STD data.
- In order to interpret the experimental data from the 2D-STD-TOCSYreF experiments on the DC-SIGN/$(2F\text{-}Man)_n$ system, a new protocol combining extensive MD simulations, CORCEMA-ST calculations and BM-Mixer post-processing has been proposed. This protocol permitted quantitatively analyzing the STD intensities of the 2D-STD-TOCSYreF experiments, determining the proportions of each ligand-protein complex coexisting in solution which best matched the observed STD data. Interestingly, the most populated complexes were similar to those found in several X-ray crystallographic structures.

Appendix

A.I R_2 equations for ^{19}F and ^{1}H

$$R_2^{\mathrm{F}} = \frac{\gamma_{\mathrm{F}}^2\gamma_{\mathrm{H}}^2\hbar^2\tau_{\mathrm{c}}}{20}\sum_{\mathrm{H}_i}\frac{1}{r_{\mathrm{FH}_i}^6}\left\{4+\frac{1}{1+(\omega_{\mathrm{F}}-\omega_{\mathrm{H}})^2\tau_{\mathrm{c}}^2}+\frac{3}{1+\omega_{\mathrm{F}}^2\tau_{\mathrm{c}}^2}+\frac{6}{1+\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}^2}+\frac{6}{1+(\omega_{\mathrm{F}}+\omega_{\mathrm{H}})^2\tau_{\mathrm{c}}^2}\right\} + \frac{2}{15}\Delta\sigma_F^2\left(1+\frac{\eta_{\mathrm{CSA,F}}^2}{3}\right)\omega_{\mathrm{F}}^2\tau_{\mathrm{c}}\left\{\frac{2}{3}+\frac{1}{2(1+\omega_{\mathrm{F}}^2\tau_{\mathrm{c}}^2)}\right\} \quad (1)$$

$$R_2^{\mathrm{H}} = \frac{\gamma_{\mathrm{H}}^4\hbar^2\tau_{\mathrm{c}}}{20}\sum_{\mathrm{H}_j\neq\mathrm{H}_i}\frac{1}{r_{\mathrm{H}_i\mathrm{H}_j}^6}\left\{5+\frac{9}{1+\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}^2}+\frac{6}{1+4\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}^2}\right\} + \frac{2}{15}\Delta\sigma_H^2\left(1+\frac{\eta_{\mathrm{CSA,H}}^2}{3}\right)\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}\left\{\frac{2}{3}+\frac{1}{2(1+\omega_{\mathrm{H}}^2\tau_{\mathrm{c}}^2)}\right\} \quad (2)$$

where γ is the gyromagnetic ratio, ω is the Larmor frequency of the nucleus, $r_{\mathrm{H}_i\mathrm{H}_j}$ is the internuclear distance between the observed proton nucleus H_i and all neighbor protons H_j, $r_{F\mathrm{H}_i}$ is the internuclear distance between the observed fluorine nucleus and all neighbor protons H_i, $\Delta\sigma$ is the chemical shift anisotropy, η_{CSA} is the asymmetry parameter, and τ_{c} is the rotational correlation time. In practice, notice that $\Delta\sigma_F \gg \Delta\sigma_H$.

A.II k_{ex} expressions

The ligand and receptor exchange rate constants for the process described in Equation 16, are:

$$k_{ex}^{L} = k_{on}[P] + k_{off} \quad (3)$$

$$k_{ex}^{E} = k_{on}[L] + k_{off} \quad (4)$$

where k_{on} is the rate constant of the complex formation, k_{off} the rate constant of the complex dissociation, and $[P]$ and $[L]$ are the concentrations of receptor and ligand respectively.

A.III Non-stochastic methods for temperature control

The link between the average kinetic energy (i.e, particle velocities) and temperature is given by the *equipartition theorem*:

$$T = \frac{2}{3}\frac{\langle K\rangle_{NVT}}{Nk_B} = \frac{1}{3}\frac{\langle mv^2\rangle}{k_B} = \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i {v_i}^2 \quad (5)$$

where $\langle K\rangle_{NVT}$ is the average kinetic energy. From Equation 5 is easy to derive a relationship for the change in temperature with a scaling factor λ:

$$\begin{aligned}\Delta T = T_f - T_0 &= \frac{1}{3Nk_B}\sum_{i=1}^{N}(m_i(\lambda v_i)^2 - m_i {v_i}^2) \\ &= \frac{1}{3Nk_B}\sum_{i=1}^{N} m_i {v_i}^2(\lambda^2 - 1) = T_0(\lambda^2 - 1)\end{aligned} \quad (6)$$

$$\lambda = \sqrt{T_f/T_0} \quad (7)$$

T_f and T_0 are the requested and current system temperature respectively. At each simulation step, velocities of all particles are adjusted using λ calculated from Equation 7.

Other typical approach consists on coupling the system to an imaginary 'heat bath' at the desired temperature. This bath is able to remove or supply heat from the original system as needed on a finite timescale [184]. This way, the strength of the coupling between the bath and the system is controlled via a coupling parameter τ, so that the scaling factor takes the form:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau}\left(\frac{T_f}{T_0} - 1\right)} \quad (8)$$

where Δt is the time step of the simulation. This method, which is equivalent to simply velocity scaling when $\tau = \Delta t$, allows the temperature to fluctuate around its requested value, so that the perturbation of the system as a consequence of the scaling can take place softer than in the previous case.

A.IV The Andersen scheme for temperature control

The Andersen method basically consists on simulating imaginary sporadic collisions of a heat bath with randomly selected particles of the system, employing a user-defined *collision frequency*. The velocity of the selected particles after each fictitious collision is drawn from a Maxwell-Boltzmann distribution at the desired temperature. Between collisions, the energy is constant and the Newtonian motion applies. Therefore, the optimal collision

rate would be one sufficiently high so that the canonical distribution was properly sampled, but without disturbing the system so frequently that the configurational space was sampled slowly, and the kinetic energy fluctuations was not able to fluctuate. Different variants of the Andersen thermostat (for instance, differences in the number of particles that experiences the imaginary collisions, or the way in which the particles undergoing a collision are chosen) can be found on different MD simulation programs.

A.V Pressure computation

The pressure is usually computed in MD simulations via the *virial theorem* of Clausius, and in the case of a real gas or liquid is given by [197]:

$$P = \frac{1}{V}\left[Nk_BT - \frac{1}{3}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\boldsymbol{F}_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)\right] \qquad 9$$

where $\boldsymbol{F}_{ij}$ is the force acting on molecule $\boldsymbol{i}$ due to molecule $\boldsymbol{j}$. To notice, the sum generally include only interactions between molecules, since intramolecular forces do not contribute to the pressure. Therefore, coordinates and forces are evaluated at each molecular center of mass. Equation 9 shows that pressure changes can take place by varying the virial through scaling of intermolecular distances.

A.VI Monte Carlo barostat

Monte Carlo barostats regulate the pressure of the system by introducing trial moves on the volume of the periodic box at some fixed time interval. These trial moves are randomly generated, and are submitted to an acceptance criteria according to the standard MC probability[32]. In this case, the true NPT ensemble is sampled, and the computation is in general substantially more efficient than for the previous method, since it is not required to calculate the virial at each time step. The use of stochastic algorithms alter the true dynamics of the system, which might be distorted to some extent, preventing the calculation of time-dependent properties.

[32] Moves are accepted only if the potential energy of the new configuration (i.e, with corrected volume) is lower than the previous one, and otherwise according to the probability given by the *Boltzmann factor* $e^{-\Delta V/kT}$.