

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE FARMACIA**



**TESIS DOCTORAL**

**Sugars and Proteins: the role of dynamics in  
molecular interactions**

**Azúcares y Proteínas: el papel de la dinámica en  
las interacciones moleculares**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

**Luca Unione**

DIRECTORES

**Jesús Jiménez Barbero**  
**F. Javier Cañada Vicinay**

Madrid, 2017



UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Farmacia

**Azúcares y Proteínas:**

**El papel de la dinámica en las interacciones moleculares**

Memoria para optar el grado de doctor presentada por

Luca Unione

Madrid, 2016

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

Bajo la codirección de

Dr. Jesús Jiménez Barbero

Dr. F. Javier Cañada Vicinay





UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Farmacia

**Sugars and Proteins:  
The Role of Dynamics in Molecular Interactions**

Memoria para optar el grado de doctor presentada por

Luca Unione

Madrid, 2016

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

Bajo la codirección de

Dr. Jesús Jiménez Barbero

Dr. F. Javier Cañada Vicinay



## ACKNOWLEDGEMENTS

The work presented in this Thesis has been carried out at the Center for Biological Research of the Spanish National Research Council (CIB-CSIC), under the supervision of Prof. Jesús Jiménez-Barbero and Prof. Francisco Javier Cañada Vicinay, and thanks to a Marie Curie Fellowship granted by the EU (ITN/ETN Marie Curie Action Dynano).

I thank the EU for the grant. I also thank the Institutions: CIB-CSIC (Madrid, Spain) and the CICbioGUNE Centre for Cooperative Research in Biosciences (Derio, Spain).

To the centres where I had the opportunity to develop part of my work also go my special acknowledgements: CEU Universidad San Pablo (Madrid), under the supervision of Dr. Sonsoles Martín Santamaría; Université de Namur (Belgium), under the supervision of Prof. Stéphane Vincent; CICbiomaGUNE (San Sebastian), under the supervision of Dr. Niels-Christian Reichardt.

My gratitude to all the authors who participated to the works presented in this Thesis.

My deep appreciation to all the Scientists with whom I have had the honour and pleasure to discuss during these years, in my work institutions and in several meeting.

To my colleagues.

Due to all the above mentioned reasons, my main thank to Jesús and Javier.



# CONTENTS

ACKNOWLEDGEMENTS	III
CONTENTS	V
ABBREVIATIONS	XI
ABSTRACT	XV
RESUMEN	XVI
ROUTE MAP OF THIS THESIS	1
CHAPTER I. INTRODUCTION	3
1.1 Glycomics	3
1.1.1 Language. The sugar code	3
1.1.2 Alphabet. Monosaccharides	4
1.1.3 Vocabulary. Oligosaccharides	8
1.1.4 The anomeric effect	10
1.2 Glycomimetics	11
1.2.1 <sup>19</sup> F: spy atom	13
1.3 Lectins	13
1.3.1 Interlocutor. Sugar Receptor	13
1.4 NMR	15
1.4.1 The reader. Nuclear Magnetic Resonance	15
1.4.2 Principles of NMR	16
1.4.3 NMR for carbohydrates	18
1.4.4 Coupling constants	19



1.4.5 Nuclear Overhauser Effect. NOE Spectroscopy (NOESY)	20
1.4.6 Long range restraints. PCSs, RDCs and PREs	21
1.5 Computational Chemistry	24
1.5.1 Quantum Chemical Models	24
1.5.2 Density Functional Models	25
1.5.3 Molecular Mechanics (MM) Models	26
1.5.4 Molecular Dynamics (MD)	28
1.6 Objectives	30
1.7 References	31
CHAPTER II. Glycomimetics: understanding conformational plasticity in sugar rings.	41
2.1 Introduction	42
2.2 Results and Discussion	44
2.2.1 NMR analysis	44
2.2.2 Computational analysis	53
2.3 Conclusions	58
2.4 Methods	58
2.4.1 NMR Spectroscopy	58
2.4.2 Computational Methods	59
2.5 Supporting Information	60
2.6 References	64
CHAPTER III. <i>gem</i> -diF-carbadiisaccharides: the importance of the stereoelectronic component for the <i>exo</i> -anomeric effect.	67

3.1 Introduction	68
3.2 Results and Discussion	70
3.2.1 <i>Ab initio</i> calculations	70
3.2.2 NMR Conformational studies	73
3.2.3 Intramolecular hydrogen bond	78
3.3 Conclusions	81
3.4 Methods	81
3.4.1 <i>Ab initio</i> calculations	81
3.4.2 NMR spectroscopy	81
3.4.3 Molecular Mechanics calculations	82
3.5 Supporting Information	83
3.6 References	84
CHAPTER IV. Sugar Receptor: unraveling the amplitude and time scale of macro- molecular motions	87
4.1 Introduction	88
4.2 Results and Discussion	91
4.2.1 Self-alignment with a paramagnetic tag	91
4.2.2 MD simulations	97
4.2.3 The synergy between structure and dynamics is essential for protein function	101
4.3 Conclusions	102
4.4 Methods	103
4.4.1 Sample preparation for paramagnetic studies	103
4.4.2 NMR Spectroscopy	103
4.4.3 Molecular Dynamic simulations	103

4.4.4 Analysis of the trajectories	104
4.5 References	105
CHAPTER V. CH/ $\pi$ interactions: elucidating fundamental weak intermolecular forces	109
5.1 Introduction	110
5.2 Results and Discussion	112
5.2.1 DFT prediction of J couplings between the nuclei involved in CH- $\pi$ intermolecular interactions	112
5.2.2 Sugar deuteration	115
5.3 Partial conclusions	116
5.4 Methods	116
5.4.1 DFT calculations	116
5.4.2 Synthesis	117
5.4.3 Acetyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-[1- $^2$ H]glucopyranoside	117
5.4.4 [1,1,1- $^2$ H]Methyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-[1- $^2$ H]glucopyranoside	118
5.4.5 [1,1,1- $^2$ H]Methyl- $\beta$ -D-[1- $^2$ H]glucopyranoside	119
5.4.6 Selective deuterium labelling	120
5.5 Supporting information	121
5.6 References	129
CONCLUSIONS	133
CONCLUSIONES	136





## ABBREVIATIONS

AFM: Atomic-force microscopy

B3LYP: Becke, three-parameter, Lee-Yang-Parr

COSY: Correlation spectroscopy

DFT: Density functional theory

DMSO: Dimethyl sulfoxide

Gal: D-galactose

GalNAc: N-Acetylgalactosamine

GGBP: glucose/galactose binding protein

Glc: D-glucose

HOESY: Heteronuclear Overhauser effect spectroscopy

HR-MAS: High-Resolution Magic Angle Spinning

HSQC: Heteronuclear single quantum coherence

Ido: L-idose

IR: Infrared spectroscopy

Kdo: 3-deoxy-D-mannoctulosonic acid

L-IdoA: L-iduronic acid

MAS: Magic Angle Spinning

Man: D-mannose

MD: molecular dynamics

MM: molecular mechanics

NMR: nuclear magnetic resonance

NOE: nuclear Overhauser effect

NOESY: nuclear Overhauser effect spectroscopy

PCS: pseudocontact chemical shift

PRE: paramagnetic relaxation enhancement

QM: quantum mechanics

RDC: residual dipolar coupling

TOCSY: total correlation spectroscopy

UV: ultraviolet-visible spectroscopy







## Resumen

Los glicanos constituyen unos de los tipos de moléculas más variadas y complejas entre los sistemas biológicos. Las diferentes ramas del árbol de la vida se pueden distinguir en base a la composición de los glicanos de los propios glicoconjugados. Sin embargo, cuánto conocemos acerca de la función y distribución de los glicanos y glicoconjugados es una cuestión todavía abierta. Hace no demasiados años, nuestro conocimiento acerca de las glicoproteínas era considerablemente escaso.

De hecho, durante mucho tiempo se creyó que la N-glicosilación de proteínas se daba sólo en los organismos eucariotas. Sin embargo, hoy está firmemente establecido que esta compleja modificación ocurre también en bacterias y arqueas. Por lo tanto, en los últimos 10 años, el campo de las glicoproteínas ha sido testigo de enormes avances en el descubrimiento de nuevos e inusuales carbohidratos, así como en la elucidación de las enzimas responsables de la construcción y procesamiento de glicanos y en la comprensión del impacto biológico que estas modificaciones tienen sobre la estructura y función de la proteína diana. La razón de este “retraso”, probablemente radica en la intrínseca complejidad estructural, heterogeneidad y flexibilidad de los glicanos. Como contrapeso, numerosos y exhaustivos trabajos en el campo de la glicómica han demostrado que es exactamente gracias a su complejidad estructural, heterogeneidad y flexibilidad el porqué los glicanos han sido seleccionados, entre otras biomoléculas, como intermedios claves para los procesos celulares de proliferación, diferenciación, adhesión, infección, comunicación, etc.

En esta tesis, hemos intentado mirar profundamente en el interior de la estructura de los glicanos, con el objetivo de conciliar sus propiedades estructurales intrínsecas, a nivel atómico, con los motivos de su flexibilidad molecular. Cuando los glicanos interactúan con sus receptores, esta flexibilidad se une a la plasticidad del sistema global para dar al lugar al proceso de reconocimiento molecular. De hecho, las dos partes que participan en la interacción intermolecular pueden adaptar su superficie de contacto de manera que se maximice la entalpía de unión. De manera alternativa o complementaria, el glicano y el receptor podrían mantener movimientos moleculares internos, incluso en el estado unido, de tal manera que se minimizase la penalización entrópica del proceso de unión. Como consecuencia de ello, el papel de la compensación entálpico/entrópica no es fácil de evaluar ni de medir.

A lo largo de esta tesis nos hemos enfocado en la interacción entre azúcares y proteínas, empezando por el estudio de la flexibilidad en los anillos de los monosacáridos, siguiendo por la investigación de los movimientos alrededor de los enlaces glicosídicos en disacáridos, para después pasar a la descripción cuantitativa de los movimientos macromoleculares internos de un receptor. Por último, se han discutido las interacciones intermoleculares de tipo CH/ $\pi$  que contribuyen a la estabilidad de los complejos de proteína-azúcar y se ha propuesto una nueva estrategia para su detección directa.

Capítulo I.

La introducción de esta tesis se ha centrado en el código de los carbohidratos, en cómo la información biológica se almacena en la estructura de los glicanos y cómo este mensaje es leído y traducido por sus receptores. También se discute el uso de glicomiméticos y su empleo como moléculas de interés bio-médico y espectroscópico. El capítulo concluye con una descripción de los métodos experimentales que se han empleado en esta tesis para elucidar la estructura y las interacciones de los glicanos.

## Capítulo II.

En este capítulo se discute la flexibilidad de los azúcares a nivel de monosacárido. Se presenta una nueva generación de glicomiméticos que contienen átomos de flúor. Hemos aplicado una combinación de espectroscopia de RMN y métodos computacionales para investigar el comportamiento dinámico de los anillos de idosa y glucosa. Más específicamente, se han utilizado experimentos de RMN a baja temperatura para ralentizar el intercambio conformacional de los anillos de idosa. En estas condiciones, el intercambio es lento en la escala de tiempos del desplazamiento químico de  $^{19}\text{F}$ , permitiendo así arrojar luz sobre las características termodinámicas y cinéticas del equilibrio.

## Capítulo III.

De los monosacáridos a los disacáridos. En este capítulo se discute el origen de la flexibilidad del enlace glicosídico. En concreto, el uso de miméticos moleculares proporciona resultados clave en el diseño de glicanos sintéticos, y permite modular las componentes estereoelectrónicas relevantes para definir la distribución de poblaciones para los posibles conformeros. Mediante el uso de métodos de RMN y de cálculos teóricos, se ha demostrado que es posible restablecer el efecto anomérico de un acetal cuando se sustituye uno de los átomos de oxígeno por un grupo  $\text{CF}_2$ .

## Capítulo IV.

Receptores de glicanos. La relación que existe entre la dinámica de proteínas y sus funciones está todavía poco caracterizada desde el punto de vista energético. En este capítulo hemos utilizado GGBP, una proteína de unión a glucosa y galactosa, como modelo para definir los movimientos conformacionales asociados a su función, tanto desde el punto de vista estructural como el energético. Para llevar a cabo este estudio hemos usado un enfoque innovador que combina los experimentos de RMN de compuestos paramagnéticos con simulaciones de Dinámica Molecular. Nuestros resultados demuestran que la proteína en su estado libre coexiste entre conformaciones abiertas y cerrada, con una velocidad de intercambio de 25 ns. A pesar de esta heterogeneidad conformacional, la presencia del ligando provee el *ímpetu* para desplazar el equilibrio hacia la forma cerrada. Estos resultados demuestran que, en algunos casos, el evento de reconocimiento molecular no se puede describir como un proceso puro de tipo “induced fit” o “conformational selection”.

## Capítulo V.

Las interacciones tipo  $\text{CH}/\pi$  aportan importantes contribuciones a los procesos de reconocimiento molecular. En complejos de glicano-proteína, estas interacciones débiles implican un sistema  $\pi$  de un residuo aromático de la proteína, como aceptor, y un grupo polar

CH en el azúcar, como donante. A pesar de su amplia difusión, caracterizada por estudios estructurales de Rayos X y de RMN, no existe evidencia experimental directa de esta interacción. Nosotros proponemos que la RMN podría proporcionar evidencias directas de la existencia de un acoplamiento débil *via* una constante de acoplamiento intermolecular J. En este capítulo se discuten los resultados preliminares dedicados al diseño de un sistema de etiquetas isotópicas apropiadas para mejorar la sensibilidad experimental necesaria a la detección directa.

#### Capítulo VI.

Aquí se presentan las conclusiones generales procedentes del trabajo que se ha realizado durante este programa de estudios de doctorado.







## Abstract

Glycans are among the most varied and complex molecules in biological systems. The different branches of the tree of life could be differentiated on the basis of the glycan composition of the own glycoconjugate molecules. However, how much we already know about glycans and glycoconjugate function and distribution is still an open question. Not so many years ago our knowledge about protein N-glycosylation was considerably scarce. In fact, while protein N-glycosylation was once believed to be limited to eukaryotes, it is now firmly established that this complex modification also occurs in bacteria and archaea. Consequently, in the past 10 years, the field of protein glycosylation has witnessed enormous strides in the discovery of new and unusual carbohydrates, in the elucidation of the enzymes involved in glycan assembly and processing, and in the understanding the biological impact that these glycan modifications have on the structure and function of target protein. The reason for this “late” discovery probably lies in the intrinsic structural complexity, heterogeneity and flexibility of glycans. As counterweight, numerous and exhaustive works in glycomics have demonstrated that it is due to their structural complexity, heterogeneity and flexibility why glycans have been selected as key intermediates for cell proliferation, differentiation, adhesion, infection, communication, etc.

With this thesis we have tried to look inside into glycan structure, with the aim to reconcile their structural features at the atomic level with the reasons of their molecular flexibility at a more complex scale. When glycans are recognized by their receptors, their intrinsic flexibility and the plasticity of the whole system has enormous effects in the molecular recognition phenomenon. In fact, both partners involved in the intermolecular interaction could adapt their contact surface in a way that enhances enthalpy-based favourable intermolecular interactions. Alternatively or simultaneously, the glycan and the receptor could strategically keep internal molecular motions, even in the bound state, in a way that minimizes the entropy penalty to the binding event. As consequence, the role of enthalpic/entropic compensation is not easy to predict and even, to assess.

Along this thesis we have explored these features, focusing our attention on sugar protein interactions, starting from the sugar flexibility at the monosaccharide level, passing then to the study of disaccharides, and later investigating the complex motions within a sugar receptor. Finally, CH/ $\pi$  intermolecular interactions, which essentially contribute to the stability of sugar-protein complexes, have also been discussed and a new strategy for their direct detection has been proposed.

### Chapter I.

The introduction of this thesis is focused on the sugar code, how key biological information is stored into glycan structure and how this message is read and is translated by specific glycan receptors. The use of stable and useful glycomimetics with bio-medical and spectroscopic interest is also presented. The chapter concludes with a description of the experimental methods that have been employed in this thesis to elucidate glycan structure and interactions.

### Chapter II.



In this chapter, we discuss sugar flexibility at the monosaccharide level. Herein, a new generation of fluorine-containing glycomimetics is presented. We have applied a combination of NMR spectroscopy and computational methods to investigate the conformational behaviour of idose- and glucose-like rings. We have used low-temperature NMR experiments to slow down the conformational exchange of the six-membered rings. Under these conditions, the exchange rate becomes slow in the  $^{19}\text{F}$  NMR chemical shift time scale and allows shedding light on the thermodynamic and kinetic features of the dynamic equilibrium.

### Chapter III.

From the monosaccharide to the disaccharide level. In this chapter, we discuss the origin of glycosidic linkage flexibility. Specifically, the use of molecular mimicry has provided key findings in glycan engineering and on how to modulate relevant stereoelectronic components that define the population distribution for the possible glycoside conformations. By using NMR methods and theoretical calculations, we have shown that it is possible to restore the anomeric effect for an acetal when replacing one of the oxygen atoms by a  $\text{CF}_2$  group.

### Chapter IV.

A flexible paradigmatic glycan receptor. Protein dynamics related to function are still ill-defined energetically. In this chapter, we have used GGBP, the glucose-galactose binding protein, as a model to define a functional conformational landscape, both structurally and energetically, by using an innovative approach that combines paramagnetic NMR experiments and MD simulations. Our results demonstrate that the protein, in its unbound state, coexists between open-like and closed-like conformations, with an exchange rate around 25 ns. However, despite such conformational heterogeneity, the presence of the ligand is the ultimate driving force to unbalance the equilibrium toward the closed form. These results demonstrate that, even in this case, ligand recognition event cannot be described by pure “induced fit” or pure “conformational selection” modes.

### Chapter V.

$\text{CH}/\pi$  interactions are essential for many intermolecular recognition processes. In glycan-protein complexes, these weak interactions involve the  $\pi$ -system on an aromatic residue on the protein, as acceptor, and the polarized CH moiety on the sugar, as donor. Despite their common spread, which has been identified from the detailed X-ray or NMR-based structural analysis of the complex, their direct detection by experimental means at atomic resolution has still been elusive. We herein propose that NMR could provide direct evidences for the existence of a weak intermolecular scalar J coupling between the donor and the acceptor. In this chapter, we discuss preliminary results devoted to the design of an appropriate isotope labelling scheme that could improve the necessary experimental sensitivity.

### Chapter VI.

The general conclusions arising from the different investigations that have been performed during this PhD student program are here presented.





## Route map of this thesis

This section provides a comprehensive link between the chapters presented in this thesis. Sugar/protein interactions arise from the delicate interplay between structure, molecular recognition features, and dynamics. Both partners involved in the recognition processes are flexible molecules. Therefore, the effectiveness and specificity of the resulting biochemical response is associated to the plasticity of binding. Respect to the classical lock and key molecular recognition model, nowadays a more complex adaptive model is largely endorsed, especially when sugars or sugar mimics act as ligands. Actually, the sugar code is significantly enriched by the intrinsic flexibility of monosaccharide rings or at higher complexity level, by the flexibility of the interglycosidic linkages. Usually, only few of all possible conformers are recognized by the proper receptor. Thus, the access to these bio-relevant structures and the possibility of unravelling the origin of sugars flexibility is of paramount importance. On the other hand, segmental dynamics in proteins are also functional and also must be taken into account to get a full picture of the ligand recognition process. In fact, in flexible proteins, one single structural model unsatisfactory describes the vast conformational space that these systems could cover in solution. The existence of induced-fit or conformational selection processes has been largely discussed.

However, studies on structure-activity relationships are only rarely complemented by a detailed analysis of dynamics related to function. In this scenario, NMR assisted by computational chemistry has the privilege to uncover the conformational and recognition features associated to sugar/protein recognition. NMR provides high-resolution structural information and is almost the only bio-physic technique that allows exploring the dynamic behaviour at different time scales of biological systems in their proper environment, in solution. However, the interpretation of the NMR data is elusive because of the involved time scales and amplitude of molecular motions and averaged observations. Thus, when the desired data are experimentally ill-defined, molecular modeling methods are essential to gain insights at the atomistic scale. In this thesis, we have combined a thorough experimental approach with computational techniques at different levels of complexity, including molecular mechanics (MM), molecular dynamics (MD), and quantum mechanics (QM), to characterize the sugar and protein dynamics. The first chapter of this thesis contains an overview about all the employed methods. Specifically, an introduction on carbohydrates, from monosaccharides to oligosaccharides, their implications as bio-relevant molecules in cell processes, and the stereoelectronic properties that are behind the origin of molecule flexibility have been addressed. Following this introduction, glycomimetics with bio-medical and spectroscopic interest are presented. In particular, the advantage of using  $^{19}\text{F}$

labelled sugar mimics is discussed. Further, an interesting sugar receptor is presented as model system. This protein has been used as model for describing the complex dynamics behavior of receptors, since the particular recognition process not only involves the typical side chain adjustments, but also large scale structure rearrangements. This chapter concludes with a brief description of the NMR spectroscopic and computational techniques used in this work. Chapter II focuses on sugar function, structure and dynamics. In this chapter, detailed studies on flexibility in monosaccharides and mimetics thereof are presented, its consequences in modulating the interaction with biological receptors, as well as approaches to extract the energy values associated to conformer interconversion. To achieve this goal, a new generation of fluorine-containing glycomimetics is presented. A combined approach of NMR and computational methods at QM level has been used to investigate the conformational behaviour of idose- and glucose-like glycomimetics. Chapter III deals with the understanding of the features governing glycoside conformation in oligosaccharides. The relative orientation of a monosaccharide unit respect to its vicinal one in a disaccharide moiety is strongly determined by the anomeric effects, despite that, its origin is still the subject of virulent scientific debates. Key findings are provided in the relevance of the stereoelectronic component of the anomeric effect, by demonstrating that  $\text{CF}_2$  sugar analogues are able to adopt the natural glycoside conformation, providing new avenues for sugar-based drug design. Chapter IV gathers the characterization of interdomain dynamics in a sugar receptor system, the glucose-galactose binding protein (GGBP). An iterative approach using experimental NMR restraints, combining pseudo contact shifts (PCSs) and residual dipolar couplings (RDCs) with molecular dynamics simulations has allowed the complete description of the energetic landscape associated to this highly dynamic protein in solution and their molecular recognition events. The energy barrier associated with open-to-closed transitions has been estimated using non equilibrium molecular dynamic calculations. Finally, chapter V refers to ongoing work where one of the key intermolecular interactions that contribute to the stability of sugar/protein complexes could be directly elucidated via NMR methods.  $\text{CH}/\pi$  interactions can be defined as a non-conventional hydrogen-bond interaction in which the  $\pi$ -electron cloud on the aromatic residue acts as an acceptor and the CH group on the sugar as a donor. We propose that NMR could be used to directly detect glucose/tryptophan interactions in the GGBP/Glc system. Though each chapter of this thesis presents partial conclusions regarding the different topics, chapter VI gathers the main conclusions arising from the particular topics.

# CHAPTER I

**1.1 Glycomics:** The study of carbohydrates and carbohydrates containing biomolecules their receptors and their functions.

## **1.1.1 Language** The sugar code

The human origin has a sweet origin. Actually, mammalian sperm-egg binding is primarily mediated by the interaction of the egg-binding protein (EBP) on the sperm plasma membrane with carbohydrate sequences expressed on glycoproteins of the egg's zona pellucida. Evidence that carbohydrate recognition plays a major role in human fertilization was disclosed when the polysaccharide fucoidan was shown to potently block this interaction.<sup>[1]</sup> The most abundant carbohydrate sequence responsible for this recognition process has been identified to be a well-known selectin ligand, the sialyl-Lewis<sup>x</sup> sequence.<sup>[2]</sup>

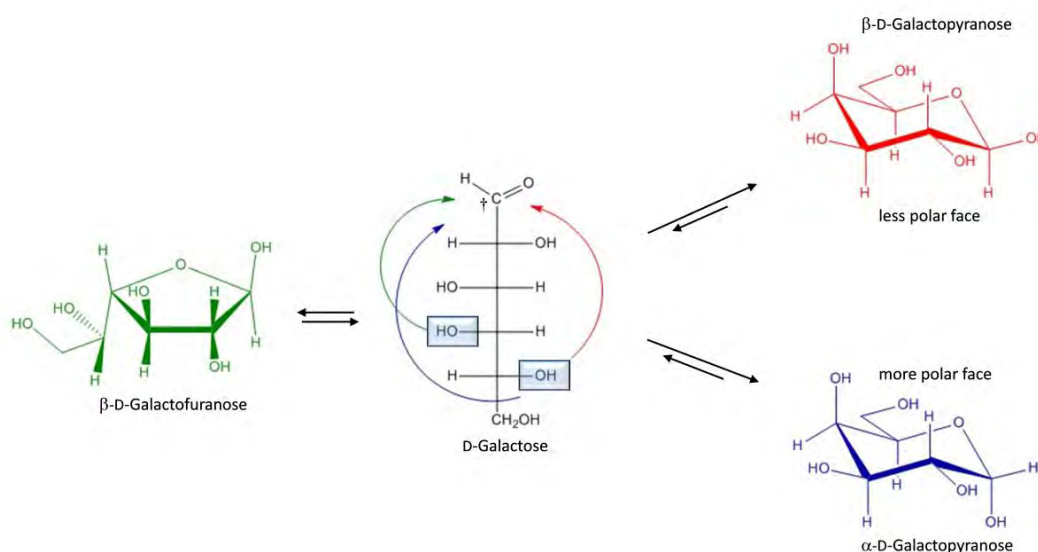
Conversely, aberrant glycosylation is a hallmark of cancer cells. The Tn antigen ({-O-GalNAc-Ser/Thr) is one of the most specific human tumor-associate structures. This chemical moiety is expressed in approximately 90% of carcinomas, and a direct correlation between the aggressiveness of the carcinoma and the occurrence of the antigen has been observed.<sup>[3]</sup> As a consequence, the Tn antigen has found useful application as biomarker and as a potential therapeutic target against cancer.<sup>[4,5,6,7,8]</sup> These two examples show how glycans conjugate to cognate biomolecules spread in almost all aspects of life as reflected by the term "sugar code".<sup>[9]</sup>

Generally speaking, molecular recognition processes are the pillars of bio-chemical communication. Subtle differences in chemical substitution, geometry and structure in the molecular partners guarantee the effectiveness and specificity of the resulting response. Of all the natural bio-molecule, glycans are unequalled in their capacity for biological coding and signal translation into cellular message.<sup>[10,11,12,13]</sup> The origin of this "power" arises in the tremendous structural variability of glycans with respect to chemical singularity, conformation, sequence and shape.<sup>[14]</sup> Thus, glycosylated molecules carry diverse information in their glycan chains, encoded at different levels of complexity.<sup>[10]</sup> Glycoconjugates, including glycoprotein, glycolipids, proteoglycans and free complex carbohydrates<sup>[15]</sup> play a relevant role in molecular recognition processes by establishing specific interactions that trigger intra- and extra-cellular physiological and pathological events. In fact, they are, very often, the primary contact points in cell/cell or cell/host pathogen interaction. In this context, knowledge inside the chemical features responsible for triggering biological responses is of paramount importance. The main goal of

glycoscience is to decipher the sugar code in order to open wide new avenues for therapeutic applications.

### 1.1.2 Alphabet Monosaccharides

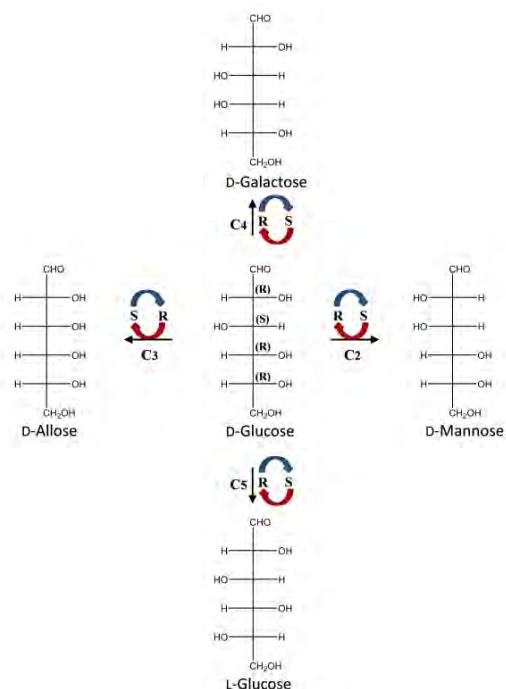
Monosaccharides are the building blocks of complex carbohydrates. The word “carbohydrate” means hydrate of carbon. Thus, formally a monosaccharide is composed by carbon atoms and water molecules. The high solubility of sugars is easily inferred; however, it would be incorrect to say that carbohydrates are exclusively hydrophilic molecules. Actually the carbon chain, usually from five to seven carbon atoms, cyclizes to form rings of different size, e. g. the hexoses (six carbon atoms) can cyclize either to six-membered pyranose or to five-membered furanose rings.<sup>[16]</sup> Fig. 1.1. Depending on the spatial orientation of each of the hydroxyl group in the sugar ring, a highly polar and a less polar face of the molecule are defined. In fact, monosaccharides and, in general, carbohydrates are amphiphilic molecule, with a dual character. In other words, carbohydrates, even at monosaccharide level, are water soluble molecules that give rise to hydrophilic as well as hydrophobic interactions with the right molecular partner.



**Figure 1.1** Structural variability of a monosaccharide (galactose, Gal). Schematic representation of D-Gal in its open-chain form, { and | anomeric cyclic pyranose forms and | anomeric cyclic furanose form. The prochiral planar  $sp^2$ -carbonyl carbon atom is labeled with the symbol †. Once cyclized, a high polar and a less polar face of the sugar are defined. The majority of the CH groups of the Gal ring point in the same direction, so defining the less polar face.

The types of monosaccharides used for building up oligo- and poly-saccharides or glycoconjugates are not exceedingly large. Commonly, around a dozen of monosaccharides occur as constituent of glycosides in higher organisms. However, the bacteria kingdom is rich in rare carbohydrates that are biomarkers of bacteria species or families, as Kdo for Gram negative bacteria.<sup>[17,18]</sup> The chemical difference between two types of sugars, e.g. glucose (Glc) or mannose (Man), resides in the absolute configuration at each chiral center of the carbon backbone. Starting from a linear hexose as reference, which contains four chiral centers, we can generate four different types of sugars simply by changing the absolute configuration of only one carbon atom. Fig. 1.2. shows the variability obtained by changing the configuration of more than one chiral center.

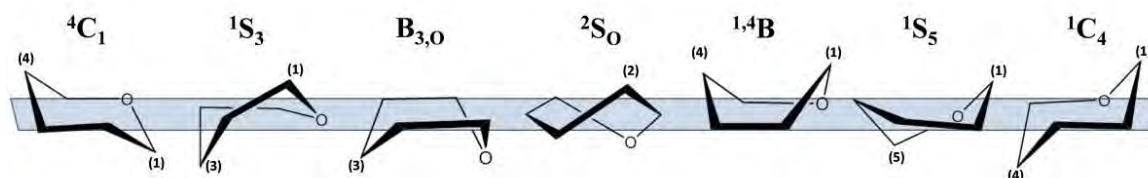




**Figure 1.2** Schematic representation of D-Glc and its epimers for the four chiral carbon atoms. The absolute configuration is labeled with the (R) or (S) descriptors. The possible permutations by changing the absolute configuration of only one stereocenter are illustrated. Not all of them are relevant in nature. Actually, L-Glc, idose, does not occur naturally, but can be synthesized.

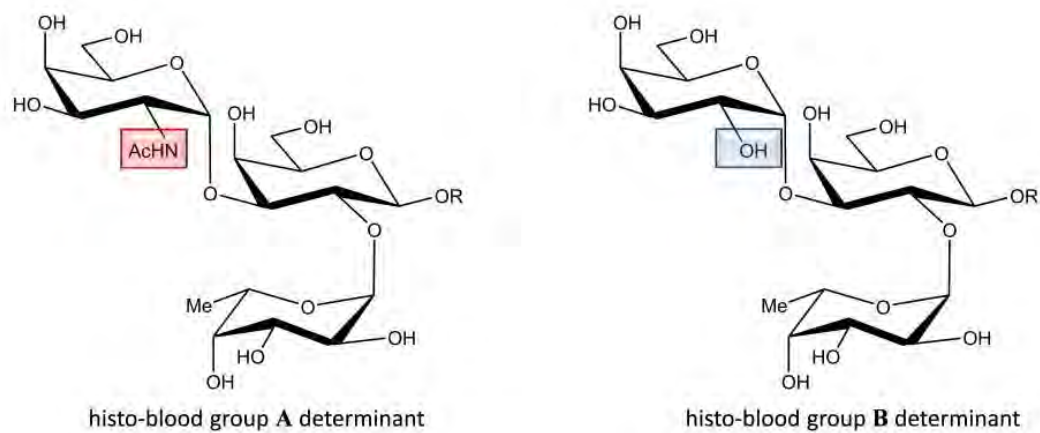
During ring formation, due to the conversion of the planar  $sp^2$ -carbonyl carbon atom into a chiral tetrahedral  $sp^3$  carbon, the glycosidic hydroxyl group can acquire two configurations, called { and | anomers. The difference resides in the orientation of the hydroxyl group in the anomeric position (C1) that can be either axial (|) or equatorial (|) with respect to the plane containing the sugar ring. However, the consequences are significant in terms of biochemical interactions; anomer selectivity has been studied in enzymes involved in carbohydrate metabolism using kinetic approaches,<sup>[19]</sup> while, more recently it has been shown a robust NMR based method to measure anomer selectivity in the ConcanavalinA/Man complex without requirement for chemical modification.<sup>[20]</sup> Furthermore, the variability of the sugar code is increased by the intrinsic structural plasticity. Once cyclized, the sugar ring experiences conformational flexibility. They may acquire different 3D ring conformations in fast equilibrium among them. The absolute configuration and the consequent spatial orientation of the hydroxymethyl and hydroxyl groups determine the population distribution of the different geometries that may take place.<sup>[21,22]</sup> In the case of | -Glc, the  ${}^4C_1$  chair conformation ensures that all the hydroxyl groups are in equatorial disposition, since this 3D arrangement minimizes the steric clashes that characterize the  ${}^1C_4$  conformation. In this case, the 1,3-diaxial hydroxyl groups are oriented in the same direction, with the consequent steric and dipolar repulsion. Due to this | -Glc exists almost exclusively in the  ${}^4C_1$  conformation. Fig. 1.3. However, other high energy conformations have

mechanistic implications in sugar/receptor recognition processes, as in glycoside hydrolysis. In particular, the distortion of the  $\beta$ -Glc sugar ring in the Michaelis complexes of several Glc/Glycoside hydrolases (GHs) determines its efficiency in the degradation of polysaccharides.<sup>[23,24]</sup>



**Figure 1.3** Ring conformations adopted by monosaccharide rings. The energetically favored chair conformations are labeled as  ${}^4C_1$  or  ${}^1C_4$ , depending on the position of carbons 1 and 4 (above or below the sugar ring plane, as noted by the position of the numeral). The other conformations, as Skew boat (labeled as S) or boat (labeled as B) are adopted by sugar rings when recognized by different glycoside hydrolases or at Michaelis complexes. However, these conformations for some sugars can be adopted also in absence of enzymes, depending on the ring substituents and configuration of the stereocenters, as for  $\beta$ -L-Idopyranose (Ido) in the  ${}^2S_0$  geometry, or as in the case of xylose.

Furthermore, chemical modification of the hydroxyl groups, e.g. acylation, sulfation, methylation and phosphorylation, contributes to glycan differentiation. More than 40 modifications are known in gangliosides (glicolipids of cellular membranes). Their biological importance is well illustrated by comparing the histo-blood group A and B determinants. Fig. 1.4. The chemical difference resides in the presence or absence of the NHAc substituent at position 2 of the terminal galactose unit of the glycoconjugate. The medical complications related to the transfusion of blood belonging to the wrong group are well known, demonstrating how subtle the carbohydrate/antibody interactions are.



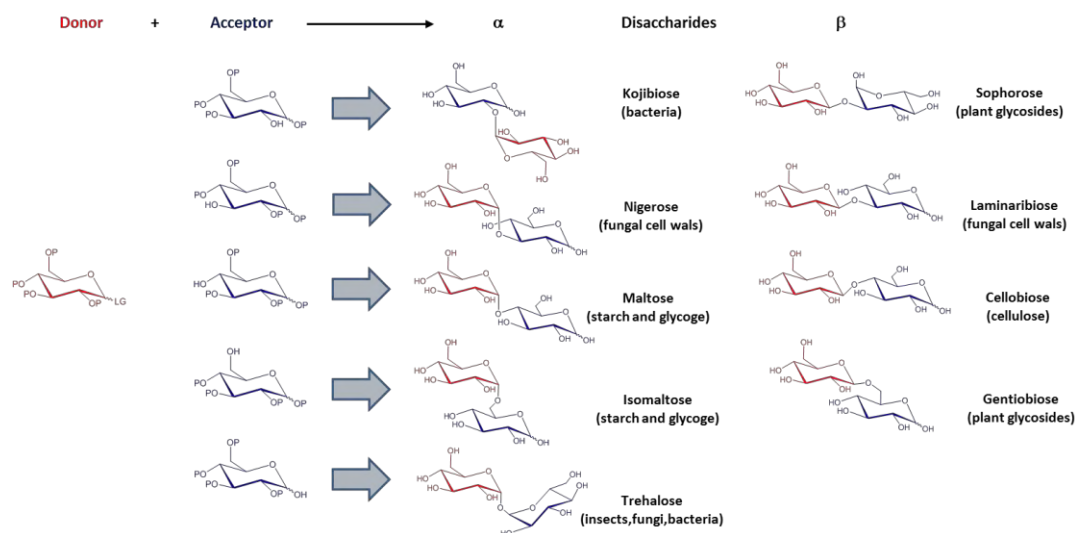
**Figure 1.4** A- and B-type trisaccharides in histo-blood group determinants. The extension at the anomeric position of the branched Gal moiety is indicated by R. The sole structural difference resides in the highlighted 2'-position of the Gal/GalNAc extension.

Summarizing, monosaccharides are the letters composing the alphabet that nature uses to build words and sentences in the sugar code. Even if, apparently, this alphabet is not as rich as that of protein, different ring size, mutarotation, ring flexibility and chemical modifications contribute to enrich the sugar alphabet of an enormous number of letters. With this potential, it appears clear that sugars provide the cell with a huge structural variability that is translated into an understandable and efficient communication.

### 1.1.3 Vocabulary Oligosaccharides

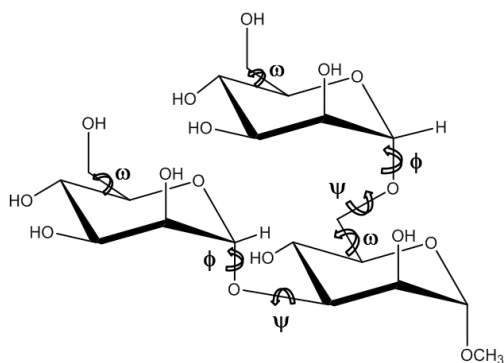
A disaccharide originates from the condensation reaction (glycosylation) between two monosaccharides. The reaction involves the so-called “glycosidic” hydroxyl group of one monosaccharide moiety (glycosyl donor) and one attachment point of the other monosaccharide (glycosyl acceptor), which can be any of the available hydroxyl groups, including the glycosidic function itself. Since the formed glycosidic bond can acquire two configurations, { and | , and due to the availability of several attaching points the glycosylation reaction between two glucose units, if not enzymatically driven, can generate 11 different disaccharides. Fig. 1.5. This potential for permutations on the level of linkage points and anomeric configuration is translated into physical-chemical property permutations. In between the eleven possible dimers of glucose,

nearly all occur in nature and their function is surprisingly various<sup>[25]</sup> and in some of the cases, their functions have been only very recently proposed.<sup>[26]</sup>



**Figure 1.5** The building blocks, donor/acceptor pairs, in disaccharides synthesis. Between all possible combinations in the glycosylation reaction between two glucose units, nearly all occurs in nature. The occurrence is specified in bracket.

When introducing another building block to one disaccharide moiety, the possibility to extend the chain competes with the alternative to branch it, since other hydroxyl groups on the original glycosyl acceptor are still available. Precisely, the ramification strongly contributes to the spatial organization of the molecule that is translated into a different biochemical message, converting the branched oligosaccharide into a different structure respect to the linear one. The conformations around the glycosidic linkage are described by the torsion angles  $\phi$ ,  $\psi$  and  $\omega$  where  $\phi$  is defined as  $(H1_{(i)}-C1_{(i)}-O1_{(i)}-C_{x(i-1)})$ ,  $\psi$  as  $(C1_{(i)}-O1_{(i)}-C_{x(i-1)}-H_{x(i-1)})$  and  $\omega_{O5}$  as  $(O5_{(i)}-C5_{(i)}-C6_{(i)}-O6_{(i)})$ ;  $\omega_{C4}$  as  $(C4_{(i)}-C5_{(i)}-C6_{(i)}-O6_{(i)})$ .<sup>[27]</sup> Fig. 1.6.



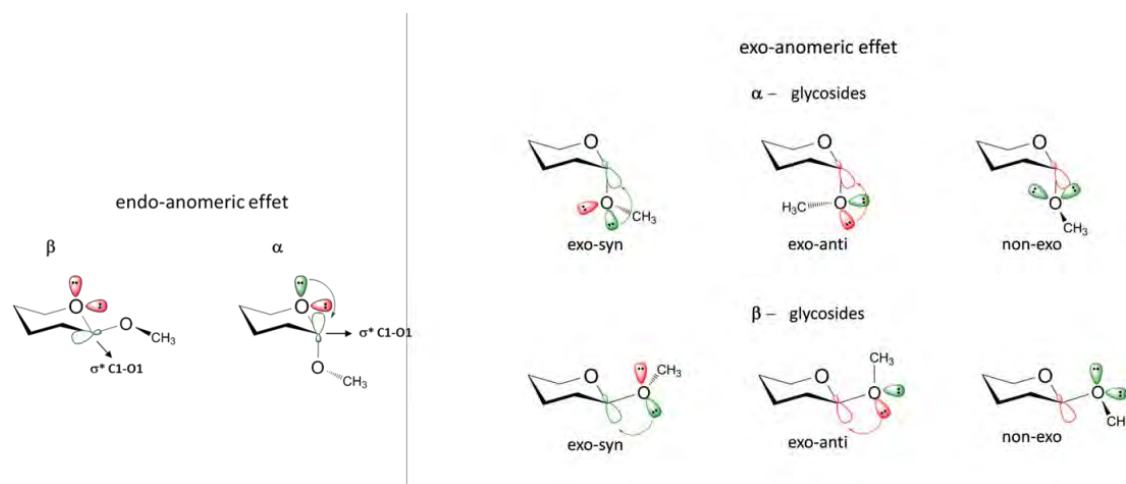
**Figure 1.6** Methyl 3,6-di-O-({-D-mannopyranosyl)-{-D-mannopyranoside. This sugar is ubiquitous in the inner core of N-glycans. Glycosidic angles are  $\phi$ ,  $\psi$  and  $\omega$ .

Again, the flexibility around the glycosidic linkage introduces a new level of complexity in carbohydrates. Of course, not all the possible geometries are equally populated, and some of them are energetically unfavorable. Specifically, the  $\psi$  angle is confined by steric and electronic clashes, while  $\phi$  is mainly defined by the *exo*-anomeric effect.

#### 1.1.4 The anomeric effect

The origin of the anomeric effect is still subject to virulent scientific debate.<sup>[28,29]</sup> Nevertheless, it plausibly lies in the favorable interaction between a lone electron pair located in a molecular orbital ( $n$ ) on either glycoside oxygen atoms and the vicinal anti-bonding molecular orbital ( $\sigma^*$ ). Thanks to the contribution of Lemieux, it is mandatory to distinguish between the *exo*-anomeric effect and the *endo*-anomeric effect.<sup>[30]</sup> The *endo*-anomeric effect arises from the favorable electronic interaction between the electronic lone pair on the *endo*-cyclic oxygen atom O5 and the periplanar anti-bonding molecular orbital ( $\sigma^*$ ) of the C1-O1 bond. The donation of electron density from the ring oxygen is possible only if the acceptor empty molecular orbital is geometrically properly oriented. This is the case of  $\{$  glycosides, but not of  $|$  glycosides. As a result, in  $\{$  glycosides the O5-C1 bond is slightly short if compared to the O5-C5 bond, because of its partial double bond character, while the C1-O1 bond is slightly longer. In contrast, the *exo*-anomeric effect is common to both  $\{$  and  $|$  glycosides. The anti-parallel arrangement of one lone pair on the *exo*-cyclic oxygen with respect to the anti-bonding O5-C1 bond may allow the electron back donation, thus stabilizing those conformations for which the values of  $\phi$  are consistent with this geometric requirement. More precisely, this interaction is more favorable than the interaction of a lone pair orbital of the same oxygen with the C1-C2  $\sigma^*$ . This difference is attributed to the difference in polarization between the C1-O5 and C1-C2 bonds, producing a larger  $\sigma^*$  orbital centered on the less electronegative atom of the polarized bond. This remark is supported by the observation that the anomeric effect increases with the electronegativity of

the substituent and decreases in solvents of high dielectric constant. As result of the *exo*-anomeric effect, the C1-O1 bond is shorter than the other C-O bonds. Fig. 1.7.



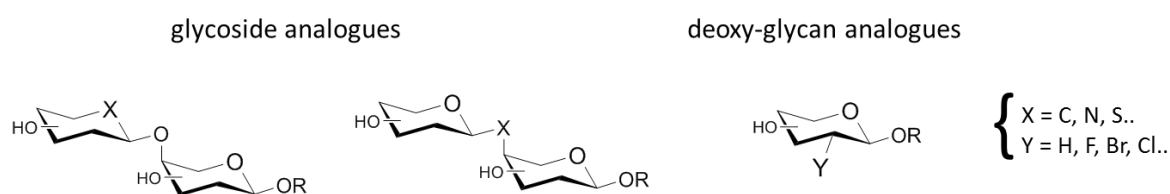
**Figure 1.7** Schematic representation of the lone pair- $\sigma^*$  interactions responsible for the *endo*- (right) or *exo*-anomeric (left) effects in { - and | - glycosides. The *endo*-anomeric effect exists only in { -glycosides due to the geometrical requirements that allow the molecular orbital overlapping. For the *exo*-anomeric effect, different conformers are represented, both for { - and | -glycosides. However, the overlapping between the  $\sigma^*$  and the lone electron pairs is possible only for two of the three conformations. The fill color of the orbitals does not refer to the orbital phase, but it is a schematic representation for full or empty orbitals.

An important consequence concerns the molecular dynamics features. Actually, due to the anomeric effect, some conformations around the glycosidic bond are negligible. As mentioned, even if the phenomenon is experimentally observable, the origins of the anomeric effect are still debated. In order to shed light on the stereo-electronic contributions behind the anomeric effect, the use of glycomimetics that lack of the chemical moiety responsible for the suggested orbital stabilization has been proposed.<sup>[31,32,33,34,35]</sup>

## 1.2 Glycomimetics

Sugar mimicry is an essential part of the development of carbohydrate-based therapeutics, an area that has already proved successful with molecules such as Miglustat, Acarbose or Voglibose.<sup>[36]</sup> The mimics are sugar scaffolds bearing a relatively minimal modification that changes its properties while still resembling the natural sugar. These modifications can be

divided into two main groups; the first includes sugar mimics where the glycosidic linkage has been modified,<sup>[37]</sup> while the second group includes sugar mimics in which the modified moiety is one or more of the pendant hydroxyl groups.<sup>[38]</sup> In the first group, when the aglycone is another glycoside, the replacement of the exocyclic-anomeric oxygen atom by another atom type, as carbon, nitrogen or Sulphur, leads to C-, N-, or S-glycoside<sup>[16,39,40,41]</sup> while the replacement of the endocyclic oxygen produces carba-, amino- and thio-sugar, in many cases these transformations leads to non-enzymatically hydrolysable glycoside analogues.<sup>[42,43]</sup> In the second group, deoxy-glycan analogues have been widely employed to monitor interaction events and to deduce the enthalpy contributions of specific hydroxyl groups of the sugar to the binding processes with different receptors. Fig. 1.8.



**Figure 1.8.** Sugar mimics. On the left, non-hydrolysable glycoside analogues. On the right, deoxy-glycans. The R group can be a methyl moiety or another glycoside.

From a bio-medical view-point, the knowledge about both the chemical and structural factors that are decisive for establishing effective interactions is of great interest for the development of new potential glycan-based drugs.<sup>[44]</sup> Hence, the field of glycosciences has experienced in the last years a blossom in the research to reveal the structure-activity relationships (SARs) that govern the effectiveness of carbohydrate-protein interactions.<sup>[45,46]</sup> Thus, different glycomimetics have been reported that compete with their natural counterparts for the same receptors or enzymes acting, consequently, as molecular probes or enzyme inhibitors.<sup>[47]</sup> In this context, the use of non endogenous atoms as molecular spies generates great expectation for clarify the binding mode of oligosaccharide to the proper receptor. Between those, fluorine seems to be the best candidate for the innocuous substitution of endogenous nuclei in saccharides.

### 1.2.1 <sup>19</sup>F as spy atom

Fluorine is not an endogenous nucleus, but medicinal chemistry studies have demonstrated its useful introduction in bioactive substances to improve their pharmacokinetics properties and to modulate its biological properties.<sup>[48,49,50,51]</sup> From the NMR perspective, F-19 NMR provides an attractive alternative. This is due to both the lack of F-containing molecules in biomatrixes<sup>[52,53,54,55]</sup> and the high intrinsic sensitivity of <sup>19</sup>F, a ½ spin nuclei with 100% natural abundance and a gyromagnetic ratio only slightly smaller than that of proton.<sup>[56]</sup> In addition, the F-19 chemical shift is highly sensitive to even subtle changes in the magnetic environment, making it an ideal probe for distinguishing closely related molecules, as saccharides, in their free or bound states.<sup>[57,58]</sup> The presence of <sup>19</sup>F atoms in sugar rings may also deliver important conformational and structural information, complementary to that provided by <sup>1</sup>H and <sup>13</sup>C nuclei.<sup>[59]</sup> For all these reasons, fluorinated glycomimetics have received careful attention.<sup>[37]</sup> Depending on the substituted position, the fluorine substituent can have a remarkable effect upon the physical and chemical properties of the molecule.<sup>[60]</sup> It could induce increase of lipophilicity,<sup>[61,38]</sup> decrease in pKa values of certain groups by OH-F electrostatic interactions,<sup>[62]</sup> modulate the hydrogen bond acceptor/donor ability,<sup>[62,63]</sup> or foster the presence of a particular ring conformation.<sup>[64]</sup> The functional/structural outcome of fluorine substitution should be explored.

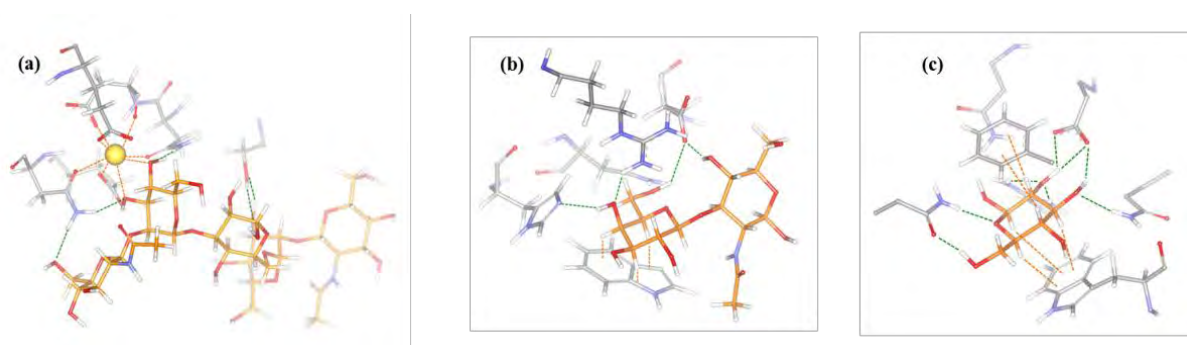
## 1.3 Lectins

### 1.3.1 *Interlocutor* Sugar Receptor

Carbohydrates are recognized by a wide pool of proteins that include lectins, antibodies and enzymes. Enzymes are catalytic proteins responsible for synthesis or degradation of carbohydrates and for the transfer of oligo- and polysaccharides to the cognate biological scaffolds. Actually, the synthesis of glycoproteins involves the action of multisubunit oligosaccharide transferase complexes, while the glycan maturation processes happens with a concerted mechanism of glycosidases and glycosyl transferases responsible for the high structural variability of glycans. For example, in mammals, the repertoire of glycan structures biosynthesized by a given cell is largely based on the expression of a subset of more than 200 glycosidases and glycosyl transferases found in the mammalian genome.<sup>[65]</sup> Anti-glycan antibodies are generated by the immune system and play a key role in the defense against pathogens, based on their ability to distinguish self-antigens from pathogen-derived antigens.<sup>[66]</sup>



Lectins are non-catalytic proteins with exquisite target specificity for carbohydrates and therefore they are being exploited as biomarkers and valuable models for the study of protein-carbohydrate interactions, in either free form or attached to lipids or proteins.<sup>[67]</sup> Both glycan and lectin structures can reach equal levels of sophistication. Fig. 1.9 illustrates two examples of carbohydrate recognition domains (CRD) of mammalian lectins showing the diversity of possible recognition modes. C-type lectins strategically host one  $\text{Ca}^{2+}$  ion in the CRD that acts as a linking point in between the ligand and negatively charged aminoacids in the binding pocket.<sup>[68]</sup> On the other hand, Galectin-1 traps the ligand, N-acetyllactosamine, by stacking the less polar face of the galactose ring against the aromatic residue tryptophan while the side-chains of conserved residues provide a hydrogen bond network with the polar groups of the sugar. These differences in recognition mode are at the origin of lectin's differentiation/classification and are the reason why lectins are considered the favorite partners of glycans. The best way to decipher the sugar code lies in the elucidation, at atomic level, of the ligand/receptor interactions that contribute to the high specificity and accuracy of bio-chemical communication. As further example of carbohydrate binding proteins, those belonging to the family of Periplasmic Binding Proteins selectively recognize and actively transport small glycans across the inner membrane in gram negative bacteria.<sup>[69]</sup> Respect to lectins, the affinity of this sensor/transport proteins for the substrate is significantly higher ( $K_D \approx \text{nM}$ ). The sugar binding site is here located in between two globular domains that ensure a tight hydrogen bond network surrounding the ligand, while two aromatic residues fix the ligand into a "sandwich" fashion.<sup>[70]</sup> Ligand binding provides the impetus to drive the protein conformational equilibrium to the bound closed geometry.



**Figure 1.9.** Ligand selection by three types of sugar binding proteins. (a) Interactions of the C-type lectin DC-SIGN with  $\text{GlcNAc}_2\text{Man}_3$ .  $\text{Ca}^{2+}$  coordination or  $\text{CH}/\pi$  interactions are shown as dashed orange lines, while hydrogen bonds as green lines. The calcium ion in the CRD acts as bridge between the ligand, OH(3) and OH(4) of the Man moiety, and the protein residues, (N377, E366, D378, E359) in the binding pocket. [PDB 1K9J]. (b) Interactions of human galectin 1 with N-Acetyllactosamine. Hydroxyl groups four and six of

the Gal moiety engage in cooperative donor/acceptor hydrogen bonds, while the less polar face of the same ring stacks against the aromatic residue W68. N-AcetylGlc participates with additional contacts. [PDB 4XBL]. (c) Binding pocket of GGBP. The selection of either epimer, Glc or Gal, is justified by the fact that the hydroxyl group in position four does not interact with the protein residues. Conversely, all the OHs are involved into a dense hydrogen bond network with residues on both protein domains. The two faces of the sugar are stacked against aromatic residues, W183 and F16. [PDB 2FVY].

---

In conclusion, sugar receptors are supplied with the chemical features necessary to modulate the specificity and effectiveness of sugar recognition. The huge number of sugar receptors, their biological ubiquity and the possibility to be organized in oligomeric forms and even in clusters, with the consequent regulation of chemical response upon sugar binding, is an additional demonstration of the role that carbohydrates cover in communication processes. Biophysical techniques as X-ray crystallography and Nuclear Magnetic Resonance (NMR) have contributed during the last decades to the structural elucidation of glycan/receptor complexes at atomic level.

## 1.4 NMR

### 1.4.1 *The reader* Nuclear Magnetic Resonance

The fruitful application of NMR in biochemical and biomedical research lies on its sensitivity to chemical structure and molecular dynamics, in addition to the non-invasive and non-destructive nature of the technique. Since the discovery of the NM moment of protons by Stern (Nobel prize in Physics 1943), the development of a method to measure NM moments by Rabi (Nobel prize in Physics 1944), and the NMR experiments performed by Bloch and Purcell (Nobel prize in Physics 1952), the technique has been enriched with numerous continuous advances that have led to a highly active and fertile field of work recognized in several Nobel prizes in other disciplines (Chemistry in 1991 and 2002 and Physiology and Medicine in 2003). In the last few years, NMR has become a powerful tool to monitor molecular interactions and to deduce features of recognition processes at different level of complexity, both from the perspective of the receptor and of the ligand.<sup>[71,72,73, 15]</sup> Nevertheless, there is still an open window for further development of new NMR methods and their applications to unravel molecular recognition problems and transient interaction and states. Actually, advances in chemical protein modification,<sup>[74]</sup> the use of paramagnetic metals as molecular tags,<sup>[75]</sup> the employment of MAS or HR-MAS spectroscopy

methods,<sup>[76,77]</sup> together with new computational tools and the access to larger magnetic fields are examples of new developments that are taking place.

#### 1.4.2 Principles of NMR <sup>[78,79]</sup>

NMR is an intrinsic property of atomic nuclei having a nuclear spin quantum number,  $I$ , different to zero. Those nuclei are defined as “NMR active” nuclei, while those having  $I=0$  are termed “NMR silent”. The spinning nuclei possess angular momentum,  $P$ , and together with its charge and motion it is associated the magnetic moment according to the equation:  $\mu = \gamma P$ , where the term  $\gamma$  is the magnetogyric ratio, which is characteristic for each nucleus and may be viewed as a measure of the magnetic sensitivity of the nucleus. The angular momentum and the magnetic moment are vector quantities, having both magnitude and direction. They are quantized, this means that both the magnitude and orientation can only adopt a discrete number of values. For a spin of magnetic quantum number  $I$  there exist  $2I+1$  possible spin states, taking values from  $-I$  to  $+I$ . For the  $\frac{1}{2}$  spin nuclei, as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{19}\text{F}$  between others, the two states correspond to  $m_s = +1/2$  ( $\uparrow$ ) and  $m_s = -1/2$  ( $\downarrow$ ). In the absence of an external magnetic field, these two levels are degenerate, both states have the same energy, this means that they are equally populated. However, in the presence of an external magnetic field,  $B_0$ , the degeneracy is broken and, with it, the population is redistributed in favor of the lowest energy level, according to the law of Boltzmann.

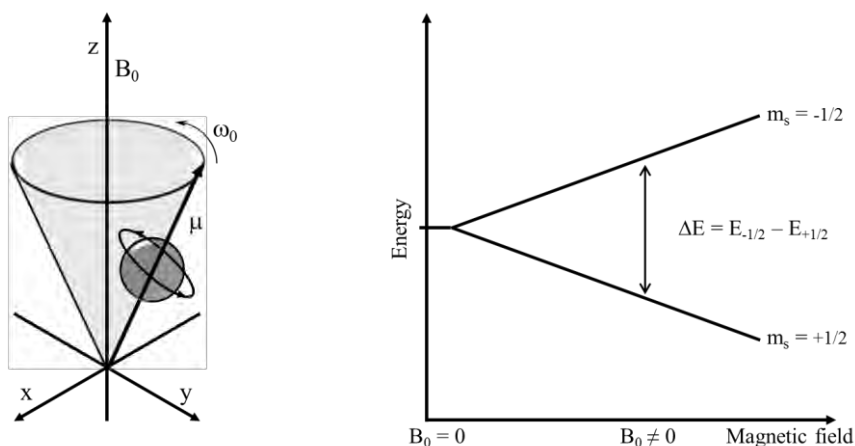
$$\frac{N_{low}}{N_{upp}} = e^{\frac{-\Delta E}{kT}} \quad eq. 1.1$$

Where  $N_{low/upp}$  is the spin state population distribution,  $k$  is the Boltzmann’s constant,  $T$  the temperature and  $\Delta E$  the energy difference between both momentum states.

The difference between the energy level is proportional to the applied magnetic field according to equation 1.2.

$$\Delta E = \gamma \hbar B_0 \quad eq. 1.2$$

Where  $\Delta E$  is the energy difference between both momentum states,  $B_0$  is the external magnetic field,  $\gamma$  is the magnetogyric ratio and  $\hbar$  is the reduced Planck constant or Dirac constant. Fig. 1.10.



**Figure 1.10** On the left, the magnetization vector rotates about the field direction, sweeping out a cone of constant angle to the z-axis; this motion is called precession. On the right, splitting of nuclei spin states in an external magnetic field.

Comprehensibly, dealing with higher magnetic fields will increase the sensitivity of the technique. Actually, the signal in NMR arises from the perturbation of population distribution when a precise energy, corresponding to the energy difference between the two levels, hits the nuclei. That is the resonance condition. It inverts the population of the energetic levels and the signal is produced when the equilibrium distribution is restored. The NMR signal, therefore, is proportional to the population difference between both states. Since this energy difference is relatively small, with respect to others spectroscopic phenomenon such as IR and UV, the NMR is a less sensitive technique. The magnetic moment of each nucleus describes its own precession motion around the magnetic vector  $B_0$ . The frequency of the precession is termed Larmor frequency,  $\omega_0$ , which is the precession rate about the axis  $B_0$ . As expressed in equation 1.3, the frequency of resonance, that coincides with the Larmor frequency value, depends from the magnetogyric ratio, this means from the kind of nucleus, and from the external magnetic field and it is equivalent to the energy difference between the two spin levels:

$$\omega_0 = \gamma B_0 = \Delta E / \hbar \quad \text{eq. 1.3}$$

Where  $\omega_0$  is the Larmor frequency.

Fortunately, each active nucleus of one type (i.e.  $^1\text{H}$ ) in a given molecule experiences different perturbation of its local magnetic field,  $B_i$ , due to the shielding or de-shielding effects produced by nearby atoms or functional groups.

$$\omega_0 = \gamma(B_0 - B_i) \quad \text{eq. 1.4}$$

Where  $B_i$  is the local magnetic field.

As result, nuclei within different chemical environments are characterized by slightly different (chemical shift) Larmor frequencies, which permits to identify them, converting NMR into a powerful tool.

### 1.4.3 NMR for carbohydrates

NMR is one of the most widely used techniques to characterize molecular features of carbohydrates,<sup>[27]</sup> thanks to its versatility, recent advances in resolution and sensitivity<sup>[80]</sup> and to the development of procedures for structural determination, conformation and dynamics of complex oligosaccharides.<sup>[81]</sup> These avenues are associated to the impressive advances in synthesis of carbohydrates and labeled or modified carbohydrates.<sup>[82]</sup> Significantly, NMR techniques provide the possibility to work in solution, in conditions similar to the physiological ones. Of note, carbohydrates are highly flexible molecules and precisely, this flexibility has important consequences on their recognition features. Among the experimental data that can be gathered for the structural determination of sugars, vicinal  $^1\text{H}$ - $^1\text{H}$  coupling constants,  $^3J_{\text{HH}}$ , are particularly useful to determine the relative orientation of the coupled nucleus. Nuclear Overhauser enhancement spectroscopy (NOESY) is employed for structural characterization in general and is a key parameter for the conformational analyses of the glycosidic linkage. Finally, paramagnetic data are used to obtain longer-range restraints.

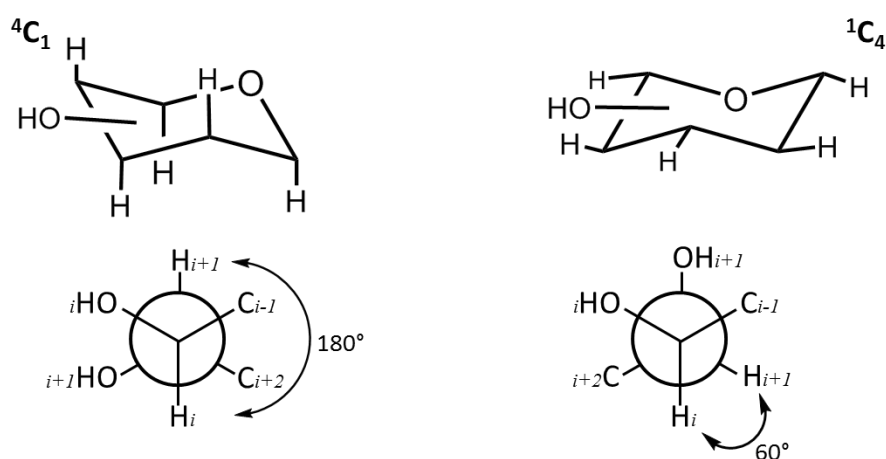
#### 1.4.4 Coupling constants

In the late fifties, Nobel Laureate Martin Karplus, after Raymond Lemieux empirical observations,<sup>83</sup> described a relationship between the dihedral torsion angle established between vicinal hydrogens and the  ${}^3J_{HH}$  coupling constant, by using a theoretical approach.<sup>[84]</sup>

$${}^3J_{HH} = A + B\cos\phi + C\cos^2\phi \quad \text{eq. 1.5}$$

Where  $\phi$  is the dihedral angle between the two protons.

According to equation 1.5, for dihedral angles of  $0^\circ$  or  $180^\circ$ , the observed coupling constant is large, while conversely, dihedral angles around  $90^\circ$  lead to  $\cos^2\phi$  values close to zero, and thus the observed coupling constant is small or null. For its part, coefficients  $A$ ,  $B$  and  $C$  are parametrized for each particular kind of molecule, atom, and substituent. In this regard, and for instance, eight new Karplus relationships have been refined for use in conformational studies of saccharides.<sup>[85]</sup> In chapter II we have made extensive use of  $J$  coupling analysis for the characterization of the conformational equilibrium in Glc and Ido mimics. Among all different found conformers, the two energy minima for Glc are dubbed as  ${}^1C_4$  and  ${}^4C_1$ . In the case of the  ${}^4C_1$  conformer, almost all the dihedral angles between protons of adjacent carbon atoms are maximum, being their respective vicinal coupling constants of around 10 Hz, while in the  ${}^1C_4$  conformer, the trans-diequatorial disposition implies a gauche geometry between vicinal protons and their  $J$  coupling values smaller than 5 Hz. Fig. 1.11.



**Figure 1.11** On the top,  $\beta$ -D-Glc in  ${}^4C_1$  conformation, left, and  ${}^1C_4$  conformation, right. The relative spatial orientation for vicinal hydrogen atoms is more clearly schematized in the Newman projection on the bottom. The dihedral angle between those nuclei is of  $180^\circ$  in  ${}^4C_1$  conformation and of  $60^\circ$  in  ${}^1C_4$ . Consequently, the  ${}^3J_{HH_{i+1}}$  coupling constants are around 10 Hz in  ${}^4C_1$  chair conformation and less than 5 Hz in  ${}^1C_4$ .

Of note, scalar coupling constants have the advantage that calculation of average values over an ensemble of conformations is simple, allowing the characterization of the conformational equilibrium in terms of population distribution.<sup>[86]</sup> Today, the Karplus equation is behind every conformational analysis studied by NMR. Variations of the Karplus equation are profusely used in elucidating the structure of organic compounds from NMR data. Furthermore, not only  $^3J_{HH}$  coupling constants, but also longer-range coupling ( $^xJ_{HH}$ ,  $x > 3$ ), as well as single and multiple-bond heteronuclear coupling<sup>[87,88]</sup> ( $^xJ_{HY}$ ,  $1 \leq x \leq 4$ , and  $Y = ^{13}\text{C}, ^{15}\text{N}, ^{19}\text{F}$ , etc.)<sup>[89]</sup> and homonuclear  $^xJ_{YY}$  coupling constants can provide a wealth of connectivity information on how molecular fragments are linked together, a geometric information that complements the one provided by NOE spectroscopy.

#### 1.4.5 Nuclear Overhauser Effect NOE Spectroscopy (NOESY)

Three-dimensional structures of carbohydrates become complicate to infer as the number of saccharide units increases. Knowledge on the spatial relationships among nuclei belonging to different sugar units can be provided by both NOESY studies and paramagnetic restraints. These two sources of structural information are based on their dependency on the interproton distance through space. In 2002, Kurt Wüthrich was awarded with the Nobel Prize in Chemistry for demonstrating that 2D NOE spectroscopy could be used to determine the 3D structures of biological macromolecules in solution. The NOE is defined as the change in intensity of a nucleus signal when the spin transitions of another nucleus cause a perturbation of its equilibrium populations. The two nuclei do not share a scalar, through bond, coupling; instead, they are sufficiently close in space to share a dipolar coupling. Thus, the NOE originates from dipolar cross-relaxation between pairs of protons, and depends on the proton-proton distance and on the molecular motion of the interproton vector:

$$I_{NOE} \approx \langle 1/r^6 \rangle f(\tau_c) \quad \text{eq. 1.6}$$

Where  $I_{NOE}$  is the NOE intensity,  $r$  is the proton-proton distance, and  $f$  is a function that depends, among others factors, on the correlation time ( $\tau_c$ ) that describes the motion of the interproton vector.

The inverse-sixth relationship in equation 1.6 implies that the NOE only develops between nuclei that are close in space, usually within 5 Å. Of course, not only homonuclear NOESY, but also heteronuclear studies are possible. For instance,  $^{19}\text{F}$ - $^1\text{H}$  heteronuclear NOE spectroscopy

(HOESY)<sup>[64]</sup> has been used in Chapter III of this thesis to estimate the conformation of disaccharides mimics at the glycosidic linkage.

#### 1.4.6 Long range restraints PCSs, RDCs and PREs

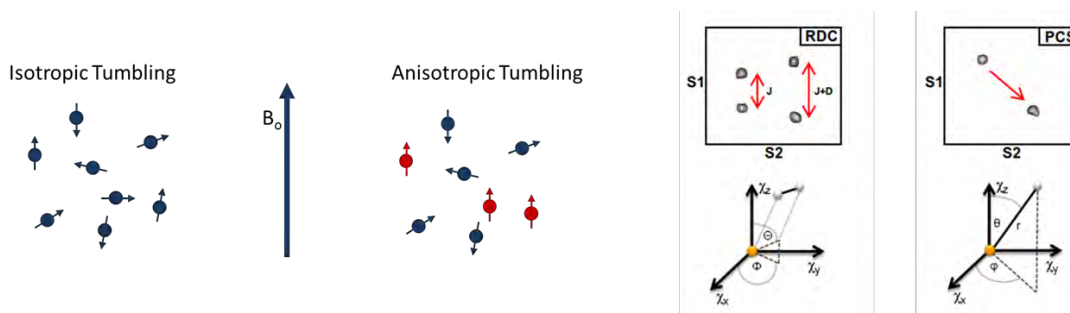
The structural complexity in saccharides proportionally increases with the number of monosaccharidic units, making the study of the conformational preferences of large oligosaccharides a challenging task. In this scenario, residual dipolar coupling (RDCs),<sup>[90,91]</sup> pseudo contact shifts (PCSs)<sup>[92]</sup> and paramagnetic relaxation enhancements (PREs)<sup>[93]</sup> actively contribute to the structural characterization of large molecules in solution.<sup>[94]</sup>

RDC provides relative orientations among internuclei vectors independently of their distance separation. This property of RDC has opened new avenues in the structural determination of large macromolecules with highly flexible functional arms or domains.<sup>[95,96]</sup> In the presence of a magnetic field, RDCs arise when the molecule in solution weakly align relative to the field, thus creating an anisotropic condition. Under such anisotropic condition, with the presence of an external field, a magnetic dipole-dipole interaction does not average to zero and yields a measurable dipolar coupling. The magnitude of the dipolar coupling depends on the angle between the internuclei vector and the external magnetic field as well as the internuclei distance. If the dipolar interaction is between two covalently bonded nuclei, then the internuclear distance is fixed and only the orientation dependence remains.<sup>[97]</sup>

$$D_{AB} = \frac{hB_0^2 \gamma_A \gamma_B}{240\pi^3 kT} r_{AB}^{-3} \left\{ \rho_{ax} (3\cos^2 \Theta - 1) + \frac{3}{2} \rho_{rh} \sin^2 \Theta \cos 2\Phi \right\} \quad \text{eq. 1.7}$$

Where  $D_{AB}$  is the residual dipolar coupling observed between nuclei  $A$  and  $B$ ,  $r_{AB}$  is the distance between nuclei  $A$  and  $B$ ,  $\Theta$  is the angle between the  $A$ - $B$  internuclear vector and the  $z$  axis of the  $\rho$ -tensor,  $\Phi$  is the angle between the projection of the  $A$ - $B$  internuclear vector on the  $xy$  plane and the  $x$  axis of the  $\rho$ -tensor,  $B_0$  is the static magnetic field strength and  $h$  is Planck's constant. Fig. 1.12.





**Figure 1.12** Vectorial representations of magnetic nuclei in an isotropic or anisotropic media. On the right, definition of the angular parameters used in equation 1.7 and 1.8, respectively.

In a normal solution, the solute tumbles almost isotropically, the interaction of intrinsic magnetic susceptibility of the macromolecule with the magnetic field produce a very weak alignment. As a result, the time average of the angular term  $\langle \cos^2\Theta \rangle$  is close to zero, leading to small values of dipolar couplings that are close to the experimental error; hence the term “residual” to refer to RDCs. Alternatively, the molecule sample has to be conjugated or mixed with an alignment medium with substantially large susceptibility anisotropy that can be aligned under an external magnetic field. With this approach, the degree of alignment is roughly one order of magnitude stronger than the former one and thus, practical for general application. The calculation of the dipolar interaction between a pair of dipolar-coupled nuclei requires the value of  $\langle \cos^2\Theta \rangle$ , which can be determined using the alignment tensor methodology.<sup>[98]</sup> The alignment tensor is described for a rigid part of the molecule using a minimum of five RDCs pointing in nonparallel direction. These values are used as restraints for structural determination of the flexible region of the molecule under study. The back calculated values for different minimized structures are compared with the experimental ones. A good fit between the experimental RDCs values and those obtained for one or for an ensemble of different conformations defines the three-dimensional structure of the macromolecule in solution.<sup>[99]</sup>

PCSs also arise from dipolar interactions. However, in this case, the effect is due to the dipolar interaction between the unpaired electrons of a paramagnetic entity, e.g. a lanthanide ion, and the nuclei in its vicinity. The PCS effect occurs only when the paramagnetic center presents a non-isotropic magnetic susceptibility, which is the case of some lanthanide cations, and it is dependent on the relative position of the paramagnetic center (its magnetic susceptibility tensor,  $\Delta$ ) and the affected nuclei. The equation that governs PCS is given in eq 1.8.

$$\Delta\delta_{PCS} = \frac{1}{12\pi} r^{-3} \left\{ \Delta\chi_{ax} (3\cos^2\theta - 1) + \frac{3}{2} \Delta\chi_{rh} \sin^2\theta \cos 2\varphi \right\} \quad eq. 1.8$$

where  $\Delta\delta_{PCS}$  is the difference in chemical shift between diamagnetic and paramagnetic samples,  $r$  is the distance between the metal ion and the nuclear spin,  $\theta$  and  $\varphi$  are the angles describing the position of the nuclear spin with respect to the principal axes (with the metal ion at the origin) of the magnetic susceptibility tensor ( $\chi$ ), and  $\Delta\chi_{ax}$  and  $\Delta\chi_{rh}$  are the axial and rhombic components, respectively, of the magnetic susceptibility tensor.

This restraint depends from the distance between the metal ion and the nuclear spin, with a  $1/r^3$  dependence. Thus, PCSs, respect to NOE, provide longer-range structural information, being observable up to 40 Å distance for certain paramagnetic metals. In this scenario, it does not surprise that the attachment of a small molecular fragment as metal chelating tag to flexible macromolecules, as oligosaccharides,<sup>[100]</sup> and dynamic proteins<sup>[101]</sup> or protein/ligand complexes<sup>[102]</sup> gives excellent results in terms of structural and dynamic determination of the macromolecular system. In chapter IV of this thesis, we have applied this strategy for unravel the conformational landscape of GGBP, a highly flexible sugar binding protein.

As consequence of the proximity to the paramagnetic metal ion, a second effect has to be taken into account. Paramagnetic relaxation enhancement also arises from magnetic dipolar interactions between a nucleus and the unpaired electrons of the paramagnetic center, which results in an increase in nuclear relaxation rates.<sup>[103]</sup> In contrast to PCSs, PRE vanishes faster, owing to its dependency with  $1/r^6$ , where  $r$  is the distance between the affected nuclei and the paramagnetic center. The key for using the PRE to study low populated states resides in the finding that the observed broadening of the NMR signals in the fast exchange regime are population weighted averages of the PREs for the major and minor species, thereby enabling structural information on the minor species to be extracted.<sup>[104,105,106]</sup> However, the main advantage of PRE is that can be detected in the presence of any paramagnetic system, being independent from the isotropic or anisotropic conditions.

Respect to the above mentioned NMR techniques, additional and more accurately described NMR based methods can be found in specialized books and reviews.<sup>[107,108]</sup> These methods, when assisted by computational protocols, yield to the elucidation of the structural and dynamics features of flexible biomolecules in solution.

## 1.5 Computational Chemistry

Although the next paragraphs are devoid to the practical and advantageous use of computational chemistry for structural and dynamic characterization of small molecules and macromolecules, it is important to highlight that the acquisition of experimental data play an essential role in validating the simulation methodology. Actually, comparisons of simulation and experimental data serve to test the accuracy of the calculated results and to provide criteria for improving the methodology. This is particularly important because the errors introduced by the use of empirical potentials are difficult to quantify. A number of studies comparing the simulations with experimental data have been used to validate computational techniques.<sup>[109]</sup> NMR data are particularly useful, as the many receptor and ligand conformations sampled computationally can be used to predict values of NMR parameters like spin relaxation, dipolar interactions or coupling constants, permitting direct comparison between experimental and theoretical techniques. Indeed, a number of studies have shown good agreement between computational and experimental measurements.<sup>[110,111,112]</sup> According to the Born-Oppenheimer approximation, the Schrodinger equation for a given molecule can be separated into a part describing the motion of the electrons, the part at which quantum mechanics (QM) is interested, and a part describing the motions of the nuclei and that these two motions can be studied independently. The structural problem associated to the latter part is the aim of Molecular Mechanics (MM) studies.

### 1.5.1 Quantum Chemical Models

When the molecular structure is critically determined by electron delocalization, it is not possible to neglect the energy contribution coming from the stabilizing molecular orbital interactions. This is the case for carbohydrates, where the glycosidic linkage is largely controlled by the anomeric effect. For these systems, the structural problem has to be solved considering the electrons as discrete particles or as a cloud, not restricted to the proper nucleus. Highly expensive computational methods are necessary. However, if the molecule is not too large, QM calculations can be achieved in a reasonable time and the resulting data are usually very accurate. QM describes the molecules in terms of interactions between nuclei and electrons, and the molecular geometry in terms of minimum energy arrangements of them. All QM methods ultimately trace back to the Schrödinger equation. For a molecular system, which obviously includes more than one electron, the Schrödinger equation cannot be solved, and thus

approximations need to be introduced to provide practical results. The different kind of possible approximations and their choice depends on the molecular system under study, and can be found in specialized books.<sup>[113,114]</sup>

### 1.5.2 Density Functional Models

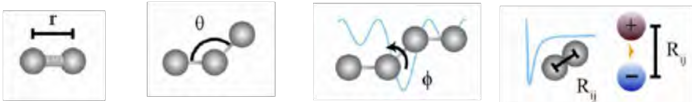
The density functional theory (DFT) assumes that the sum of the exchange and correlation energies of a uniform electron gas can be calculated knowing only its density. The main advantage in DFT is that, unlike the wave function, the electron density is an observable and can be measured experimentally, e. g. by X-ray diffraction. The approximation assumes that we can reduce the numbers of all the electrons that are in the molecule to a single electron density, reducing considerably the computational cost that would be otherwise too demanding. The complete physical-mathematic description of the equations behind DFT is beyond the scope of this thesis, just a brief discussion of the main properties of the electron density is presented.

- i)* The density integrates to the number of electrons; this means that the density is a function of all the electrons in the molecule.
- ii)* The density has maxima only at the position of the nuclei; in others terms, the density take into account the attractive force exerted by the positive charge of the nuclei.
- iii)* The density at the position of the nucleus contains information about the nuclear charge  $Z$ . Defining, in this way, the chemical nature of molecular constituent elements.
- iv)* The theory includes a term for the exchange and correlation electron interaction. This means that exist solutions for the problems of Coulombic repulsion and exclusion Pauli principle.

Thus, the electron density already provides all the ingredients necessary for a complete characterization of electronic structure in molecule.<sup>[115]</sup> The success of DFT in computational chemistry field is demonstrated by the increasing number of publication where the method has been applied. Among these, numerous are the application for characterization of mono- and oligosaccharides where the stereoelectronic effects strongly define the molecular geometry and energy, as for the anomeric effect.<sup>[116,117,118,119,120]</sup>

### 1.5.3 Molecular Mechanics (MM) Models

When the molecular system under study is composed of a considerable number of atoms the more sophisticated quantum molecular modeling techniques are too demanding of computer time resources to be of general use. In these cases, MM calculations, based on empirically parameterized force field and classic mechanic calculations are of practical use in the qualitative descriptions of the molecular model. MM focus on the structural aspect and not on the electronic properties of the molecule. In this, the motions of the nuclei are studied and the electrons are not explicitly examined at all, but are assumed to find an optimal distribution about the nuclei. The resulting potential energy surface (PES) is a multi-dimensional surface describing the potential energy of the molecule in function of the nuclear position. In MM, the atoms in the molecule are considered as a series of points with mass and charge linked by a spring, the bond. Chemical bonds and atomic angles are modeled using simple virtual springs, and dihedral angles are modeled using a sinusoidal function that approximates the energy differences between eclipsed and staggered conformations.<sup>[121]</sup> Non-bonded forces arise due to van der Waals interactions, modeled using the Lennard-Jones 6-12 potential,<sup>[122]</sup> and charged (electrostatic) interactions, modeled using Coulomb's law. The laws of classic mechanics are used in order to calculate the energy of the system, while simple mathematical derivations are used to find the energy minimum. The total molecular energy is a sum of different terms:

$$E_{total} = \underbrace{\sum_{bonds} K_r (r - r_{eq})^2}_{E_s} + \underbrace{\sum_{angles} K_\theta (\theta - \theta_{eq})^2}_{E_b} + \underbrace{\sum_{dihedrals} \frac{V_n}{2} [1 + \cos(\phi - \gamma)]}_{E_w} + \underbrace{\sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]}_{E_{nb}} \quad eq. 1.9$$


Where  $E_s$  is the energy for bond deformation, stretching or compression,  $E_b$  is for angle bending,  $E_w$  is the torsional angle energy, and  $E_{nb}$  is the energy for not bonded interactions and contains van der Waals, Coulombic and hydrogen bond interactions.

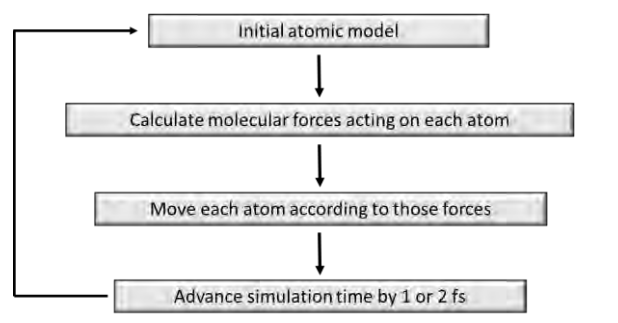
Each one of these energy contribution is calculated as *function* of the position of the nuclei, in practical terms the *function* is a force field. In order to reproduce the actual behavior of real molecules, in the force field all the parameters defining the atom type, charge and their relative position are experimentally defined or, when experimentally ill-defined, they are computationally determined using high level molecular modeling methods (QM calculations). This parametrization includes identifying the ideal stiffness and lengths of the springs that describe chemical bonding and atomic angles, determining the best partial atomic charge used for calculating electrostatic-interaction energies, identifying the proper van der Waals atomic

radii, and so on. Different force fields are now available and their choice depends on the specific molecular system. Some advanced force fields include additional types of terms, such as dipole interactions or stretch-bend couplings. In a conformational analysis the final goal is to locate the minimum energy structures, assuming that these are the best representation of the molecule. However, the search for the global minimum starts from a specific set of atomic coordinates, those used as input for the energy calculation, and the program finally finds the minimum which is close to the starting point coordinates. In others terms, the simulation is not able to cross transition states or saddle points those are in between two energy minima. This means that the found minimum could be a relative minimum and not the absolute one. A way to overcome this issue consists in an iterative variation, step by step, of the critical molecular coordinates without changing the others, followed by energy minimizations of the entire structures (conformational scanning for example). Specifically for carbohydrates, the higher flexible points are those that define the glycosidic linkages. The torsions  $\phi$  and  $\psi$  are monitored independently and the molecular potential energy is computed. As a result, a three dimensional plot is obtained, where the two variables,  $\phi$  and  $\psi$ , are plotted against the structural associated energy. The map provides practical information on the global and relative minima of the system. The obtained structures simplify significantly the conformational analysis excluding such conformations that are not populated on the basis of steric hindrances. This information has to be taken carefully, being some general force fields inadequate for describe molecular structures in which the stereoelectronic effects, as the anomeric effect, are determinant.<sup>[123,124,125,126]</sup> In order to overcome this lack a specific force field for carbohydrates has been developed. GLYCAM06 force field<sup>[127]</sup> implements the so called CHI Energy Functions. The carbohydrate intrinsic (CHI) energy functions were developed to score oligosaccharide structures based on glycosidic linkage conformations. QM calculations were used to obtain the torsional energy profiles of glycosidic torsion angles using several tetrahydropyran-based disaccharide models. Similar energy profiles were grouped, and an average energy curve for each group was calculated, resulting in final four CHI energy equations. Currently, CHI energy functions are applied to all standard glycosidic linkages between pyranoses in the chair conformation.

Finally, the calculated structures have to be validated experimentally.

### 1.5.4 Molecular Dynamics (MD)

The term MD refers to a simulation that analyzes biomolecule motions and determines conformation and stability patterns. These simulations provide a complete description of the system's evolution as function of time, what is called trajectory.<sup>[128,129,121]</sup> MD fill the gap that exists in between the most powerful techniques that are currently used in structural characterization of biomolecules. Actually, if X-ray crystallography gives with the structural details at atomic resolution of the molecule under study, it provides a static representation of the molecular system. On the other side, NMR has the privilege to provide dynamic features of the system in its proper environment, in solution. However, the time scales and the amplitude of the motions may be ill-defined by such techniques, often because some molecular motions are too fast and the different states cannot be singularly detected and weighted mean values are observed. When this happens, MD provides key dynamic information on the structure, dynamics, and even thermodynamics of biological molecules and their complexes.<sup>[130]</sup> The simplest description of the dynamics between two atoms, arms or domains of the molecule in solution should define the speed at which they move, as well as the time scale, and the amplitude of the motions. MD is based on Newtonian physics to simulate atomic motions. The general protocol used is outlined in scheme 1.1.



**Scheme 1.1.** How a MD simulation is performed. First, a computer model of the molecular system is prepared. An equation like eq. 1.9 is used to estimate the force acting on each of the system atoms. The positions of the atoms are moved according to Newton's laws of motion. The simulation time is advanced, and the process is repeated many times.

First, a computer model of the molecular system is prepared from NMR, X-ray crystallography or homology-modeling data. The forces acting on each of the atoms of the system are then estimated from the same equation described for MM (eq. 1.9). In brief, forces arising from interactions between bonded and non-bonded atoms contribute. According to the second law or

the equation of motion,  $F = ma$ , where  $F$  is the force exerted on the particle,  $m$  is its mass and  $a$  is its acceleration. From knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of the motion then yields a trajectory that describes the positions, velocities and acceleration of the particles as they vary with time. From this trajectory, the average values of properties can be determined. The method is deterministic; once the positions and velocities of each atom are known, the state of the system can be predicted at any time. The two dramatic limitations of MD are the time scale and the accuracy of the selected force field. MD are time consuming and computationally expensive. However, recent advances in hardware, software, and algorithms have increased the timescales accessible to simulation by several orders of magnitude, enabling millisecond-scale simulations and allowing MD to capture many critical biochemical processes as active transport or channel gating.<sup>[131]</sup> Recently, it has been proposed that long MD simulations in the millisecond regime are necessary to well describe the conformational behavior of oligosaccharides in solution.<sup>[132]</sup> Many reactions and conformational transitions exhibit long time scales because they consist of one or more activated processes, such as local conformational changes associated with ligand binding.<sup>[133]</sup> In these cases, conventional MD is unsuitable for investigating activated processes because of the simulation time scale. Steered Molecular Dynamics (SMD) simulations introduce a time-dependent or position-dependent force.<sup>[134]</sup> The purpose of this force is to steer systems along particular pathways. This allows focusing on dynamic events of interest while keeping computational expense to a minimum.<sup>[135,136,137]</sup> In chapter IV of the thesis, an example where external forces drive the binding event between GGBP and Glc is analyzed. SMD is the computational analogue of the experimental techniques which apply external mechanical manipulations to biomolecules, as AFM, optical tweezer or dynamic force spectroscopy experiments.<sup>[138,139,140]</sup> Free energy differences can be obtained as a function of the work done on the system. Albeit, accelerated, targeted or steered MD simulations give only an estimation of the energetic cost associated to the transition, their benefit in significant reduction of the computational cost is a reason of its growing use.



## 1.6 Objectives

The general objective of this Thesis is the study of the dynamic behavior of sugars and proteins and to advance into the knowledge of its functional role in their interactions. Special focus has been placed on the development of novel NMR procedures that permit to advance in the knowledge of carbohydrates structure and recognition features.

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

Chapter 2:

-Characterize the dynamic behavior of sugar rings. Define the thermodynamic and kinetic features of conformational exchange in pyranose rings.

Chapter 3:

-Demonstrate the importance of the stereoelectronic component for the *exo*-anomeric effect. Show that *gem*-Difluorocarbadisaccharides are suitable glycomimetics able to restore the *exo*-anomeric effect that is absent in carbadisaccharides.

Chapter 4:

-By combining paramagnetic NMR experiments and MD simulations, it is possible to define a functional conformational landscape for a protein model, both structurally and energetically.

Chapter 5:

-Design and synthesis of a strategically isotope-labeled sugar, useful for the direct detection of weak intermolecular interaction.

## 1.7 References

1. Ambrosetti, A., Ferri, N. & Tkatchenko, A. Wavelike charge density fluctuations and van der Waals interactions at the nanoscale. *Science*. **351**, 1171–1176 (2016).
2. Pang, P.-C. *et al.* Human Sperm Binding Is Mediated by the Sialyl-Lewisx Oligosaccharide on the Zona Pellucida. *Science*. **333**, 1761–1764 (2011).
3. Desai, P. R. Immunoreactive T and Tn antigens in malignancy: role in carcinoma diagnosis, prognosis, and immunotherapy. *Transfus. Med. Rev.* **14**, 312–25 (2000).
4. Ju, T., Otto, V. I. & Cummings, R. D. The Tn antigen—structural simplicity and biological complexity. *Angew. Chemie - Int. Ed.* **50**, 1770–1791 (2011).
5. Buskas, T., Thompson, P. & Boons, G.-J. Immunotherapy for cancer: synthetic carbohydrate-based vaccines. *Chem. Commun.* **36**, 5335–5349 (2009).
6. Lakshminarayanan V. *et al.* Immune recognition of tumor-associated mucin MUC1 is achieved by a fully synthetic aberrantly glycosylated MUC1 tripartite vaccine. *Proc Natl Acad Sci U S A.* **109**, 261–6 (2012).
7. Wilson, R. M. & Danishefsky, S. J. A vision for vaccines built from fully synthetic tumor-associated antigens: From the laboratory to the clinic. *J. Am. Chem. Soc.* **135**, 14462–14472 (2013).
8. Cai, H. *et al.* Self - Adjuvanting synthetic antitumor vaccines from MUC1 glycopeptides conjugated to T - Cell epitopes from tetanus toxoid. *Angew. Chemie - Int. Ed.* **52**, 6106–6110 (2013).
9. Gabius, H. J., André, S., Kaltner, H. & Siebert, H. C. The sugar code: Functional lectinomics. *Biochim. Biophys. Acta - Gen. Subj.* **1572**, 165–177 (2002).
10. Cummings, R. D. The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **5**, 1087–1104 (2009).
11. Seeberger, P. H. & Werz, D. B. Synthesis and medical applications of oligosaccharides. *Nature* **446**, 1046–1051 (2007).
12. Gringhuis, S. I., Kaptein, T. M., Wevers, B. A., Mesman, A. W. & Geijtenbeek, T. B. H. Fucose-specific DC-SIGN signalling directs T helper cell type-2 responses via IKK $\epsilon$ - and CYLD-dependent Bcl3 activation. *Nat. Commun.* **5**, 3898 (2014).
13. Seeberger, P. H. Chemical glycobiology: why now? *Nat. Chem. Biol.* **5**, 368–372 (2009).
14. Gabius, H.-J. & Kayser, K. Introduction to glycopathology: the concept, the tools and the perspectives. *Diagn. Pathol.* **9**, 4 (2014).
15. Ardá, A. *et al.* Recent advances in the application of NMR methods to uncover the conformation and recognition features of glycans. *Carbohydr. Chem.* **42**, 47–82 (2017).
16. Cano, M. E. *et al.* Synthesis of divalent ligands of | -thio- and | -N-galactopyranosides and related lactosides and their evaluation as substrates and inhibitors of Trypanosoma cruzi trans-sialidase. *Beilstein J. Org. Chem.* **10**, 3073–3086 (2014).

17. Molinaro, A. *et al.* Chemistry of lipid A: at the heart of innate immunity. *Chemistry* **21**, 500–519 (2015).
18. Madala, N. E., Leone, M. R., Molinaro, A. & Dubery, I. A. Deciphering the structural and biological properties of the lipid A moiety of lipopolysaccharides from *Burkholderia cepacia* strain ASP B 2D, in *Arabidopsis thaliana*. *Glycobiology* **21**, 184–194 (2011).
19. Benkovic, S. J. in *Enzyme Kinetics and Mechanism Part A Initial Rate and Inhibitor Methods* (ed. Enzymology, B. T.-M. in) **Volume 63**, 370–379 (Academic Press, 1979).
20. Ribeiro, J. *et al.* Fluorinated Carbohydrates as Lectin Ligands: 19F-Based Direct STD Monitoring for Detection of Anomeric Selectivity. *Biomolecules* **5**, 3177–3192 (2015).
21. Sattelle, B. M. *et al.* Dependence of pyranose ring puckering on anomeric configuration: Methyl idopyranosides. *J. Phys. Chem. B* **116**, 6380–6386 (2012).
22. Rönnols, J., Manner, S., Ellervik, U. & Widmalm, G. Conformational effects due to stereochemistry and C3-substituents in xylopyranoside derivatives as studied by NMR spectroscopy. *Org. Biomol. Chem.* **12**, 8031–8035 (2014).
23. Speciale, G., Thompson, A. J., Davies, G. J. & Williams, S. J. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr. Opin. Struct. Biol.* **28**, 1–13 (2014).
24. Biarnés, X. *et al.* The conformational free energy landscape of beta- D- glucopyranose. Implications for substrate preactivation in beta- glucoside hydrolases. *J. Am. Chem. Soc.* **129**, 10686 (2007).
25. Rüdiger, H. *et al.* Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets. *Curr. Med. Chem.* **7**, 389–416 (2000).
26. DeBosh, B.J. Trehalose inhibits solute carrier 2A (SLC2A) proteins to induce autophagy and prevent hepatic steatosis. *Sci Signal.* **9**, 21 (2016).
27. Solís, D. *et al.* A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochimica et Biophysica Acta.* **1850**, 186-235 (2015).
28. Hasanzadeh, N., Nori-Shargh, D., Farzipour, M. & Ahmadi, B. The origin of the anomeric effect: probing the impacts of stereoelectronic interactions. *Org. Biomol. Chem.* **13**, 6965–6976 (2015).
29. Filloux, C. M. The Problem of Origins and Origins of the Problem: Influence of Language on Studies Concerning the Anomeric Effect. *Angew. Chemie Int. Ed.* **54**, 8880–8894 (2015).
30. Thøgersen, H., Lemieux, R., Bock, K. & Meyer, B. Further justification for the exo-anomeric effect. Conformational analysis based on nuclear magnetic resonance spectroscopy of oligosaccharides. *Can. J. Chem.* **60**, 44–57 (1981).
31. Jiménez-Barbero, J., Demange, R., Schenk, K. & Vogel, P. Synthesis and Solution Conformational Analysis of 2, 3-Anhydro-3-C-[(1R)-2,6-anhydro-1-deoxy-1-fluoro-D-glycero-D-gulo-heptitol-1-C-yl ]- D -D-gulo-furanose: first Example of a Monofluoromethylene-Linked C-disaccharide. *J. Org. Chem.* **66**, 5132–5138 (2001).

32. Kolym padi, M. *et al.* Synthesis and conformational analysis of (alpha-D-galactosyl)phenylmethane and alpha-,beta-difluoromethane analogues: interactions with the plant lectin viscumin. *Chemistry* **15**, 2861–2873 (2009).
33. Shang, F. *et al.* Crystallographic Complexes of Surfactant Protein A and Carbohydrates reveal Ligand-induced Conformational Change. *J. Biol. Chem.* **286**, 757–765 (2010).
34. Pérez-Castells, J. *et al.* The conformational behaviour and P-selectin inhibition of fluorine-containing sialyl LeX glycomimetics. *Org. Biomol. Chem.* **5**, 1087–1092 (2007).
35. Jiménez-Barbero, J. *et al.* The conformation of C-glycosyl Compounds. *ChemInform.* **32**, 1522-2667 (2001).
36. Derosa, G. & Maffioli, P.  $\alpha$ -Glucosidase inhibitors and their use in clinical practice. *Arch. Med. Sci.* **8**, 899–906 (2012).
37. Leclerc, E., Pannecoucke, X., Ethève-Quelquejeu, M. & Sollogoub, M. Fluoro-C-glycosides and fluoro-carbasugars, hydrolytically stable and synthetically challenging glycomimetics. *Chem. Soc. Rev.* **42**, 4270–83 (2013).
38. Van Straaten, K. E. *et al.* Structural basis of ligand binding to UDP-galactopyranose mutase from *Mycobacterium tuberculosis* using substrate and tetrafluorinated substrate analogues. *J. Am. Chem. Soc.* **137**, 1230–1244 (2015).
39. Vidal, P. *et al.* Conformational selection in glycomimetics: Human galectin-1 only recognizes syn- $\psi$ -type conformations of  $\alpha$ -1,3-linked lactose and its C-glycosyl derivative. *Chem. - A Eur. J.* **19**, 14581–14590 (2013).
40. Lin, C.-K., Chen, K.-T., Hu, C.-M., Yun, W.-Y. & Cheng, W.-C. Synthesis of 1-C-Glycoside-Linked Lipid II Analogues Toward Bacterial Transglycosylase Inhibition. *Chem. – A Eur. J.* **21**, 7511–7519 (2015).
41. Tam, H. K. *et al.* Structural characterization of O- and C-glycosylating variants of the landomycin glycosyltransferase LanGT2. *Angew. Chemie - Int. Ed.* **54**, 2811–2815 (2015).
42. Sutkeviciute, I. *et al.* Unique DC-SIGN clustering activity of a small glycomimetic: A lesson for Ligand Design. *ACS Chem. Biol.* **9**, 1377-1385 (2014).
43. Andriotis, V. M. E. *et al.* Iminosugar inhibitors of carbohydrate - active enzymes that underpin cereal grain germination and endosperm metabolism. *Biochem. Soc. Trans.* **44**, 159–165 (2016).
44. Gruner, S. A. W., Locardi, E., Lohof, E. & Kessler, H. Carbohydrate-based mimetics in drug design: Sugar amino acids and carbohydrate scaffolds. *Chem. Rev.* **102**, 491–514 (2002).
45. Fan, Q. H., Claunch, K. A. & Striegler, S. Structure-activity relationship of highly potent galactonoamidine inhibitors toward  $\beta$ -galactosidase (*Aspergillus oryzae*). *J. Med. Chem.* **57**, 8999–9009 (2014).
46. Ardá, A. *et al.* Molecular recognition of complex-type biantennary N-glycans by protein receptors: A three-dimensional view on epitope selection by NMR. *J. Am. Chem. Soc.* **135**, 2667–2675 (2013).
47. García-Herrero, A. *et al.* Conformational selection of glycomimetics at enzyme catalytic sites: Experimental demonstration of the binding of distinct high-energy distorted

- conformations of C-, S-, and O-glycosides by E. Coli  $\beta$ -galactosidases. *J. Am. Chem. Soc.* **124**, 4804–4810 (2002).
48. Müller, K., Faeh, C. & Diederich, F. Fluorine in Pharmaceuticals: Looking Beyond Intuition. *Science (80-. )*. **317**, 1881–1886 (2007).
  49. Biffinger, J. C., Kim, H. W. & DiMugno, S. G. The polar hydrophobicity of fluorinated compounds. *ChemBioChem* **5**, 622–627 (2004).
  50. Böhm, H.-J. *et al.* Fluorine in medicinal chemistry. *Chembiochem* **5**, 637–43 (2004).
  51. Chubarov, A. S. *et al.* Design of protein homocystamides with enhanced tumor uptake properties for  $^{19}\text{F}$  magnetic resonance imaging. *Bioorganic Med. Chem.* **23**, 6943–6954 (2015).
  52. Mishra, N. K., Urick, A. K., Ember, S. W. J., Schonbrunn, E. & Pomerantz, W. C. Fluorinated aromatic amino acids are sensitive  $^{19}\text{F}$  NMR probes for bromodomain-ligand interactions. *ACS Chem. Biol.* **9**, 2755–2760 (2014).
  53. Urick, A. K. *et al.* Dual Screening of BPTF and Brd4 Using Protein-Observed Fluorine NMR Uncovers New Bromodomain Probe Molecules. *ACS Chem. Biol.* **10**, 2246–2256 (2015).
  54. Wagner, G., Ph, D. & Macher, J. Brain Pharmacokinetics and Tissue Distribution In Vivo of Fluvoxamine and Fluoxetine by Fluorine Magnetic Resonance Spectroscopy. *Neuropsychopharmacology*. **23**, 428-439 (2000).
  55. Yang, F. *et al.* Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and  $^{19}\text{F}$ -NMR. *Nat. Commun.* **6**, 1–15 (2015).
  56. Hu, H., Kulanthaivel, P. & Krishnamurthy, K. Simultaneous characterization of a mixture of fluorochemicals using three-dimensional  $^{19}\text{F}$ - $^1\text{H}$  heteronuclear TOCSY filtered/edited NMR experiments. *J. Org. Chem.* **72**, 6259–6262 (2007).
  57. Matei, E. *et al.* Fluorinated Carbohydrates as Lectin Ligands: Dissecting Glycan–Cyanovirin Interactions by Using  $^{19}\text{F}$  NMR Spectroscopy. *Chem. – A Eur. J.* **19**, 5364–5374 (2013).
  58. Braitsch, M. *et al.* Synthesis of fluorinated maltose derivatives for monitoring protein interaction by  $^{19}\text{F}$  NMR. *Beilstein J. Org. Chem.* **8**, 448–455 (2012).
  59. 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.* **131**, 12949–12959 (2009).
  60. Keddie, N. S., Slawin, A. M. Z., Lebl, T., Philp, D. & O’Hagan, D. All-cis 1,2,3,4,5,6-hexafluorocyclohexane is a facially polarized cyclohexane. *Nat. Chem.* **7**, 483–488 (2015).
  61. Linclau, B. *et al.* 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.* **55**, 674–678 (2015).
  62. Linclau, B. *et al.* Intramolecular OH...Fluorine Hydrogen Bonding in Saturated, Acyclic Fluoroalcohols: The  $\gamma$ -Fluoropropanol Motif. *Chem. – A Eur. J.* **21**, 17808–17816 (2015).
  63. Bernet, B. & Vasella, A. Hydrogen bonding of fluorinated saccharides in solution: F acting

- as H-bond acceptor in a bifurcated H-bond of 4-fluorinated levoglucosans. *Helv. Chim. Acta* **90**, 1874–1888 (2007).
64. Combettes, L. E. *et al.* Conformational analysis of fluorinated pyrrolidines using <sup>19</sup>F-<sup>1</sup>H scalar couplings and heteronuclear NOEs. *Chem. - A Eur. J.* **18**, 13133–13141 (2012).
  65. Wolfert, M; Boons, G. Adaptive immune activation: Glycosylation does matter. *Nat Chem Biol* **9**, 776–784 (2013).
  66. Varki, a. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130 (1993).
  67. Gabius, H. J., André, S., Jiménez-Barbero, J., Romero, A. & Solís, D. From lectin structure to functional glycomics: Principles of the sugar code. *Trends Biochem. Sci.* **36**, 298–313 (2011).
  68. Imberty, A., Wimmerová, M., Mitchell, E. P. & Gilboa-Garber, N. Structures of the lectins from *Pseudomonas aeruginosa*: Insights into the molecular basis for host glycan recognition. *Microbes Infect.* **6**, 221–228 (2004).
  69. Shilton, B. H., Flocco, M. M., Nilsson, M. & Mowbray, S. L. Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins. *J. Mol. Biol.* **264**, 350–363 (1996).
  70. 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.* **16**, 1032–1041 (2007).
  71. del Carmen Fernández-Alonso, M. *et al.* Protein-carbohydrate interactions studied by NMR: from molecular recognition to drug design. *Curr. Protein Pept. Sci.* **13**, 816–30 (2012).
  72. Unione, L., Galante, S., Díaz, D., Cañada, F. J. & Jiménez-Barbero, J. NMR and molecular recognition. The application of ligand-based NMR methods to monitor molecular interactions. *Medchemcomm.* **5**, 1280–1289 (2014).
  73. Berbís, M. Álvaro *et al.* *Carbohydrate Chemistry: State of the Art and Challenges for Drug Development* (Ed. Laura Cipolla) IMPERIAL COLLEGE PRESS, **2015**, pp. 121-146.
  74. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–2195 (2015).
  75. Camacho-Zarco, A. R. *et al.* Multiple paramagnetic effects through a tagged reporter protein. *Angew. Chemie - Int. Ed.* **54**, 336–339 (2015).
  76. Andreas, L. B., Le Marchand, T., Jaudzems, K. & Pintacuda, G. High-resolution proton-detected NMR of proteins at very fast MAS. *J. Magn. Reson.* **253**, 36–49 (2015).
  77. Lee, D., Hediger, S. & De Paëpe, G. Is solid-state NMR enhanced by dynamic nuclear polarization? *Solid State Nucl. Magn. Reson.* **66**, 6–20 (2015).
  78. Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry. Tetrahedron Organic Chemistry Series* **27**, (2009).
  79. Zheltikov, A. *Understanding NMR Spectroscopy* (Eds. James Keeler, John Wiley & Sons Ltd, Chichester), **2005**, pp. 459.

80. Zierke, M., Smies, M., Rabbani, S., Aeschbacher, T. & Cutting, B. Stabilization of Branched Oligosaccharides : Lewis x Bene fi ts from a Nonconvantional C-H...O Hydrogen Bond. *J. Am. Chem. Soc.* **135**, 13464-13472 (2013).
81. Yamaguchi, T. *et al.* Exploration of conformational spaces of high-mannose-type oligosaccharides by an NMR-validated simulation. *Angew. Chemie - Int. Ed.* **53**, 10941–10944 (2014).
82. Kamiya, Y. *et al.* Application of Metabolic <sup>13</sup>C Labeling in Conjunction with High-Field Nuclear Magnetic Resonance Spectroscopy for Comparative Conformational Analysis of High Mannose-Type Oligosaccharides. *Biomolecules* **3**, 108–23 (2013).
83. Coxon, B. Developments in the Karplus equations as they relate to the NMR coupling constant of Carbohydrates. *Adv. Carbohydr. Chem. Biochem.* **62**, 17–82 (2009).
84. Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **30**, 11 (1959).
85. Stenutz, R., Carmichael, I., Widmalm, G. & Serianni, A. S. Hydroxymethyl Group Conformation in Saccharides: Structural Dependencies of <sup>2</sup>J<sub>HH</sub>, <sup>3</sup>J<sub>HH</sub>, and <sup>1</sup>J<sub>CH</sub> Spin–Spin Coupling Constants. *J. Org. Chem.* **67**, 949–958 (2002).
86. Wang, X. & Woods, R. J. Insights into furanose solution conformations: beyond the two-state model. *J. Biomol. NMR.* **64**, 291-305 (2016).
87. Cano, F. Experimental evidence of deviations from a Karplus-like relationship of vicinal carbon-proton coupling constants in some conformationally rigid carbohydrate. *J. Org. Che.* **52**, 3367–3372 (1987).
88. Tafazzoli, M. & Ghiasi, M. New Karplus equations for <sup>2</sup>J<sub>HH</sub>, <sup>3</sup>J<sub>HH</sub>, <sup>2</sup>J<sub>CH</sub>, <sup>3</sup>J<sub>CH</sub>, <sup>3</sup>J<sub>COCH</sub>, <sup>3</sup>J<sub>CSC</sub>H, and <sup>3</sup>J<sub>CC</sub>CH in some aldohexopyranoside derivatives as determined using NMR spectroscopy and density functional theory calculations. *Carbohydr. Res.* **342**, 2086–2096 (2007).
89. Michalik, M., Hein, M. & Frank, M. NMR spectra of fluorinated carbohydrates. *Carbohydr. Res.* **327**, 185–218 (2000).
90. Azurmendi, H. F., Martin-Pastor, M. & Bush, C. A. Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings. *Biopolymers* **63**, 89–98 (2002).
91. Martin-Pastor, M. & Bush, C. A. The use of NMR residual dipolar couplings in aqueous dilute liquid crystalline medium for conformational studies of complex oligosaccharides. *Carbohydr. Res.* **323**, 147–155 (1999).
92. Zhang, Y., Yamamoto, S., Yamaguchi, T. & Kato, K. Application of Paramagnetic NMR-Validated Molecular Dynamics Simulation to the Analysis of a Conformational Ensemble of a Branched Oligosaccharide. *Molecules* **17**, 6658 (2012).
93. Jain, N. U., Venot, A., Umemoto, K., Leffler, H. & Prestegard, J. H. Distance mapping of protein-binding sites using spin-labeled oligosaccharide ligands. *Protein Sci.* **10**, 2393–2400 (2001).
94. Martin-Pastor, M. & Bush, C. A. Refined structure of a flexible heptasaccharide using <sup>1</sup>H-<sup>13</sup>C and <sup>1</sup>H-<sup>1</sup>H NMR residual dipolar couplings in concert with NOE and long range scalar

- coupling constants. *J. Biomol. NMR* **19**, 125–139
95. Tian, F., Al-Hashimi, H. M., Craighead, J. L. & Prestegard, J. H. Conformational analysis of a flexible oligosaccharide using residual dipolar couplings. *J. Am. Chem. Soc.* **123**, 485–492 (2001).
  96. Snider, J. D. *et al.* New strategy for RDC-assisted diastereotopic proton assignment using a combination of J-scaled BIRD HSQC and J-scaled BIRD HMQC/HSQC. *Magn. Reson. Chem.* **50 Suppl 1**, S86-91(2012).
  97. Chen, K. & Tjandra, N. The Use of Residual Dipolar Coupling in Studying Proteins by NMR. *Top. Curr. Chem.* **326**, 47–67 (2012).
  98. Wirz, L. N. & Allison, J. R. Fitting alignment tensor components to experimental RDCs, CSAs and RQCs. *J. Biomol. NMR* **62**, 25–29 (2015).
  99. Pham, T. N., Hinchley, S. L., Rankin, D. W. H., Liptaj, T. & Uhrin, D. Determination of sugar structures in solution from residual dipolar coupling constants: methodology and application to methyl beta-D-xylopyranoside. *J. Am. Chem. Soc.* **126**, 13100–13110 (2004).
  100. Canales, A. *et al.* Breaking pseudo-symmetry in multiantennary complex N-glycans using lanthanide-binding tags and NMR pseudo-contact shifts. *Angew. Chemie - Int. Ed.* **52**, 13789–13793 (2013).
  101. Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S. & Griesinger, C. Interdomain dynamics explored by paramagnetic NMR. *J. Am. Chem. Soc.* **135**, 17111–17120 (2013).
  102. Brath, U. *et al.* Paramagnetic Ligand Tagging To Identify Protein Binding Sites. *J. Am. Chem. Soc.* **137**, 11391–11398 (2015).
  103. Clore, G. M. & Iwahara, J. Theory, Practice and Applications of Paramagnetic Relaxation Enhancement for the Characterization of Transient Low-Population States of Biological Macromolecules and Their Complexes. *Chem. Rev.* **109**, 4108–4139 (2009).
  104. Theillet, F. X., Binolfi, A., Liokatis, S., Verzini, S. & Selenko, P. Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: Application to intrinsically disordered proteins. *J. Biomol. NMR* **51**, 487–495 (2011).
  105. Gottstein, D., Reckel, S., Dötsch, V. & Güntert, P. Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* **20**, 1019–1027 (2012).
  106. Seetaha, S. *et al.* Application of Site-Specific Spin Labeling for NMR Detecting Inhibitor-Induced Conformational Change of HIV-1 Reverse Transcriptase. *ChemMedChem* **11**, 363–366 (2016).
  107. Cavanagh, J., Fairbrother W. J., Palmer, A. G., Rance, M., Skelton, N. J. *Protein NMR Spectroscopy* (2nd Edition, Elsevier Science), **2006**.
  108. Bertini, I., Luchinat, C., Parigi, G., *Solution NMR of Paramagnetic Molecules* (1st Edition, Elsevier Science), **2001**.
  109. van Gunsteren, W. F., Dolenc, J. & Mark, A. E. Molecular simulation as an aid to experimentalists. *Curr. Opin. Struct. Biol.* **18**, 149–153 (2008).



110. Peter, C. *et al.* Molecular Dynamics Simulations of Small Peptides: Can One Derive Conformational Preferences from ROESY Spectra? *Chem. - A Eur. J.* **9**, 5838–5849 (2003).
111. Showalter, S. & Bruschweiler, R. Validation of MD simulations using NMR Spin Relaxation as benchmarks: Application to the AMBER99SB Force Field. *J. Chem. Theory Comput.* **3**, 961-975 (2007).
112. Markwick, P. R. L. *et al.* Enhanced conformational space sampling improves the prediction of chemical shifts in proteins. *J. Am. Chem. Soc.* **132**, 1220–1221 (2010).
113. Hehre, W. J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations.* (Wavefunction, Inc.) **2003**.
114. Koch, W. & Holthausen, M. C. *A Chemist's Guide to Density Functional Theory.* (WILEY-VCH) **2001**.
115. Dibble, D. J., Ziller, J. W. & Woerpef, K. A. Spectroscopic and X-ray crystallographic evidence for electrostatic effects in 4-substituted cyclohexanone-derived hydrazones, imines, and corresponding salts. *J. Org. Chem.* **76**, 7706–7719 (2011).
116. Cocinero, E. J., Carçabal, P., Vaden, T. D., Simons, J. P. & Davis, B. G. Sensing the anomeric effect in a solvent-free environment. *Nature* **469**, 76–79 (2011).
117. Hricovini, M., Driguez, P. A. & Malkin, O. L. NMR and DFT Analysis of Trisaccharide From Heparin Repeating-Sequence. *J. Phys. Chem. B.* **118**, 11931-11942 (2014).
118. Roslund, M. U., Tähtinen, P., Niemitz, M. & Sjöholm, R. Complete assignments of the <sup>1</sup>H and <sup>13</sup>C chemical shifts and JH,H coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides. *Carbohydr. Res.* **343**, 101–112 (2008).
119. Juaristi, E. & Notario, R. Theoretical examination of the S-C-P anomeric effect. *J. Org. Chem.* **80**, 2879–2883 (2015).
120. Csonka, G. I. Proper basis set for quantum mechanical studies of potential energy surfaces of carbohydrates. *J. Mol. Struct. THEOCHEM* **584**, 1–4 (2002).
121. Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **9**, 71 (2011).
122. Jones, J. E. On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Proc. R. Soc. London A Math. Phys. Eng. Sci.* **106**, 463–477 (1924).
123. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B. & Woods, R. J. Carbohydrate Docking : Sorting the Wheat from the Chaff. *J. Comput. Chem.* **35**, 526–539 (2015).
124. Foley, B. L., Tessier, M. B. & Woods, R. J. Carbohydrate force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 652–697 (2012).
125. Monticelli, L. & Tieleman, D. P. *Biomolecular Simulations.* (Eds. Monticelli, L. and Salonen, E. Springer) **2013**.
126. Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B. & Woods, R. J. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J. Comput. Chem.* **35**, 526–539 (2014).
127. Kirschner, K. N. *et al.* GLYCAM06: A generalizable biomolecular force field. carbohydrates.

- J. Comput. Chem.* **29**, 622–655 (2008).
128. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys.* **41**, 429–52 (2012).
  129. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
  130. Stockner, T., Vogel, H. J. & Tieleman, D. P. A salt-bridge motif involved in ligand binding and large-scale domain motions of the maltose-binding protein. *Biophys. J.* **89**, 3362–3371 (2005).
  131. Shaikh, S. A. *et al.* Visualizing functional motions of membrane transporters with molecular dynamics simulations. *Biochemistry* **52**, 569–587 (2013).
  132. Topin, J. *et al.* The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins. *ACS Chem. Biol.* **11**, 2011–2020 (2016).
  133. Unione, L. *et al.* Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations. *ACS Chem. Biol.* (2016). doi:10.1021/acscchembio.6b00148
  134. Adcock, S. A. & McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chem. Rev.* **106**, 1589–1615 (2006).
  135. Gao, M., Wilmanns, M. & Schulten, K. Steered molecular dynamics studies of titin I1 domain unfolding. *Biophys J* **83**, 3435–45 (2002).
  136. Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Current Opinion in Structural Biology* **11**, 224–230 (2001).
  137. Isralewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D. & Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.* **19**, 13–25 (2001).
  138. Huang, J. Exploration of molecular interactions in cholesterol superlattices: effect of multibody interactions. *Biophys. J.* **83**, 1014–25 (2002).
  139. Gu, J., Li, H. & Wang, X. A self-adaptive steered molecular dynamics method based on minimization of stretching force reveals the binding affinity of protein-ligand complexes. *Molecules* **20**, 19236–19251 (2015).
  140. Perišić, O. & Lu, H. On the improvement of free-energy calculation from steered molecular dynamics simulations using adaptive stochastic perturbation protocols. *PLoS One* **9**, (2014).

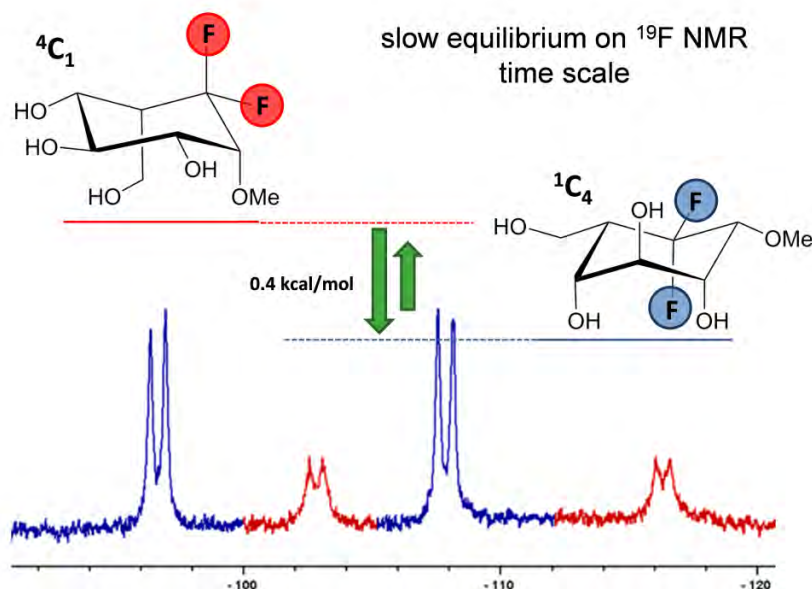


## CHAPTER II

Glycomimetics: understanding conformational plasticity in sugar rings

### Fluoro-carba-methyl-L-idopyranosides mimic the intrinsic dynamic behavior of natural idose rings

The work presented in this chapter has been performed in collaboration with the group of Dr. Matthieu Sollogoub, Sorbonne Université in Paris, France. Dr. Bixue Xu has been responsible for



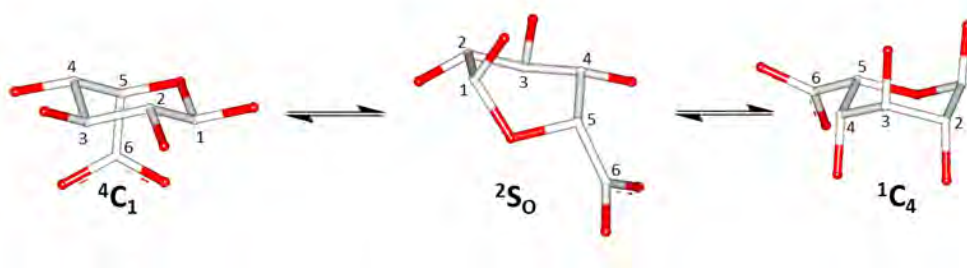
the synthesis of the glycomimetics discussed herein.

*“It is important to note that the definition of equilibrium as fast or slow is not a molecular feature, because it depends on the difference in frequency between the two states in exchange, which is different for each spin of the molecule”*

*Prof. Jorge Santoro*

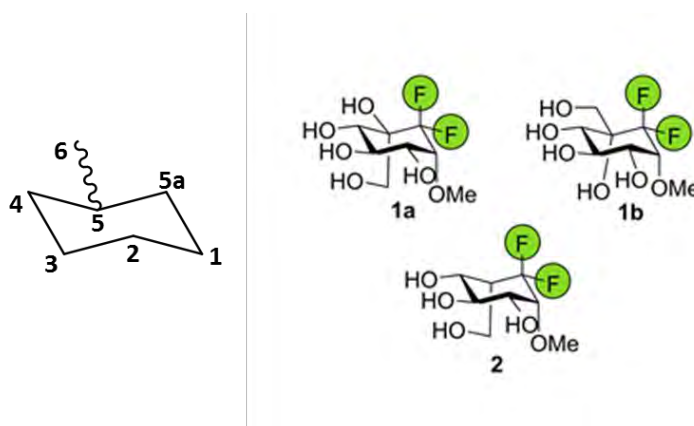
## 2.1 Introduction

One of the paradigmatic cases of conformational dynamics in glycosciences<sup>[1]</sup> is that of the Iduronic acid (L-IdoA) moiety present in heparin, L-IdoA. The extent of conformational mobility of L-IdoA in heparin oligosaccharides in both the free and bound states has also been matter of debate,<sup>[2]</sup> especially focused on the  ${}^4C_1$  chair-skew boat- ${}^1C_4$  chair equilibrium of the pyranose rings<sup>[3]</sup> in the free and protein-bound states. Fig. 2.1. Interestingly, depending on the protein receptor, distinct conformations of the L-IdoA ring are recognized.<sup>[4]</sup> Indeed, AT-III recognizes the skew boat conformer,<sup>[5]</sup> while the IdoA rings of a heparin hexasaccharide maintain the chair-skew boat flexibility when bound to FGF-1.<sup>[6]</sup> Fittingly, for AT-III case, Sinaÿ et al. prepared skew-boat conformationally-locked compounds that keep the biological activity, thus providing direct evidence on the recognition of these conformers by AT-III.<sup>[7]</sup> Obviously, this dynamic behavior has key implications in the kinetics and thermodynamics of the molecular recognition event.<sup>[8]</sup> Nevertheless, the access to idose (Ido) mimics that retain conformational plasticity and the quantification of the experimental values of the energy barriers and free energy differences for the chair-skew boat interconversion processes in water solution has remained elusive. Recent efforts using O-substituted Ido compounds have provided energy values in organic solvents.<sup>[9,10]</sup> However, given the intrinsic relative low energy barrier for this equilibrium, NMR experiments in water using hydroxylated natural compounds have failed to slow down the equilibrium to provide quantitative and non-ambiguous values.



**Figure 2.1.** Sugar ring conformations of  $\alpha$ -IdoA unit. Structural representation of the sugar ring conformations (chairs, C, and skew-boat, S) of the  $\alpha$ -IdoA unit, observed in glycosaminoglycans, such as heparin and heparin sulfate. Carbon atoms are numbered according to their positions within the sugar ring.

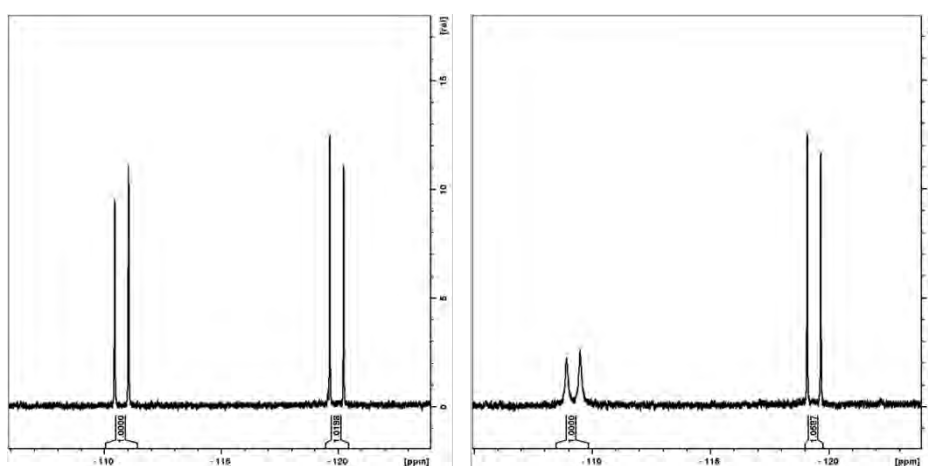
In this chapter, a new generation of fluorine-containing glycomimetics is presented. We have applied a combination of NMR and computational methods to investigate the conformational behavior of Ido- and Glc-like rings. We have used low temperature NMR experiments to slow down the conformational exchange of the Ido-like rings. Under these conditions, the exchange rate becomes slow in the  $^{19}\text{F}$  NMR chemical shift time scale and allows shedding light on the thermodynamic and kinetic features of the equilibrium. Despite the minimal structural differences between these compounds, a remarkable difference in their dynamic behavior indeed occurs. The importance of introducing fluorine atoms in these sugars mimics is also highlighted. Only the use of  $^{19}\text{F}$  NMR experiments has permitted to unveil key features of the conformational equilibrium that would have been otherwise remained unobserved. On this basis, we present the conformational analysis of different fluorine-containing sugar-mimetics. Thus, we have investigated two gem-difluorocarbascugars, **1a** and **2**, both analogues to methyl- $\beta$ -L-idopyranoside and possessing either a quaternary or a ternary C5 respectively. As model compound, we have also studied the corresponding Glc analogue **1b** (Fig. 2.2), since Glc pyranose rings are usually conformationally stable.  $^{19}\text{F}$  and  $^1\text{H}$  homo and heteronuclear NMR methods have been applied in water and dimethyl sulfoxide solutions to determine their intrinsic conformational and structural properties. The experimental NMR data have been supported by computational methods in order to unambiguously unravel the structural and conformational effects of the di-fluoro-methylene function.



**Figure 2.2.** Schematic representation of glycomimetics discussed in this work. C atoms follow the carbohydrate related numbering where C5a corresponds to the endocyclic oxygen.

## 2.2 Results and Discussion

**2.2.1 NMR analysis.** Compounds **1a** and **1b** only differ in the configuration of the stereogenic center at C5. Fig. 2.2. Strikingly, the observed  $^1\text{H}$ , and especially the  $^{19}\text{F}$  NMR spectra for both molecules, are dramatically different. Figure 2.3. The assignment of the two  $^{19}\text{F}$  resonances permitted to assess that the broad signal observed in **1a** corresponds to the axially oriented fluorine atom. As deduced from the visual inspection of the shape of the equatorial fluorine in **1a**, as well as of those of both fluorine atoms in **1b**, the behavior of the axial fluorine of **1a** is rather unique.



**Figure 2.3.**  $^{19}\text{F}$  NMR (470 MHz) spectra of **1b** (left) and **1a** (right) in the solvent  $\text{D}_2\text{O}$ , 298 K. Notice the broad shape of the axial fluorine resonance signal of **1a**. The integrals (1:1) are also shown.

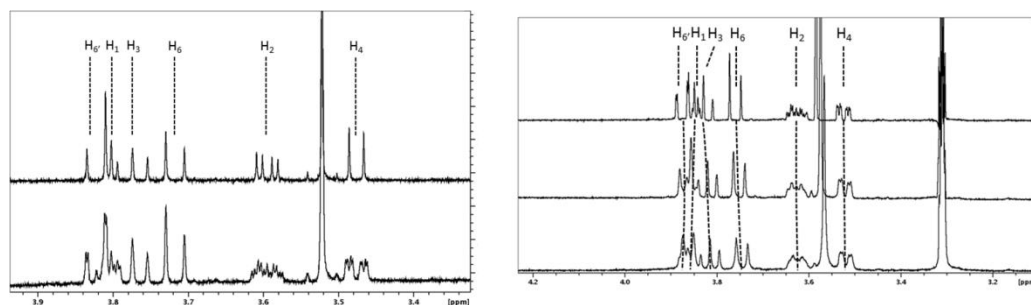
Given this particular feature, a detailed conformational analysis by using NMR methods had to be performed. Tables 2.1-2.2 gather the  $J$ -coupling and chemical shift data for the three studied molecules compared to the data for natural methyl-L-idopyranoside<sup>[11]</sup> and methyl-D-glucopyranoside<sup>[12]</sup> used as reference compounds.

The observations for the different molecules are given below:

### **Compound 1b** (gluco-like, difluoro, C5 is quaternary)

Evidences on the major shape of the six-membered ring of **1b** have been extracted from the analysis of the vicinal  $^3J_{\text{HH}}$  coupling constants (Table 2.1), measured with and without  $^{19}\text{F}$ -decoupling conditions. Fig. 2.4 left. As expected, very large  $^3J_{\text{H}2\text{H}3}$  and  $^3J_{\text{H}3\text{H}4}$  values were observed in water solution, thus demonstrating that it adopts a very major  $^4\text{C}_1$  chair conformation. No significant variations of chemical shifts and coupling constants were

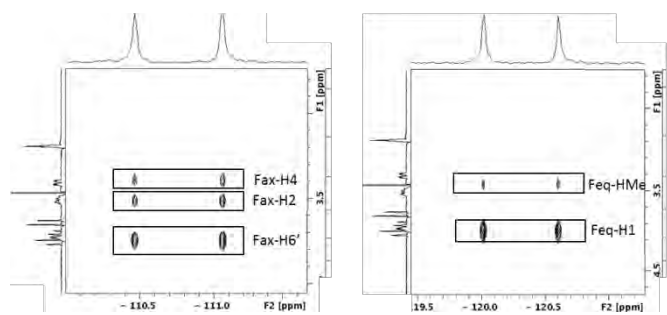
appreciated upon decreasing the temperature (Fig. 2.4 right) from 298 down to 248 K (adding 20% of deuterated methanol).



**Figure 2.4.** On the left, (top)  $^1\text{H}\{-^{19}\text{F}\}$  decoupled spectrum. (bottom)  $^1\text{H}$  NMR spectrum of **1b** at 298 K in  $\text{D}_2\text{O}$ . Notice the solved H2, H4, H1 and H6' signals. On the right,  $^1\text{H}$  NMR spectra between 248 (bottom), 263 (middle) and 298 K (top) in  $\text{D}_2\text{O}$  in the presence of 20% methanol. Methanol has been used as internal reference. 500 MHz Bruker spectrometer.

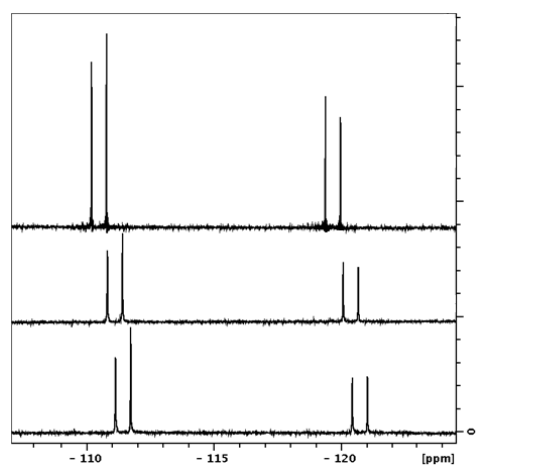
Similar couplings have been observed for **1b** in  $\text{DMSO-d}_6$  solution. Intermediate values for the  $^3J_{\text{H,OH}}$  coupling constants have been measured, ranging between 5.0 and 6.6 Hz (Table 2.3). Temperature coefficient factors are also of medium size for all the hydroxyl groups, between 4.9 and 7.0 ppb/ $^\circ$  (Table 2.4). These facts suggest that no particular orientations of the hydroxyl groups are favored. No strong intramolecular hydrogen bond between the hydroxyl moieties is present in  $\text{DMSO-d}_6$ . Obviously, the presence of competing water molecules in the water samples further precludes this possibility. Through-space coupling constants between the axial fluorine atom with H2, H4 and H6' have been also deduced. Indeed, these couplings are also supported by heteronuclear  $^1\text{H}\text{-}^{19}\text{F}$  NOEs between the corresponding atom pairs in the HOESY spectra. Fig. 2.5.





**Figure 2.5.**  $^1\text{H}$ - $^{19}\text{F}$  Hoesy 2D spectrum of **1b** in  $\text{D}_2\text{O}$ . Key heteronuclear NOEs cross peaks are highlighted.

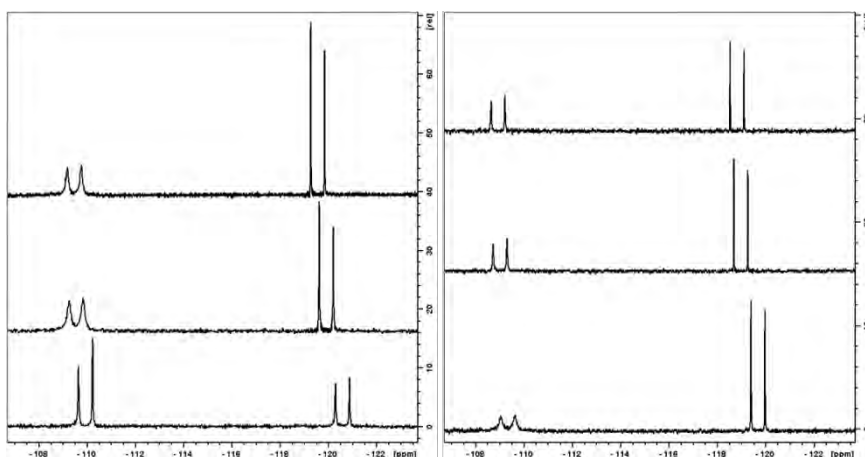
No significant features have been observed in the  $^{19}\text{F}$  NMR spectra of **1b** acquired at low temperature (Fig. 2.6). Therefore, all the NMR parameters and observations were in agreement with the existence of a  $^4\text{C}_1$  chair conformation, with no additional experimental observations worth of mentioning.



**Figure 2.6.** From top to bottom, the  $^{19}\text{F}$  NMR spectra of **1b** at 248, 263, and 298 K in  $\text{D}_2\text{O}$  in the presence of 20% of methanol. Trifluoroethanol has been used as internal reference. No significant changes are appreciated.

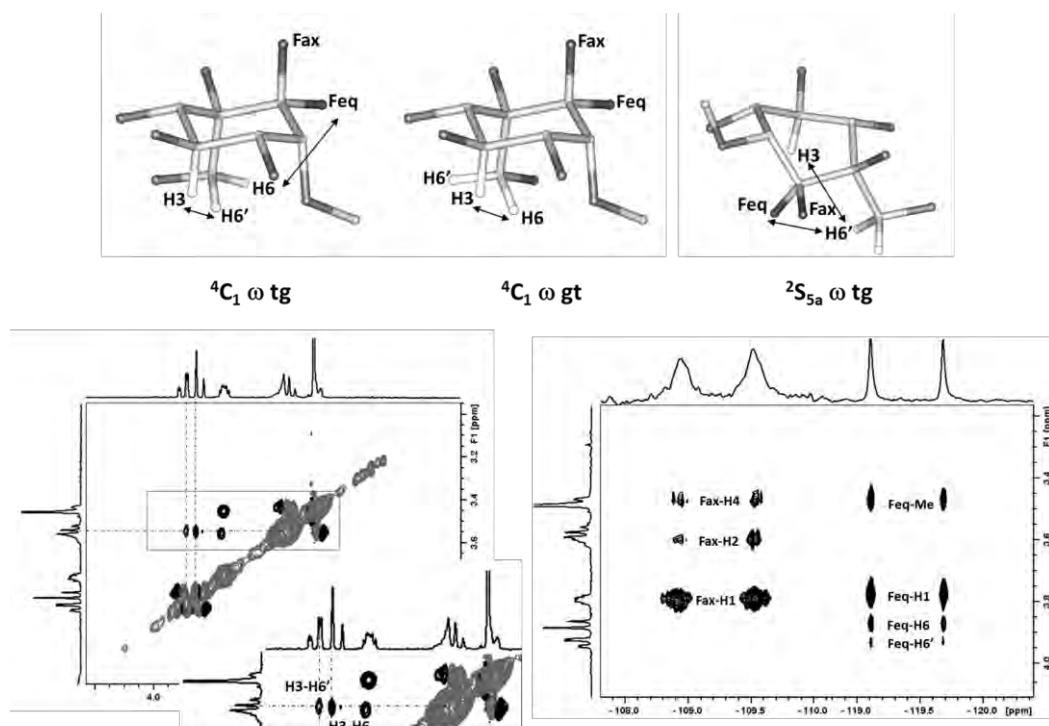
### **Compound 1a** (Ido-like, difluoro, C5 is quaternary)

$^{19}\text{F}$  NMR spectrum of **1a** is drastically different to that of **1b** (Fig. 2.3) at room temperature. Strikingly, the signal of the axial fluorine sharpened in a noticeable manner upon decreasing the temperature down to 238 K, using 20% of methanol, or increasing it at 333 K. Fig. 2.7. This fact suggests the existence of a dynamic process, which especially affects the transverse relaxation features of Fax.



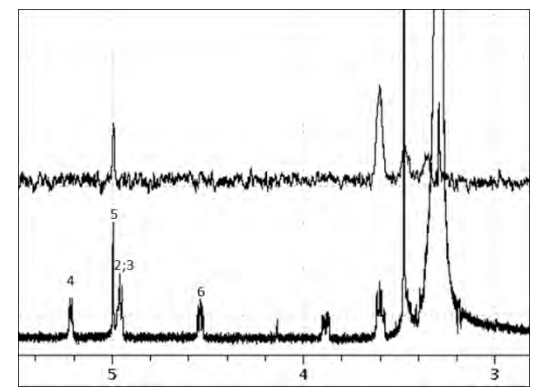
**Figure 2.7.** From bottom to top.  $^{19}\text{F}$  NMR (470 MHz) spectra of **1a** at 238 K, 273 K and 298 K (left) and at 283 K, 323 K and 333 K (right) in  $\text{D}_2\text{O}$  in presence of 20% of methanol. Left doublet, and right doublet correspond to axial and equatorial Fluorines in  ${}^4\text{C}_1$  conformation respectively.

Curiously, the  ${}^3J_{\text{H}_2\text{H}_3}$  and  ${}^3J_{\text{H}_3\text{H}_4}$  coupling constant values (Table 2.1) are also relatively large in water solution, suggesting that a major  ${}^4\text{C}_1$  chair conformation indeed exists at room temperature (see below, in discussion). As for **1b**, similar couplings are observed in  $\text{DMSO-d}_6$  solution, together with medium-size values (between 4.1 and 7.2 Hz) for the  ${}^3J_{\text{H},\text{OH}}$  couplings constants, (Table 2.3), and a narrow range of temperature coefficients (between 4.6 and 6.9 ppb/ $^\circ$ ) for the hydroxyl groups, (Table 2.4) suggesting the presence of conformational averaging around the corresponding C-O bonds. Again, there is no strong intramolecular hydrogen bond between the hydroxyl moieties in the employed solvents, in contrast with the observations for protected fluorine-containing carbohydrates in non-polar solvents.<sup>[13]</sup> Long range coupling constants could also be detected, from the inspection of the  ${}^1\text{H}$  NMR spectrum, between the axial fluorine with H2 (medium), H4 (small) and H6' (2.8 Hz). Again, the corresponding heteronuclear  ${}^1\text{H}$ - ${}^{19}\text{F}$  NOEs have been observed for the  ${}^1\text{H}/{}^{19}\text{F}$  pairs in the HOESY spectrum (Fig. 2.8).



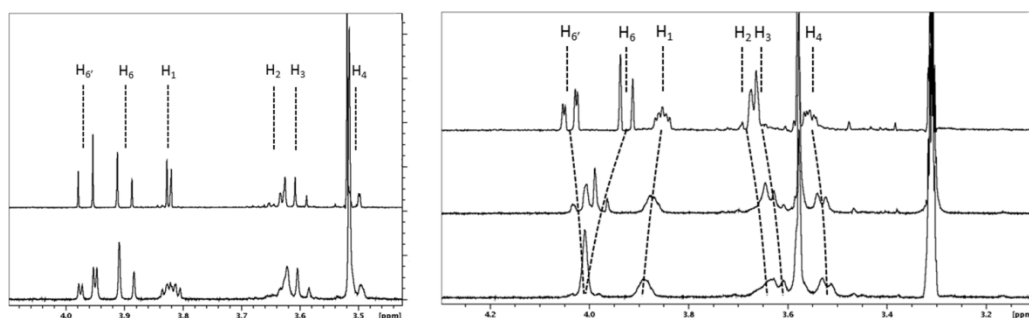
**Figure 2.8.** Top panel, schematic representation of the three major conformations present for **1a**. The arrows highlight the existence of NOE cross-peaks. No unique geometry is able to completely explain the NOE data, but a combination of all three represented structures does it.  $\omega$  is defined by (O6-C6-C5-C5a) and (O6-C6-C5-C4) torsion angles. The  ${}^4C_1$  gt conformer also justifies the H6'/Fax long range coupling constant given its relative W-like arrangement between the two coupled nuclei. Left, 2D NOESY spectra (700 ms mixing time) of compound **1a** at 600 MHz and 298 K in  $D_2O$ . Key NOEs are highlighted in the strip taken at H3 frequency. Right,  ${}^1H$ - ${}^{19}F$  HOESY spectrum (470/500 MHz Bruker spectrometer) of **1a** in  $D_2O$  in the presence of 20% methanol (800 ms mixing time).

Indeed, additional information on the geometry of **1a** has been obtained through the careful analysis of the  ${}^1H/{}^1H$  homo and  ${}^1H/{}^{19}F$  heteronuclear NOE experiments. Two key cross peaks have been observed for **1a**. There is a medium size H3/H6 NOE, and a weak H3/H6' one. Moreover, a clear heteronuclear H6/Feq NOE has been also observed, while the corresponding H6'/Feq NOE is rather weak. The vicinal couplings between the two H6 protons in **1a** and the corresponding OH6 are somehow different (2.6 Hz of difference). These findings are only compatible with the existence of conformational averaging in solution. The HOESY spectrum in DMSO (Fig. 2.9) showed the Fax/OH5 NOE, besides those observed in  $D_2O$  solution. No NOEs with other hydroxyl groups have been observed for the two  ${}^{19}F$  atoms.



**Figure 2.9.** On the top, HOESY 1D spectrum of compound **1a**, in DMSO- $d_6$ . Below, proton spectrum. The hydroxyl protons are labeled. The NOEs signals at upper field are related to ring protons. From lower to upper field: H1, H2, and H4 respectively. The NOE signal at lower field is in between the axial fluorine and the proton of the hydroxyl group 5.

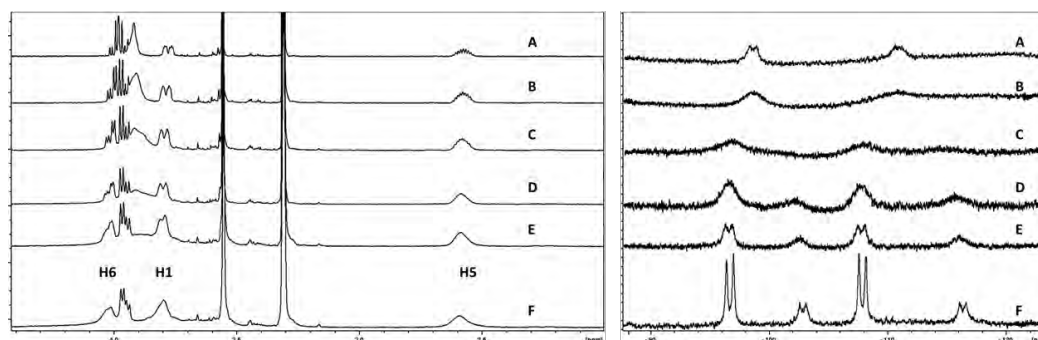
The observed data suggests that the differences in the  $^{19}\text{F}$  NMR spectra between **1b** and **1a** derive from a different dynamic behavior of both molecules. Nevertheless, still for **1a**, the relatively large  $^3J_{\text{H}_2\text{H}_3}$  and  $^3J_{\text{H}_3\text{H}_4}$  coupling constant values point out the existence of a major  $^4\text{C}_1$  conformation for the six-membered ring. Moreover, the simultaneous spatial proximity between the H3/H6 and Feq/H6 pairs, together with the through space coupling constant between Fax and H6' strongly suggest the existence of two orientations around C5-C6, as schematized in figure 2.8. In order to justify the weak H6'/Feq NOE, a minor contribution of the  $^2\text{S}_{5a}$  conformer has to be also considered (see below). Low-temperature NMR experiments in water solution (with 20% methanol) provided additional information on the nature of the conformational equilibrium. It has been observed that H1 shifted downfield upon decreasing temperature, while H2, H3 and H4 shifted upfield (Figure 2.10). Strikingly, H6 and H6' interchanged chemical shifts during the cooling process. As mentioned above, the axial  $^{19}\text{F}$  NMR signal became sharper at low temperature and no additional  $^{19}\text{F}$  signals appear.



**Figure 2.10.** On the left, (top)  $^1\text{H}\{-^{19}\text{F}\}$  decoupled spectrum. (bottom)  $^1\text{H}$  NMR spectrum of **1a** at 298 K in  $\text{D}_2\text{O}$ . Notice the solved H2, H4, H1 and H6' signals. On the right,  $^1\text{H}$  NMR spectra between 248 (bottom), 263 (middle) and 298 K (top) in  $\text{D}_2\text{O}$  in the presence of 20% methanol. Methanol has been used as internal reference. 500 MHz Bruker spectrometer. Notice that H1 move downfield upon decreasing temperature, while H2, H3 and H4 move upfield. Strikingly, H6 and H6' interchanged chemical shifts during the cooling process.

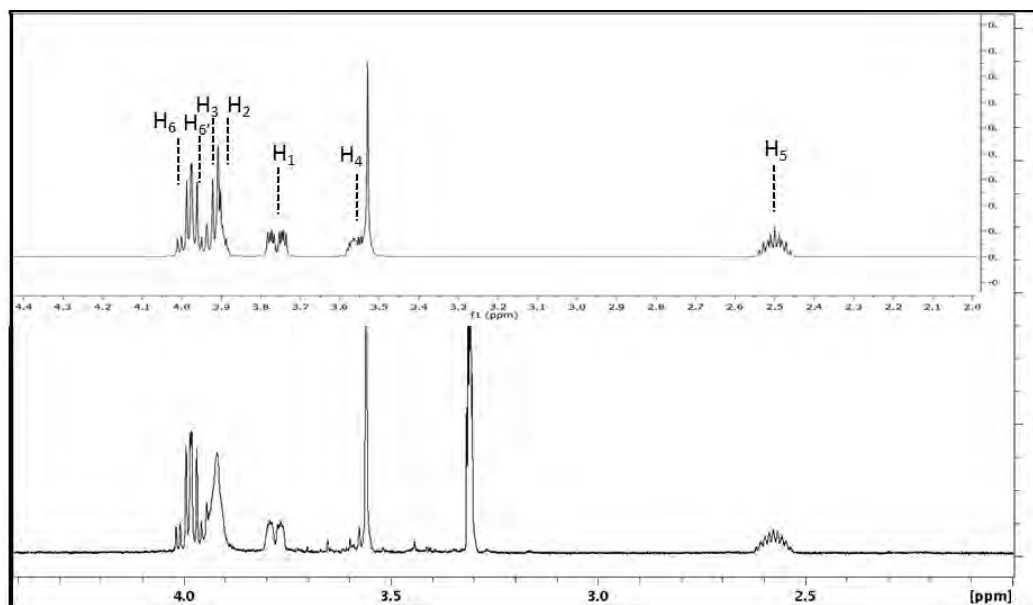
## Compound 2

The  $^{19}\text{F}$  NMR spectrum of **2** displayed two broad  $^{19}\text{F}$  NMR signals (Fig. 2.11) at room temperature. Interestingly, the signals sharpened in a noticeable manner upon decreasing the temperature and two new clear  $^{19}\text{F}$  NMR signals appeared at 253 K, using a 20% of methanol. The conformational equilibrium is slow in the  $^{19}\text{F}$  NMR chemical shift scale at this temperature. Using the observed coalescence temperature between 283 K and 298 K and the estimated chemical shift difference of the two set of fluorine signals at 253 K, the energy barrier was calculated to be ca.  $11.8 \pm 0.4 \text{ kcal}\cdot\text{mol}^{-1}$ .<sup>[14]</sup> The relative populations of the two signals were 70:30, indicating that the free energy difference was of only ca.  $0.4 \text{ kcal}\cdot\text{mol}^{-1}$ .



**Figure 2.11.** Variable temperature  $^1\text{H}$  NMR (left) and decoupled  $^{19}\text{F}\{-^1\text{H}\}$  NMR spectrum (right) of **2** in  $\text{D}_2\text{O}$  (with 20% methanol) at 500 ( $^1\text{H}$ ) or 470 ( $^{19}\text{F}$ ) MHz. A) 318 K; B) 298 K; C) 283 K; D) 273 K; E) 263 K; F) 253 K. Notice that the chemical shifts of the  $^1\text{H}$  NMR signals do not show major shifts with temperature.

Because of the severe proton overlapping, the  $J$  coupling constants were extracted from spectral simulation. (Fig. 2.12). The  ${}^3J_{H_2H_3}$  and  ${}^3J_{H_3H_4}$  (Table 2.1) values were intermediate (ca. 5.5 Hz for both) in solution, suggesting the presence of a conformational equilibrium, as shown by the two sets of  ${}^{19}\text{F}$  signals at 253 K. In contrast to the observations for **1a**, in this case, it was observed that none of the  ${}^1\text{H}$  NMR signals was significantly shifted at low temperature. As will be described below, this behavior also contains key conformational information.



**Figure 2.12.** Top. Simulated  ${}^1\text{H}$  NMR spectrum using Mestre Nova. Bottom  ${}^1\text{H}$  NMR spectrum of **2** in  $\text{D}_2\text{O}$  in the presence of 20% methanol. Methanol used as internal reference.

**Table 2.1.**  ${}^3J_{HH}$  and  ${}^nJ_{HF}$  coupling constants for **1a**, **1b**, **2** [Hz] in  $\text{D}_2\text{O}$  solution at 300 K and 500 MHz.

Molecule	${}^3J_{H_1H_2}$	${}^3J_{H_2H_3}$	${}^3J_{H_3H_4}$	${}^3J_{H_4H_5}$	${}^2J_{H_6H_6'}$	${}^3J_{H_1F_{ax}}$	${}^3J_{H_1F_{eq}}$	$J_{H_2F_{ax}}$	$J_{H_4F_{ax}}$	$J_{H_6F_{ax}}$	${}^4J_{H_6'F_{eq}}$	${}^2J_{F_{eq}F_{ax}}$
<b>1a</b>	4.1	9.0	9.5	-	12.5	3.8 <sup>§</sup>	7.3 <sup>§</sup>	3.3	2.1	<1	2.8	270.0
<b>1b</b>	3.6	10.0	10.0	-	12.5	1.7	3.7	2.5	3.4	1.2	-	278.0
<b>2</b>	3.3	5.5	5.5	4.3	11.5	2.0	9.0	2.8	2.8	n.a.	-	270.0
<b>Glc</b>	3.8	9.8	9.1	10.1	12.3	-	-	-	-	-	-	-
<b>Ido</b>	1.7	4.1	4.6	2.5	-	-	-	-	-	-	-	-

The observed values are in agreement with those expected for a very predominant  ${}^4\text{C}_1$  conformation for **1a** and **1b** isomers, while intermediate values for **2** are observed. The experimental  $J$ -coupling values for natural compounds are also provided. § Estimated from the corresponding  ${}^1\text{H}$  NMR spectroscopic signals.

**Table 2.2.**  $^1\text{H}$  and  $^{19}\text{F}$  NMR chemical shifts for **1a**, **1b**, **2** [Hz] in  $\text{D}_2\text{O}$  solution at 300 K and 500 MHz.

Molecule	$\delta\text{H1}$	$\delta\text{H2}$	$\delta\text{H3}$	$\delta\text{H4}$	$\delta\text{H5}$	$\delta\text{H6}$	$\delta\text{H6}'$	$\delta\text{Fax}$	$\delta\text{Feq}$	$\delta\text{Ome}$
<b>1a</b>	3.82	3.64	3.60	3.50	-	3.89	3.96	-109.10	-119.40	3.48
<b>1b</b>	3.75	3.54	3.72	3.42	-	3.66	3.76	-110.20	-119.40	3.47
<b>2</b>	3.80	3.90	3.92	3.55	2.58	4.00	3.96	-98.70	-110.0	3.56
<b>Glc</b>	4.79	3.54	3.65	3.38	3.63	3.85	3.74	-	-	3.50
<b>Ido</b>	4.69	3.53	3.73	3.75	4.09	3.79	3.82	-	-	3.45

Experimental chemical shift data for Methyl- $\beta$ -L-idopyranoside and Methyl- $\alpha$ -D-glucopyranoside are also provided.

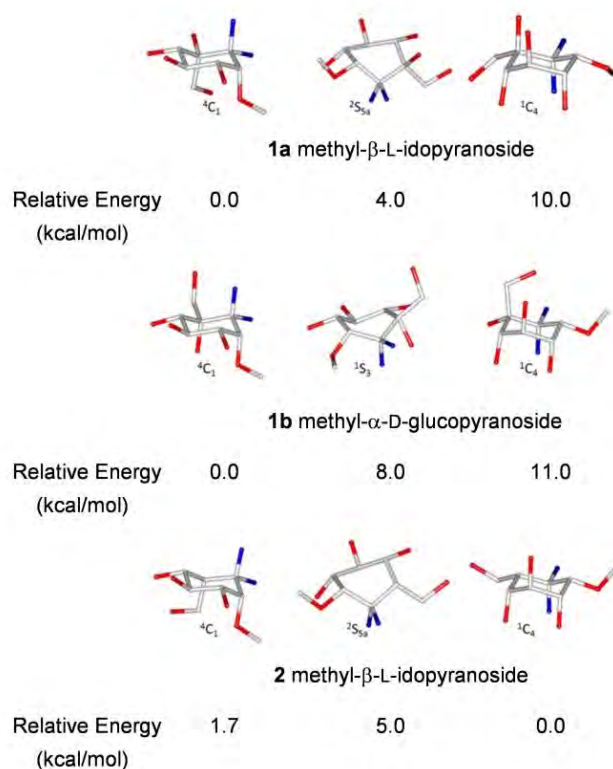
**Table 2.3.** Observed  $^3J_{\text{H},\text{OH}}$  coupling constants [Hz] in  $\text{DMSO-d}_6$  solution at 300 K and 500 MHz.

Molecule	Coupling constants [Hz]				
	$^3J_{\text{H4},\text{OH4}}$	$^3J_{\text{H2},\text{OH2}}$	$^3J_{\text{H3},\text{OH3}}$	$^3J_{\text{H6}',\text{OH6}}$	$^3J_{\text{H6},\text{OH6}}$
<b>1a</b>	5.5	4.1	5.6	4.6	7.2
<b>1b</b>	6.1	5.0	6.6	5.9	5.9

**Table 2.4.** Temperature coefficients measured for the different hydroxyl groups of **1a** and **1b**, from the analysis of the  $^1\text{H}$  NMR spectra recorded in  $\text{DMSO-d}_6$  between 298 and 343 K. The temperature coefficients of the fluorine atoms were deduced in  $\text{D}_2\text{O}$  using the same temperature range.

Molecule	$\Delta\delta/\Delta T$ (ppb/K)						
	$\text{F}_{\text{ax}}$	$\text{F}_{\text{eq}}$	$\text{OH}_4$	$\text{OH}_2$	$\text{OH}_3$	$\text{OH}_6$	$\text{OH}_5$
<b>1a</b>	8.4	17.2	6.9	6.5	6.5	5.1	4.6
<b>1b</b>	18	20	6.2	7.0	7.0	5.7	4.9

**2.2.2 Computational analysis.** The molecular modelling protocol described in the Methods section was adopted for structure assessment, providing different possible conformers for the molecules discussed here. Fig. 2.13.



**Figure 2.13.** Ball and stick representation and relative steric energy values of the major conformers of **1a**, **1b**, and **2**, according to DFT calculations.

For the Glc-configured **1b**, the theoretical calculations satisfactorily predicted the <sup>4</sup>C<sub>1</sub> as unique possible conformer. The other two local minima, <sup>1</sup>C<sub>4</sub> and <sup>1</sup>S<sub>3</sub>, displayed much larger relative free energy, ca. 11.0 kcal·mol<sup>-1</sup> and 8.0 kcal·mol<sup>-1</sup>, respectively. For this system, the transition state structures were also deduced. The activation energies to reach intermediate states were estimated as ca. 15.0 kcal·mol<sup>-1</sup> for the <sup>3</sup>E envelope and 13.5 kcal·mol<sup>-1</sup> for the <sup>5a</sup>H<sub>5</sub> half-chair shapes. (see SI Table 2.1)

For **2**, the alternative <sup>1</sup>C<sub>4</sub> chair was predicted as the global minimum, with the <sup>4</sup>C<sub>1</sub> chair destabilized in ca. 1.7 kcal·mol<sup>-1</sup>. In this case, the equatorial orientation of the two bulky groups is probably the driving force for the calculated energy values. In this case, the <sup>2</sup>S<sub>5a</sub> conformer displayed the less favorable energy value.



For **1a**, the  ${}^1C_4$  form was strongly destabilized with respect to the global minimum, the  ${}^4C_1$  conformer. The presence of several 1,3-diaxially oriented groups provides the structural basis for these predictions. The  ${}^2S_{5a}$  skew boat displays most of the bulky groups in pseudo equatorial orientations, resulting in a relative low destabilizing energy value ( $4.0 \text{ kcal}\cdot\text{mol}^{-1}$ ).

As mentioned above, the conformational behavior of idose rings and derivatives thereof has been a matter of investigation for years.<sup>[2,10,15,16]</sup> It is well known that for L-idopyranoses, the theoretically more favorable  ${}^1C_4$  chair displays three axially oriented hydroxyl groups, with the corresponding steric consequences. The alternative  ${}^4C_1$  chair places the bulky hydroxymethyl group at the axial orientation, with the corresponding collapse. Therefore, alternative skew-boat conformers are also present in the conformational equilibrium, depending on the hydroxyl substitution, chemical environment and solvent.<sup>[17,18,19]</sup> Therefore, the description of the conformational flexibility of these rings in terms of thermodynamic parameters as activation free energy, and entropic and enthalpy contributions to the Gibbs free energy difference represents a challenge to the experimental study. Such a description implies access to a detailed and accurate proton-proton coupling constant analysis, which is hampered by signal broadening and/or overlapping. In fact, the situation becomes even more arduous, and most of the times inaccessible, in cases of fast and medium conformational equilibria in the NMR chemical shift time scale, especially when it involves entropically favored isoforms such as skew boat conformers, which it is often the case of idose ring derivatives.

For ido-like sugars, flexibility is intrinsically related to biological activity.<sup>[20,21]</sup> The sulfated L-iduronic rings represent the paradigmatic example of how plasticity modulates the interaction with biological receptors. From the chemical perspective, these molecular recognition processes are the consequence of the balance between enthalpy and entropy factors and solvation/desolvation effects. In this context, since conformational entropy becomes an issue, it is essential to consider that the conformational entropy of chair and skew boat conformers is intrinsically different. Chair conformers are defined in well-characterized potential energy wells, while the conformational entropy of skew boat conformers is larger, due to the low-energy cost geometry interconversions that conduct to basically the same conformer. The energy well for skew boat conformers is much wider than that for the chairs.

Under these premises, the obtained results can now be accounted for in a satisfactory manner. Compound **1b** (Glc-like) displays exclusively the  ${}^4C_1$  chair conformer with a very well defined geometry, as in the natural compound. (Table 2.5). Compound **2**, similarly to the natural Ido-like molecule, displays significant conformational flexibility.  ${}^{19}\text{F}$ -based variable temperature experiments demonstrate the existence of a conformational equilibrium between two forms

with a ca. 70:30 population distribution. The observed coupling constants are in fact in agreement with a 70:30 distribution between the canonical  ${}^4C_1$  (minor) and  ${}^1C_4$  (major) chair forms (Table 2.6). No changes in either chemical shifts or coupling constants are observed with temperature, indicating that the conformational distribution is temperature-independent. Therefore, the conformational entropy of the contributing geometries is similar, as expected for the two alternative  ${}^4C_1$  and  ${}^1C_4$  chair conformers. The free energy difference between the two forms is ca.  $0.4 \text{ kcal}\cdot\text{mol}^{-1}$ , favoring the  ${}^1C_4$  chair in the same trend as the energy differences estimated by the calculations. The energy barrier for interconversion is relatively low (ca.  $11.8 \text{ kcal}\cdot\text{mol}^{-1}$ ), extremely difficult to access by variable temperature  ${}^1\text{H}$  NMR experiments in water solution. The use of  ${}^{19}\text{F}$ -NMR has permitted to access this value, due to the wide chemical shift difference between both conformers (ca 3 ppm, 1500 Hz), making possible, at low temperature, to reduce the exchange rate below such frequency thus appearing the system under slow exchange regime in the  ${}^{19}\text{F}$  chemical shift scale (it was not possible to determine the amplitude of proton chemical shift difference between both conformers but, as a gross approximation, comparing delta of protons H2 and H3 in compound **1b** ( ${}^4C_1$ ) and **2** (preferentially  ${}^1C_4$ ) the differences are in the order of 0.3 ppm or 150 Hz (table 2.2), one order of magnitude smaller than in the  ${}^{19}\text{F}$  chemical shifts). Indeed, computational chemistry calculations using ab initio methods (DFT (B3LYP) in vacuum with the different basis sets: 6-31++G, for the  ${}^4C_1$  and  ${}^1C_4$  geometries; 6-31++G +freq, for  ${}^2S_{5a}$ ; and 6-31++GTS +freq, for  ${}^3E$  and  ${}^5aH_5$ ), provided energy values fairly similar to those experimentally detected ( $12.5 \text{ kcal}\cdot\text{mol}^{-1}$  for  ${}^3E$  and  $13.0$  for  ${}^5aH_5$ ). (see SI Table 2.1).

In contrast, compound **1a** shows a singular conformational behavior. The chemical shift and coupling constant values drastically changed upon temperature variation (Fig. 2.10) indicating that its conformational distributions depend on the temperature. This observation strongly suggests that the conformational entropy of the contributing geometries is different. At low temperature, the enthalpy-favored conformer should be predominant, since the entropy contribution to free energy will be largely attenuated ( $\Delta G = \Delta H - T\Delta S$ ). Fittingly, H2, H3 and H4 shift upfield more than 0.1 ppm upon decreasing temperature. Concomitantly, H1 shifted downfield. This fact evidences that, at low temperature, the predominant conformer of **1a** displays H2, H3 and H4 in axial orientation, while H1 shows an equatorial arrangement. Therefore, the major and enthalpy-favored conformer is the  ${}^4C_1$  chair. The other participating conformer should display a skew boat geometry since its contribution to the conformational equilibrium strongly decreases at low temperature. Computational chemistry calculations found the  ${}^2S_{5a}$  conformer as the most stable skew boat form, with a relative energy of ca.  $4.0 \text{ kcal}\cdot\text{mol}^{-1}$  with respect to the

${}^4C_1$  chair. The  ${}^1C_4$  form was strongly destabilized, in more than  $10 \text{ kcal}\cdot\text{mol}^{-1}$ . In fact, for the  ${}^2S_{5a}$  conformer, H2, H3 and H4 display a quasi-axial orientation, providing large  ${}^3J_{H2H3}$  and  ${}^3J_{H3H4}$  couplings. Since these observed  ${}^3J_{H2H3}$  and  ${}^3J_{H3H4}$  couplings at room temperature for both molecules were already rather large (above 9 Hz), the contribution of the skew boat conformers would have been clearly neglected from the inspection of the  ${}^1H$  and  ${}^{13}C$  NMR spectra, unless the  ${}^{19}F$  NMR spectra would not have shown dramatically broad signals. (Table 2.7).

Interestingly, the conformational behavior of these  ${}^{19}F$ -containing glycomimetics remarkably resembles the intrinsic flexibility of the natural Ido-configured sugars. Although this fact might not be completely surprising, as a matter of fact, regular Ido-like carbasugars,<sup>[22]</sup> with a  $CH_2$  group mimicking the endocyclic oxygen, did not show any conformational plasticity (see SI Table 2.1). In contrast, the molecules presented herein, with  $CF_2$  moieties, show important conformational plasticity resembling the behavior of endocyclic oxygen in natural carbohydrates. The dynamic process has been quantified in terms of energy barriers and free energy differences. The destabilizing energies for the  ${}^4C_1$  conformer in **2** is  $0.4 \text{ kcal}\cdot\text{mol}^{-1}$  above that for the  ${}^1C_4$  chair. However, when C5 is modified (as in **1a**, with one additional OH substituent), there is a participation of skew boat conformers, while the proportion of the  ${}^1C_4$  conformer is strongly diminished. The OH3 group displays an equatorial orientation in the  ${}^4C_1$  and  ${}^2S_{5a}$  geometries, minimizing the influence of steric conflicts with the additional OH5. However, it would adopt an axial disposition in the  ${}^1C_4$  form, provoking important additional steric clashes. For the Glc-like molecules, the conformational behavior of the  $CF_2$  analogue also mimics that of natural glucopyranosides. Therefore, these glycomimetics can behave as conformational bioisosters.

**Table 2.5.** Experimental and calculated  ${}^3J_{HH}$  and  ${}^nJ_{HF}$  coupling constants for different conformers of **1b** [Hz] in D<sub>2</sub>O solution at 300 K and 500 MHz. The observed values are in agreement with those expected for a unique  ${}^4C_1$  conformation.

Conformer	${}^3J_{H1H2}$	${}^3J_{H2H3}$	${}^3J_{H3H4}$	${}^2J_{H6H6'}$	${}^3J_{H1Fax}$	${}^3J_{H1Fecq}$	$J_{H2Fax}$	$J_{H4Fax}$	$J_{H6Fax}$	${}^4J_{H6Fecq}$	$J_{H6'Fax}$	$J_{H2Fecq}$	$J_{H4Fecq}$	$J_{H6'Fecq}$	${}^2J_{FecqFax}$
${}^1C_4$	3.6	3.6	3.1	13.3	3.8	6.5	3.9	4.3	0.5	2.0	1.3	2.2	0.4	4.3	255.0
${}^4C_1$	4.1	9.2	9.1	12.8	1.5	2.6	2.6	2.4	1.6	0.5	0.3	1.9	1.2	0.5	280.0
${}^1S_3$	8.5	9.3	6.8	8.3	7.6	7.1	0.3	2.6	1.7	1.2	1.0	0.3	0.9	1.8	239.0
<b>Experimental</b>	3.6	10.0	10.0	12.5	1.7	3.7	2.5	3.4	1.2	n.a	n.a	2.0	n.a	n.a	278.0

**Table 2.6.** Experimental and calculated  ${}^3J_{HH}$  and  ${}^nJ_{HF}$  coupling constants for different conformers and ensemble average structure of **2** [Hz] in D<sub>2</sub>O solution at 300 K and 500 MHz. The observed values are in agreement with those expected for an 70:30 average  ${}^1C_4$ : ${}^4C_1$  conformation.

Conformer	${}^3J_{H1H2}$	${}^3J_{H2H3}$	${}^3J_{H3H4}$	${}^3J_{H4H5}$	${}^3J_{H5H6}$	${}^3J_{H5H6'}$	${}^3J_{H5Fecq}$	${}^3J_{H1Fecq}$	${}^2J_{H6H6'}$	${}^3J_{H1Fax}$	${}^3J_{H5Fax}$	$J_{H2Fax}$	$J_{H2Fecq}$	$J_{H4Fax}$	$J_{H4Fecq}$
${}^1C_4$	3.0	3.0	2.9	1.8	5.5	7.5	22.5	13.5	10.2	0.3	0.6	5.1	2.0	3.8	0.4
${}^4C_1$	3.9	9.2	9.1	8.6	5.2	7.2	2.9	3.6	13.2	2.4	1.5	2.6	2.5	0.7	1.7
${}^1C_4(70\%):{}^4C_1(30\%)$	3.3	5.0	4.7	3.8	5.4	7.4	16.6	10.5	11.1	1.0	0.9	4.3	2.0	2.8	0.8
<b>Experimental</b>	3.3	5.5	5.5	4.3	5.3	6.5	14.0	9.0	11.5	2.0	5.0	2.8	n.a	2.8	1.5

**Table 2.7.** Experimental and calculated  ${}^3J_{HH}$  and  ${}^nJ_{HF}$  coupling constants for different conformers and ensemble average structure of **1a** [Hz] in D<sub>2</sub>O solution at 300 K and 500 MHz. The observed values are in agreement with those expected for an average  ${}^4C_1$ : ${}^2S_{5a}$  conformation.

Conformer	${}^3J_{H1H2}$	${}^3J_{H2H3}$	${}^3J_{H3H4}$	${}^2J_{H6H6'}$	${}^3J_{H1Fax}$	${}^3J_{H1Fecq}$	$J_{H2Fax}$	$J_{H4Fax}$	$J_{H6Fax}$	${}^4J_{H6Fecq}$	$J_{H6'Fax}$	$J_{H2Fecq}$	$J_{H4Fecq}$	$J_{H6'Fecq}$	${}^2J_{FecqFax}$
${}^1C_4$	3.0	3.0	2.9	8.1	2.6	9.2	4.2	4.1	1.4	1.4	0.8	2.4	0.8	0.0	246.0
${}^4C_1$	3.9	9.2	9.1	12.3	0.5	4.5	2.3	3.0	1.0	0.4	4.2	1.7	1.4	0.2	256.0
${}^2S_{5a}$	5.3	8.0	9.6	14.4	8.4	12.4	6.2	1.5	0.4	0.5	0.1	1.6	0.9	4.6	230.0
${}^4C_1(75\%):{}^2S_{5a}(25\%)$	4.2	8.9	9.2	12.9	2.5	6.5	3.4	2.6	0.8	0.4	3.2	1.7	1.2	1.3	248.0
<b>Experimental</b>	4.1	9.0	9.5	12.5	3.8	7.3	3.3	2.1	<1	0.0	2.8	2.0	1.2	0.0	270.0

## 2.3 Conclusions

A new generation of fluorine-containing glycomimetics is presented. The importance of introducing fluorine atoms in these glycomimetics is also highlighted. First, only the use of  $^{19}\text{F}$  NMR experiments has permitted to detect a dynamic process of paramount significance that would have been otherwise remained unobserved. Additionally, only in the presence of fluorine atoms at C5a, the Ido-like six-membered ring recovers its required flexibility, absent in regular  $\text{CH}_2$ -Ido-carbasugars,<sup>[11]</sup> while the presence of a bulky substituent at position C5 strongly reduces the ring flexibility and introduces important steric clashes. Herein, we demonstrate that the presence of the fluorine in the ring restores the plasticity of Ido-like six-membered rings.

Thus, the combination of NMR experiments and computational methods has permitted to show that these idose-like analogues resemble the conformational plasticity of the natural parent molecules that is anticipated to be required for key molecular recognition process and ultimately for biological activity.

## 2.4 Methods

**2.4.1 NMR Spectroscopy:**  $^{19}\text{F}$  NMR experiments were performed at 470 MHz with a Bruker AVANCE spectrometer equipped with the proper fluorine probe SEF, at 298 K unless otherwise stated while low temperature experiments were done with Bruker DRX 500 MHz equipped with BBOF plus probe.  $^1\text{H}$  NMR experiments were performed at 600 and 700 MHz with a Bruker AVANCE spectrometer equipped with TXI probe. Experiments were performed in  $\text{D}_2\text{O}$ ,  $\text{DMSO-d}_6$  and in  $\text{D}_2\text{O}$  in the presence of 20% methanol for low temperature analysis. The concentration employed was 2mM for all the discussed molecules. In addition to standard 1D  $^1\text{H}$  NMR spectra, COSY, TOCSY, NOESY and HOESY (800 ms mixing time) and  $^1\text{H}/^{13}\text{C}$  HSQC experiments based on the standard BRUKER sequences were also acquired, in order to assign the resonance of all NMR signals. Because of the severe proton overlapping, the  $J$  coupling constants were extracted from spectral simulation using MestreNova software. The method of determining activation energy parameters is through the estimation of the coalescence temperature and chemical shift difference for each fluorine signals measured in the spectrum at lower temperature (253 K) giving 2900 Hz for low field signals and 4010 Hz for higher field signals.<sup>[14]</sup> The exchange rate for both fluorine signals can be now estimated applying the equation:

—

Observing the fluorine NMR spectra at different temperature (Figure 2.11) it was possible to enclose the activation energy barrier between the limiting values calculated for 283 K and 298 K applying the Eyring equation:

$$\ln \frac{k_c}{k} = \frac{E_a}{RT_c} - \ln \left( \frac{h}{k_B T_c} \right)$$


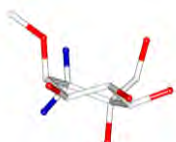

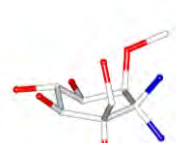

Where  $R$  is the gas constant,  $T_c$  the coalescence temperature,  $k$  the Boltzmann constant,  $h$  the Planck constant and  $k_c$  the determined exchange rate.


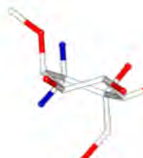

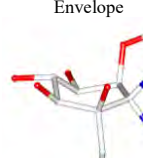

With this equation the activation energy barrier relative to the low field signal lies between the limiting values of 11.6 kcal·mol<sup>-1</sup> at 283 K and 12.2 kcal·mol<sup>-1</sup> at 298 K, giving an average value of 11.9 kcal·mol<sup>-1</sup> at the estimated coalescence temperature of 290 K, while for high field signal it lies between 11.4 and 12.0 kcal·mol<sup>-1</sup> with an average value of 11.7 kcal·mol<sup>-1</sup> at 290 K. The final estimation for activation energy is  $\Delta G^\ddagger = 11.8 \text{ kcal}\cdot\text{mol}^{-1} \pm 0.4$ .


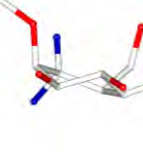
**2.4.2 Computational Methods:** A conformational search on these molecules **1a**, **1b**, and **2** was performed by using MacroModel at the Maestro suite of programs, with the MM3\* force field. Different local minima conformers were chosen within a conservative 20 kcal·mol<sup>-1</sup> threshold from the global minimum. Their expected coupling constant values and proton-proton distances (related to NOEs) were estimated from the corresponding structures using Maestro. The transition-state geometries for the interconversion process were chosen based on the well known Cremer-Pople sphere conformational routes. Then, the MM3\*-optimized structures were used as starting conformations for additional calculations. Thus, density functional theory (DFT) geometry optimizations were performed with the Gaussian 03 program using the hybrid B3LYP functional and the 6-31++G(d,p) basis set followed by vibrational frequency analysis. This protocol allowed assessing whether the optimized structures were true energy minima, transition states, or saddle points. For the transition states structures, the final geometry optimization was achieved by applying the TS Bery algorithm. In all cases, the presence of solvent was accounted for by the integral equation formalism polarizable continuum model (IEFPCM). The NMR isotropic shielding constants were calculated using the standard Gauge Independent Atomic Orbital (GIAO) approach. The experimental and calculated NMR coupling constants were then compared.

## 2.5 Supporting Information


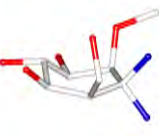

**Table 2.1 SI.** Optimized Conformations and Relative Thermodynamic Stabilities (related to the lowest energy conformer) of Compounds **1a-1b-2** and of Compounds **Ido** (Methyl- $\beta$ -L-Idopyranoside) and **CH<sub>2</sub>-Ido** (Carbasugar). DFT calculations with B3LYP/6-31++G(d,p) indicates minor ring flexibility for methylene carbasugars respect to fluoro-carbasugars. In the case of the natural sugar, Methyl- $\beta$ -L-Idopyranoside, the plasticity of the ring is evidenced.





Structure	Ring conformation	Flap atoms	$\Delta G^{\circ}_{m,calc}$ (kcal/mol)	Selected bond lengths (Å)	Selected atoms charge
<b>1a</b> <sup>4</sup> C <sub>1</sub>		C4; C1	0.0	Fax-C5a 1.43956; Feq-C5a 1.41316 C5a-C1 1.53842; C1-C2 1.54228 C2-C3 1.53565; C3-C4 1.52543 C4-C5 1.54362; C5-C6 1.54701 C5-C5a 1.53889; C1-O1 1.44815 C5-O5 1.45741	Fax -0.342323; Feq -0.326481 C5a 0.344813; C1 -0.011405 C5 0.170801; C6 -0.200980 O1 -0.334405; O5 -0.585938 O6 -0.510305
		Imaginary Frequency			
		0			
<b>1a</b> <sup>5a</sup> H <sub>5</sub>		C5a; C5	13.5	Fax-C5a 1.45058; Feq-C5a 1.41583 C5a-C1 1.55688; C1-C2 1.53422 C2-C3 1.52607; C3-C4 1.51820 C4-C5 1.55686; C5-C6 1.55093 C5-C5a 1.58954; C1-O1 1.44813 C5-O5 1.46562	Fax -0.218683; Feq -0.132654 C5a 0.197905; C1 -0.165365 C5 -0.022661; C6 -0.271305 O1 -0.254422; O5 -0.380270 O6 -0.519556
		Imaginary Frequency			
		1			
<b>1a</b> <sup>2</sup> S <sub>5a</sub>		C2; C5a	4.0	Fax-C5a 1.35723; Feq-C5a 1.35514 C5a-C1 1.53712; C1-C2 1.54693 C2-C3 1.55376; C3-C4 1.53638 C4-C5 1.56245; C5-C6 1.54561 C5-C5a 1.55756; C1-O1 1.43879 C5-O5 1.44583	Fax -0.331818; Feq -0.283616 C5a 0.396607; C1 -0.395338 C5 0.121481; C6 -0.209145 O1 -0.382483; O5 -0.531553 O6 -0.571920
		Imaginary Frequency			
		0			
<b>1a</b> <sup>3</sup> E		C3	14.0	Fax-C5a 1.45067; Feq-C5a 1.41571 C5a-C1 1.55676; C1-C2 1.53407 C2-C3 1.52610; C3-C4 1.51821 C4-C5 1.55699; C5-C6 1.55092 C5-C5a 1.58974; C1-O1 1.44812 C5-O5 1.46570	Fax -0.21859; Feq -0.132500 C5a 0.19515; C1 -0.16671 C5 -0.023495; C6 -0.270054 O1 -0.254093; O5 -0.380422 O6 -0.519513
		Imaginary Frequency			
		1			
<b>1a</b> <sup>1</sup> C <sub>4</sub>		C1; C4	10.0	Fax-C5a 1.38720; Feq-C5a 1.38334 C5a-C1 1.53665; C1-C2 1.54124 C2-C3 1.54213; C3-C4 1.54850 C4-C5 1.56387; C5-C6 1.53463 C5-C5a 1.54567; C1-O1 1.41709 C5-O5 1.43459	Fax -0.315879; Feq -0.284173 C5a 0.716041; C1 -0.512203 C5 -0.467328; C6 +0.167238 O1 -0.269825; O5 -0.531553 O6 -0.660548
		Imaginary Frequency			
		0			



<b>1b</b> ${}^4C_1$	Chair 	C4; C1	0.0	Fax-C5a 1.35706; Feq-C5a 1.35560 C5a-C1 1.54588; C1-C2 1.53954 C2-C3 1.54211; C3-C4 1.54389 C4-C5 1.54414; C5-C6 1.54461 C5-C5a 1.55109; C1-O1 1.43734 C5-O5 1.44214	Fax -0.269793; Feq -0.341537 C5a 0.416180; C1 -0.161477 C5 0.010982; C6 0.239725 O1 -0.248329; O5 -0.513954 O6 - 0.607389
		Imaginary Frequency			
		0			
<b>1b</b> ${}^5aH_5$	Half Chair 	C5a; C5	13.5	Fax-C5a 1.42635; Feq-C5a 1.42178 C5a-C1 1.54274; C1-C2 1.56530 C2-C3 1.52390; C3-C4 1.52624 C4-C5 1.57202; C5-C6 1.53262 C5-C5a 1.55224; C1-O1 1.44121 C5-O5 1.43986	Fax -0.187195; Feq -0.139305 C5a 0.336574; C1 -0.181908 C5 -0.542419; C6 -0.131974 O1 -0.140100; O5 -0.393859 O6 - 0.668517
		Imaginary Frequency			
		1			
<b>1b</b> ${}^1C_4$	Chair 	C1; C4	11.0	Fax-C5a 1.35305; Feq-C5a 1.35576 C5a-C1 1.54863; C1-C2 1.53963 C2-C3 1.53732; C3-C4 1.54383 C4-C5 1.54051; C5-C6 1.53709 C5-C5a 1.52492; C1-O1 1.44089 C5-O5 1.41001	Fax -0.373125; Feq -0.298781 C5a 0.811003; C1 -0.151437 C5 0.313122; C6 0.046358 O1 -0.375614; O5 -0.566436 O6 - 0.651274
		Imaginary Frequency			
		0			
<b>1b</b> ${}^3E$	Envelope 	C3	15.0	Fax-C5a 1.44420; Feq-C5a 1.41051 C5a-C1 1.56234; C1-C2 1.57107 C2-C3 1.53472; C3-C4 1.52026 C4-C5 1.53581; C5-C6 1.55698 C5-C5a 1.54155; C1-O1 1.44034 C5-O5 1.45677	Fax -0.117127; Feq -0.161451 C5a -0.102798; C1 -0.552383 C5 0.354594; C6 0.155486 O1 -0.327271; O5 -0.487561; O6 - 0.503988
		Imaginary Frequency			
		1			
<b>1b</b> ${}^1S_3$	Skew boat 	C1; C3	8.0	Fax-C5a 1.35687; Feq-C5a 1.35827 C5a-C1 1.55281; C1-C2 1.55596 C2-C3 1.53445; C3-C4 1.54534 C4-C5 1.56233; C5-C6 1.54124 C5-C5a 1.54948; C1-O1 1.44253 C5-O5 1.45100	Fax -0.343562; Feq -0.303648 C5a 0.931232; C1 -0.112874 C5 -0.152003; C6 -0.081032 O1 -0.381354; O5 -0.542436 O6 - 0.651124
		Imaginary Frequency			
		0			

<b>2</b> ${}^1C_4$	Chair 	C1; C4	0.0	Fax-C5a 1.35305; Feq-C5a 1.35576 C5a-C1 1.54863; C1-C2 1.53963 C2-C3 1.53732; C3-C4 1.54382 C4-C5 1.54051; C5-C6 1.53708 C5-C5a 1.52492; C1-O1 1.44089	Fax -0.357254; Feq -0.299348 C5a 0.374882; C1 -0.024751 C5 0.289206; C6 -0.258116 O1 -0.378896; O6 -0.612554
		Imaginary Frequency			
		0			
<b>2</b> ${}^5aH_5$	Half Chair 	C5a; C5	13.0	Fax-C5a 1.44077; Feq-C5a 1.39573 C5a-C1 1.55454; C1-C2 1.53422 C2-C3 1.51803; C3-C4 1.52620 C4-C5 1.54884; C5-C6 1.54923 C5-C5a 1.59152; C1-O1 1.45012	Fax -0.322553; Feq -0.313336 C5a 0.354992; C1 -0.030081 C5 0.290105; C6 -0.261114 O1 -0.376789; O6 -0.62253
		Imaginary Frequency			
		1			



$2^4C_1$	Chair 	C4; C1	1.7	Fax-C5a 1.35713; Feq-C5a 1.35556 C5a-C1 1.54591; C1-C2 1.53950 C2-C3 1.54215; C3-C4 1.543882 C4-C5 1.54410; C5-C6 1.54453 C5-C5a 1.55110; C1-O1 1.43732	Fax -0.309334; Feq -0.333848 C5a 0.227588; C1 -0.043661 C5 0.148949; C6 -0.251668 O1 -0.318747; O6 -0.536749
		Imaginary Frequency			
		0			
$2^3E$	Envelop 	C3	12.5	Fax-C5a 1.44288; Feq-C5a 1.42684 C5a-C1 1.550221; C5a-C1 1.55767; C1-C2 1.52994; C2-C3 1.52803; C3- C4 1.51602; C4-C5 1.554989; C5-C6 1.550844; C5-C5a 1.588947; C1-O1 1.44784	Fax -0.22464; Feq -0.172542 C5a 0.207588; C1 -0.144362 C5 0.153628; C6 -0.261586 O1 -0.289446; O6 -0.516738
		Imaginary Frequency			
		1			
$2^2S_{5a}$	Skew boat 	C2; C5a	5.0	Fax-C5a 1.35723; Feq-C5a 1.35514 C5a-C1 1.53712; C1-C2 1.54693 C2-C3 1.55376; C3-C4 1.53638 C4-C5 1.56245; C5-C6 1.54561 C5-C5a 1.55756; C1-O1 1.43879	Fax -0.335046; Feq -0.305884 C5a 0.181858; C1 -0.191234 C5 0.125636; C6 -0.280518 O1 -0.384653; O6 -0.571941
		Imaginary Frequency			
		0			

Structure	Ring conformation	Flap atoms	$\Delta G^{\circ}_{m,calc}$ (kcal/mol)	Selected bond lengths (Å)	Selected atoms charge
<b>Ido</b> $^4C_1$	Chair 	C4; C1	0.3	O5-C1 1.41375; C1-C2 1.53701 C2-C3 1.53018; C3-C4 1.52566 C4-C5 1.53987; C5-C6 1.53648 C5-O5 1.44222; C1-O1 1.41240	O1 -0.324480; O5 -0.238686; C1 -0.031640 C2 -0.365494; C3 0.320474 C4 -0.216140; C5 0.038237 C6 -0.233577
		Imaginary Frequency			
		0			
<b>Ido</b> $^1S_3$	Skew boat 	C1; C3	3.0	O5-C1 1.42055; C1-C2 1.55182 C2-C3 1.53199; C3-C4 1.52718 C4-C5 1.55888; C5-C6 1.53436 C5-O5 1.43907; C1-O1 1.38965	O1 -0.357819; O5 -0.353587; C1 0.158253 C2 -0.169525; C3 0.174443 C4 -0.198965; C5 -0.233237 C6 -0.024668
		Imaginary Frequency			
		0			
<b>Ido</b> $^1C_4$	Chair 	C1; C4	0.0	O5-C1 1.42108; C1-C2 1.53479 C2-C3 1.53561; C3-C4 1.54573 C4-C5 1.53965; C5-C6 1.52276 C5-O5 1.43523; C1-O1 1.39676	O1 -0.344493; O5 -0.349919; C1 0.086719 C2 -0.129037; C3 0.087769 C4 -0.062110; C5 -0.349550 C6 -0.060004
		Imaginary Frequency			
		0			
<b>CH<sub>2</sub>Ido</b> $^1C_4$	Chair 	C1; C4	12.0	Hax-C5a 1.09470; Heq-C5a 1.09531 C5a-C1 1.53784; C1-C2 1.53383 C2-C3 1.54095; C3-C4 1.54550 C4-C5 1.55883; C5-C6 1.53162 C5-C5a 1.53415; C1-O1 1.43243 C5-O5 1.444874	Hax 0.132963 Heq 0.206132 C5a -0.574254 C1 -0.545914 C5 0.442489 C6 -0.220086 O1 -0.322411 O5 -0.530056 O6 -0.554663
		Imaginary Frequency			
		0			

<b>CH<sub>2</sub>Ido</b> <sup>4</sup> C <sub>1</sub>	Chair 	C4; C1	0.0	Hax-C5a 1.09665; Heq-C5a 1.09301 C5a-C1 1.53914; C1-C2 1.53599 C2-C3 1.53457; C3-C4 1.52571 C4-C5 1.53763; C5-C6 1.54539 C5-C5a 1.53865; C1-O1 1.43229 C5-O5 1.44325	Hax 0.179992 Heq 0.143318 C5a -0.430828 C1 0.282794 C5 -0.165423 C6 -0.122076 O1 -0.343676; O5 -0.474192; O6 -0.441194
		Imaginary Frequency			
		0			
<b>CH<sub>2</sub>Ido</b> <sup>2</sup> S <sub>5a</sub>	Skew boat 	C2;C5a	6.4	Hax-C5a 1.09732; Heq-C5a 1.09486 C5a-C1 1.54804; C1-C2 1.54823 C2-C3 1.53063; C3-C4 1.52714 C4-C5 1.55832; C5-C6 1.54633 C5-C5a 1.53777; C1-O1 1.43439 C5-O5 1.44513	Hax 0.136309 Heq 0.209956 C5a -0.466162 C1 -0.151929 C5 -0.073664 C6 0.056863 O1 -0.324407 O5 -0.410273; O6 -0.457687
		Imaginary Frequency			
		0			

## 2.6 References

1. Mayes, H. B., Broadbelt, L. J. & Beckham, G. T. How sugars pucker: Electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis. *J. Am. Chem. Soc.* **136**, 1008–1022 (2014).
2. Casu, B. *et al.* Controversial glycosaminoglycan conformations. *Nature* **322**, 215–216 (1986).
3. Plazinski, W., Drach, M. & Plazinska, A. Ring inversion properties of 1→2, 1→3 and 1→6-linked hexopyranoses and their correlation with the conformation of glycosidic linkages. *Carbohydr. Res.* **423**, 43–48 (2016).
4. Turnbull, J., Powell, A. & Guimond, S. Heparan sulfate: decoding a dynamic multifunctional cell regulator. *Trends Cell Biol.* **11**, 75–82 (2016).
5. Li, W., Johnson, D. J. D., Esmon, C. T. & Huntington, J. A. Structure of the antithrombin-thrombin-heparin ternary complex reveals the antithrombotic mechanism of heparin. *Nat Struct Mol Biol.* **11**, 857–862 (2004).
6. Canales, A. *et al.* Solution NMR structure of a human FGF-1 monomer, activated by a hexasaccharide heparin-analogue. *FEBS J.* **273**, 4716–4727 (2006).
7. Das, S. K. *et al.* Synthesis of Conformationally Locked Carbohydrates: A Skew-Boat Conformation of L-Iduronic Acid Governs the Antithrombotic Activity of Heparin. *Angew. Chemie Int. Ed.* **40**, 1670–1673 (2001).
8. Sattelle, B. M., Shakeri, J. & Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **14**, 1149–1159 (2013).
9. Ronnols, J., Manner, S., Siegbahn, A., Ellervik, U. & Widmalm, G. Exploration of conformational flexibility and hydrogen bonding of xylosides in different solvents, as a model system for enzyme active site interactions. *Org. Biomol. Chem.* **11**, 5465–5472 (2013).
10. Ronnols, J., Manner, S., Ellervik, U. & Widmalm, G. Conformational effects due to stereochemistry and C3-substituents in xylopyranoside derivatives as studied by NMR spectroscopy. *Org. Biomol. Chem.* **12**, 8031–8035 (2014).
11. Sattelle, B. M. *et al.* The dependence of pyranose ring puckering on anomeric configuration: methyl idopyranosides. *J. Phys. Chem. B* **116**, 6380–6386 (2012).
12. Roslund, M. U., Tähtinen, P., Niemitz, M. & Sjöholm, R. Complete assignments of the <sup>1</sup>H and <sup>13</sup>C chemical shifts and J<sub>H,H</sub> coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides. *Carbohydr. Res.* **343**, 101–112 (2008).
13. Giuffredi, G. T., Gouverneur, V. & Bernet, B. Intramolecular OH⋯FC Hydrogen Bonding in Fluorinated Carbohydrates: CHF is a Better Hydrogen Bond Acceptor than CF<sub>2</sub>. *Angew. Chemie Int. Ed.* **52**, 10524–10528 (2013).
14. Feigel, M. *Dynamic NMR Spectroscopy.* (Ed. Von J. Sandström. Academic Press, London) **1982**.
15. Snyder, J. R. & Serianni, A. S. D-Idose: a one- and two-dimensional NMR investigation of solution composition and conformation. *J. Org. Chem.* **51**, 2694–2702 (1986).

16. Canales, A. *et al.* Conformational Flexibility of a Synthetic Glycosylaminoglycan Bound to a Fibroblast Growth Factor. FGF-1 Recognizes Both the 1C4 and 2SO Conformations of a Bioactive Heparin-like Hexasaccharide. *J. Am. Chem. Soc.* **127**, 5778–5779 (2005).
17. Sattelle, B. M., Hansen, S. U., Gardiner, J. & Almond, A. Free Energy Landscapes of Iduronic Acid and Related Monosaccharides. *J. Am. Chem. Soc.* **132**, 13132–13134 (2010).
18. Angulo, J., Nieto, P. M. & Martin-Lomas, M. A molecular dynamics description of the conformational flexibility of the L-iduronate ring in glycosaminoglycans. *Chem. Commun.* **13**, 1512–1513 (2003).
19. Biarnés, X. *et al.* The Conformational Free Energy Landscape of  $\beta$ -d-Glucopyranose. Implications for Substrate Preactivation in  $\beta$ -Glucoside Hydrolases. *J. Am. Chem. Soc.* **129**, 10686–10693 (2007).
20. Hricovini, M. Solution Structure of Heparin Pentasaccharide: NMR and DFT Analysis. *J. Phys. Chem. B* **119**, 12397–12409 (2015).
21. Hricovini, M., Driguez, P. A. & Malkin, O. L. NMR and DFT Analysis of Trisaccharide From Heparin Repeating-Sequence. *J. Phys. Chem. B.* **118**, 11931–11942 (2014).
22. Säwén, E., Roslund, M. U., Cumpstey, I. & Widmalm, G. Synthesis and conformational analysis of carbasugar bioisosteres of  $\alpha$ -l-iduronic acid and its methyl glycoside. *Carbohydr. Res.* **345**, 984–993 (2010).

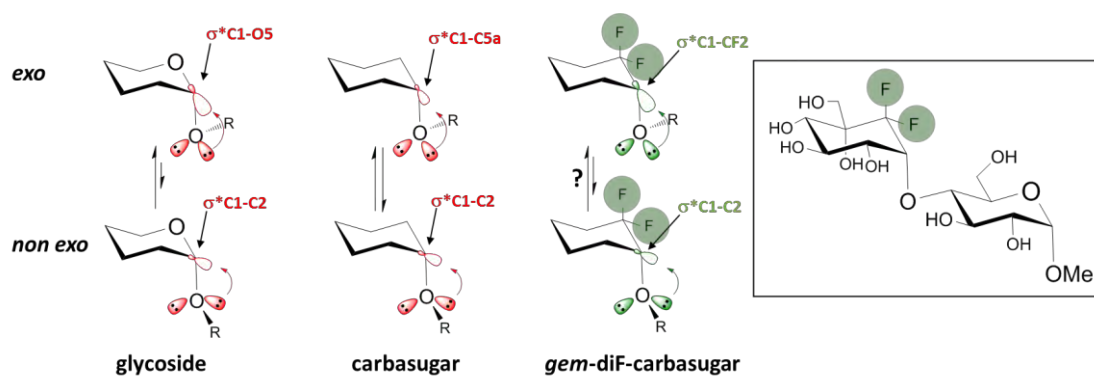


## CHAPTER III

*gem*-diF-carbadiisaccharides: the importance of the stereoelectronic component for the *exo*-anomeric effect

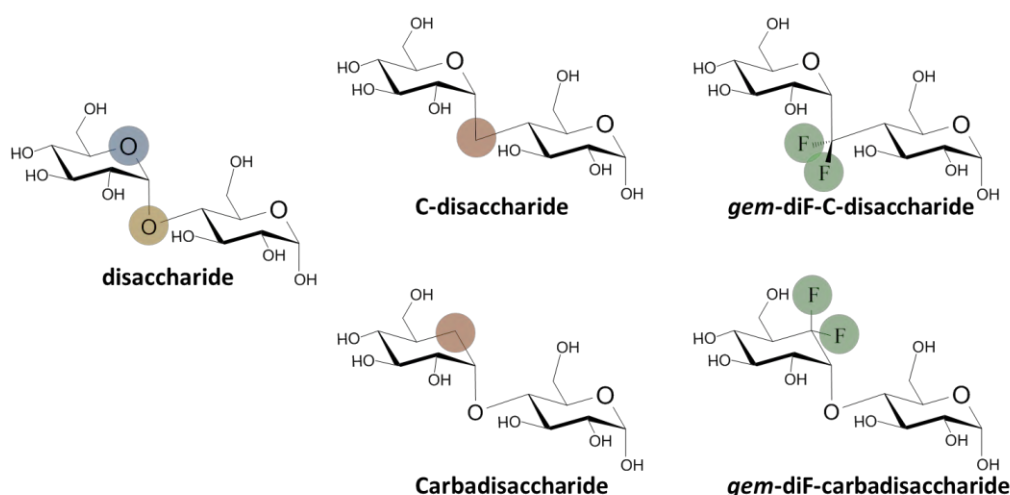
### *gem*-Difluorocarbadiisaccharides: Restoring the *exo*-Anomeric Effect

The work presented in this chapter has been performed in collaboration with the group of Dr. Matthieu Sollogoub, Sorbonne Université in Paris, France. Dr. Bixue Xu has performed the synthesis of the glycomimetics discussed herein.



### 3.1 Introduction

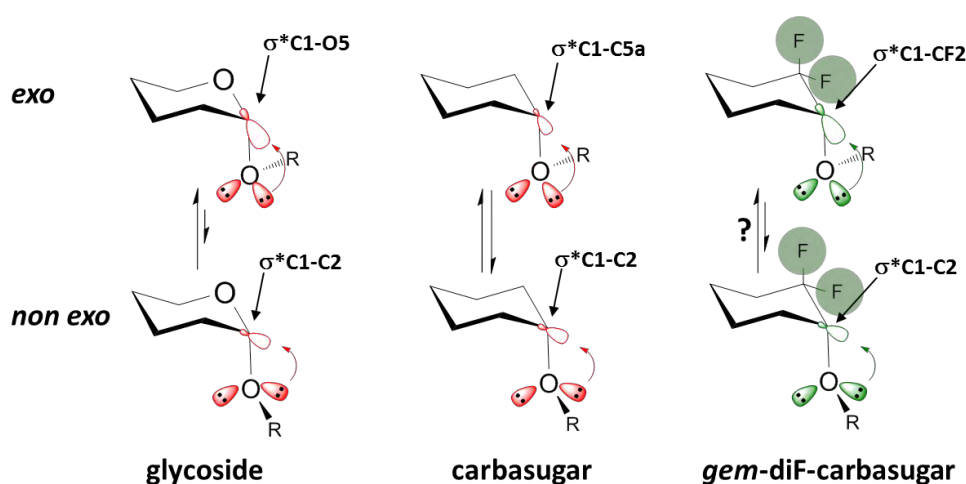
Molecular mimicry is an essential part of the development of drugs and molecular probes. In the chemical glycobiology field, although many glycomimetics have been developed in the past years, it has been considered that many failures in their use are related to the lack of the anomeric effects in these analogues. Additionally, the origin of the anomeric effects is still the subject of virulent scientific debates. The precise understanding of the parameters governing those effects is therefore essential to the design of new and more efficient therapeutic molecules. More specifically, the replacement of the exocyclic-anomeric oxygen atom by a carbon leads to C-glycosides<sup>[1]</sup> while replacement of the endocyclic one produces carbasugars.<sup>[2]</sup> In both cases, when the aglycone is another glycoside, this transformation leads to non-hydrolysable disaccharide analogues: C-disaccharides and carbadisaccharides. Figure 3.1.



**Figure 3.1.** Schematic perspective of the different glycomimetics mentioned in the text.

In both cases, the conformational behavior of those molecules changes drastically: more flexibility in the interglycosidic linkage as well as the population of unnatural conformations are observed.<sup>[3,4,5,6,7]</sup> These changes are often detrimental to the efficient interaction of such molecules with target proteins, mainly due to the entropic penalty it induces. This phenomenon has been tentatively attributed to the absence of the anomeric effects, especially the *exo*-anomeric one.<sup>[8]</sup> The anomeric effects, introduced in the late fifties, still is subject of scientific debate,<sup>[9,10,11]</sup> (see introduction page 10). In the case of the *exo*-anomeric effect, plausibly, as mentioned in the introduction, it has origin in the favorable interaction between a lone pair of the exocyclic anomeric oxygen atom with the parallel  $\sigma^*$  orbital of the adjacent C1-O5 bond.

This favourable interaction with C1-O5 over C1-C2 can be attributed to the different polarization between the C1-O5 and C1-C2 bonds, producing a larger  $\sigma^*$  orbital centered on the less electronegative atom of the polarized bond. Therefore, the search for closer “stereoelectronic” mimics retaining this feature is essential. According to the previous statement, replacement of the endocyclic oxygen atom by a  $\text{CF}_2$  group instead of a  $\text{CH}_2$  should induce a polarization of the C1- $\text{CF}_2$  bond and restore the *exo*-anomeric effect. Figure 3.2. Conversely, when the *exo*-cyclic oxygen is replaced by a  $\text{CF}_2$  rather than by a  $\text{CH}_2$  it populates the unnatural non-*exo* conformation due to hyperconjugation of the C1-H1 and C1-C2 bonds with the C-F bonds.<sup>[12,13]</sup>



**Figure 3.2.** Schematic representation of the lone pair- $\sigma^*$  interactions responsible for the *exo*-anomeric effect. The *exo*-anomeric conformers are represented in the top row. For the regular glycoside (left), a good overlapping between the  $\sigma^*$  and the lone pair is possible. The possibility of restoration of the *exo*-anomeric effect in *gem*-difluoro-carbadiisaccharides is presented at the right-hand side. The non-*exo*-anomeric geometries are depicted in the bottom row. In this case, no proper overlapping between the  $\sigma^*$  and the lone pair is possible for any molecule.

Herein, by combining NMR, and theoretical calculations, we show that it is possible to restore the *exo*-anomeric effect for an acetal when replacing the endocyclic oxygen atom by a  $\text{CF}_2$  group. This result provides key findings in Glycosciences. On the one hand, it strongly suggests the key relevance of the stereoelectronic component of the anomeric effect. On the other hand, as far as the the  $\text{CF}_2$  analogue can adopt the natural glycoside conformation, it might provide new avenues for sugar-based drug design.

In this chapter we therefore embarked in the study of *gem*-diF-carbadiisaccharides to explore the possibility to restore the *exo*-anomeric effect in a sugar mimic and to, consequently, narrow the



range of accessible  $\Phi$  dihedral angles at the glycosidic linkage similarly as happen with natural O-glycosides.

### 3.2 Results and Discussion

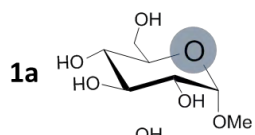
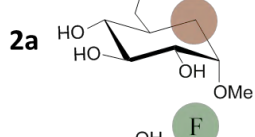
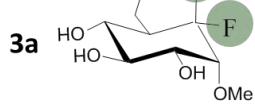
**3.2.1 *Ab initio* calculations.** First, we tested our hypothesis in silico. Hence, we performed Density functional theory (DFT) calculations on simple methyl  $\alpha$ -D-glucopyranoside (**1a**) and its CH<sub>2</sub> (**2a**) and CF<sub>2</sub> (**3a**) counterparts. Solvent effects were included using the polarizable continuum model (PCM) representing water. The obtained geometries were then submitted to Natural Bonding Orbital<sup>[14,15,16]</sup> (NBO) analysis. We focused our attention on investigating how the *endo*- and *exo*-anomeric effects were affected by these structural modifications, since NBO analysis allows the elucidation of the role of intramolecular orbital interactions. The protocol considers all possible interactions between the filled donors and empty acceptors and estimates their energetic importance using second-order perturbation theory. For each donor NBO (*i*) and acceptor NBO (*j*), the stabilization energy  $E^{(2)}$  associated with the corresponding electron delocalization is estimated as:

$$E^{(2)} = -\frac{F_{ij}^2}{\epsilon_j - \epsilon_i}$$

Where  $q_i$  is the orbital occupancy,  $\epsilon_i$ ,  $\epsilon_j$  are the diagonal elements (orbital energies), and  $F_{ij}$  is the off-diagonal NBO Fock matrix element.

Table 3.1 lists the calculated stabilization energies corresponding to the anomeric effects. As expected, they coexist in the natural glycoside **1a**, but disappear in the analogous CH<sub>2</sub> carbasugar **2a**. However, the *exo*-anomeric effect reappears in the *gem*-diF-carbaglycoside **3a**. Furthermore, a small but unexpected interaction between a lone pair of the axial fluorine atom and the  $\sigma^*_{C1-O1}$  is also observed, slightly mimicking the *endo*-anomeric effect.

**Table 3.1.** Second-order interaction energy ( $E^{(2)}$ , kcal·mol<sup>-1</sup>) between donor and acceptor orbitals in a natural sugar (**1a**) and analogous carbasugars (**2a**; CH<sub>2</sub>, and **3a**; CF<sub>2</sub>).

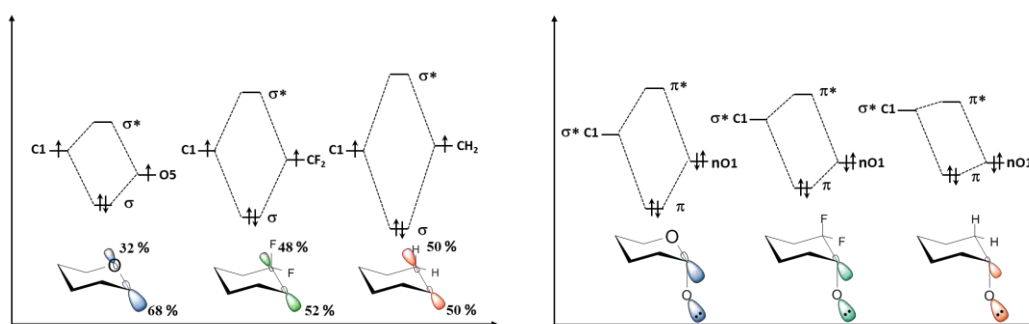
Molecule	Stabilization Energy (E <sup>2</sup> ) Kcal/mol	
	<i>exo</i> -anomeric effect	<i>endo</i> -anomeric effect
		$n_{O_{exo}} \rightarrow \sigma^*_{C1-O_{endo}}$ <b>17.39</b>
	$n_{O_{exo}} \rightarrow \sigma^*_{C1-C5a}$ <b>0.84</b>	N.P.
	$n_{O_{exo}} \rightarrow \sigma^*_{C1-C5a}$ <b>9.24</b>	$N_{Fax(LP)} \rightarrow \sigma^*_{C1-O_{exo}}$ <b>1.33</b>

The energy contribution to the *exo*-anomeric effect in the natural sugar compared to that in the corresponding CH<sub>2</sub> and CF<sub>2</sub> carbasugars clearly suggests that the fluorine atoms, due to their high electronegativity, favor the recovery of the *exo*-anomeric effect through polarization of the C1-CF<sub>2</sub> bond. The optimized structural parameters of the studied molecules obtained at the DFT (B3LYP/6-31++G PCM) level<sup>[17]</sup> are given in Table 3.2.

**Table 3.2.** Selected bond lengths (in Å) for compounds **1a-3a**.

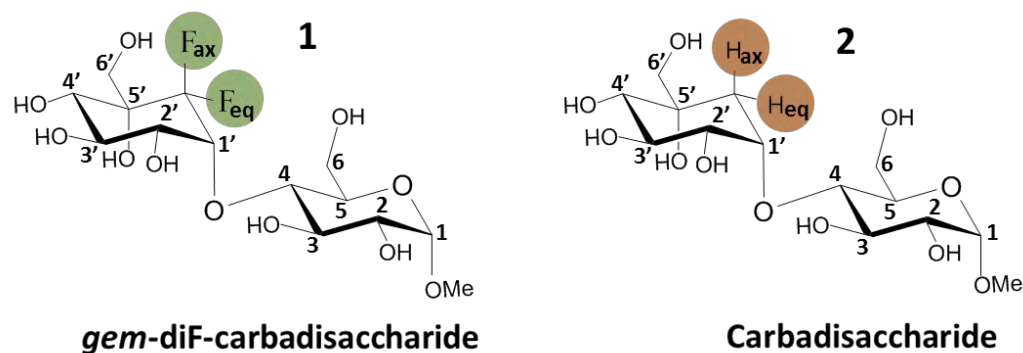
Compound	1a	2a	3a
C1-C2	1.52816	1.53367	1.54290
C2-C3	1.52883	1.52779	1.53860
C3-C4	1.52891	1.52889	1.53044
C4-C5	1.53293	1.54297	1.54082
C5-C6	1.52512	1.53131	1.55162
C1-O1	1.43150	1.47008	1.44414
C1-O5	1.46277	-	-
C1-C5a	-	1.53380	1.53438
C5-O5	1.47763	-	-
C5-C5a	-	1.54779	1.53899
C5a-Hax	-	1.10037	-
C5a-Heq	-	1.09619	-
C5a-Fax	-	-	1.45071
C5a-Feq	-	-	1.42382

The calculated bond lengths for **2a** and **3a**, compared to those of **1a**, are in agreement with the presence of *exo*-anomeric effect in **3a** (Table 3.2). According to the calculated molecular orbital diagrams (Figure 3.3), the uneven electron density distribution in the C1-X5a  $\sigma$  bonding orbitals for **3a** and **1a** compared to those of **2a**, translates in a better orbital overlapping between the (n) lone pair molecular orbital of the exocyclic oxygen and the unoccupied anti-bonding ( $\sigma^*$ ) molecular orbital at C-1. Moreover, the inductive effect of the electron-withdrawing fluorine atoms at C5a also contributes to the preservation of a large stereoelectronic *exo*-anomeric contribution in **3a**.



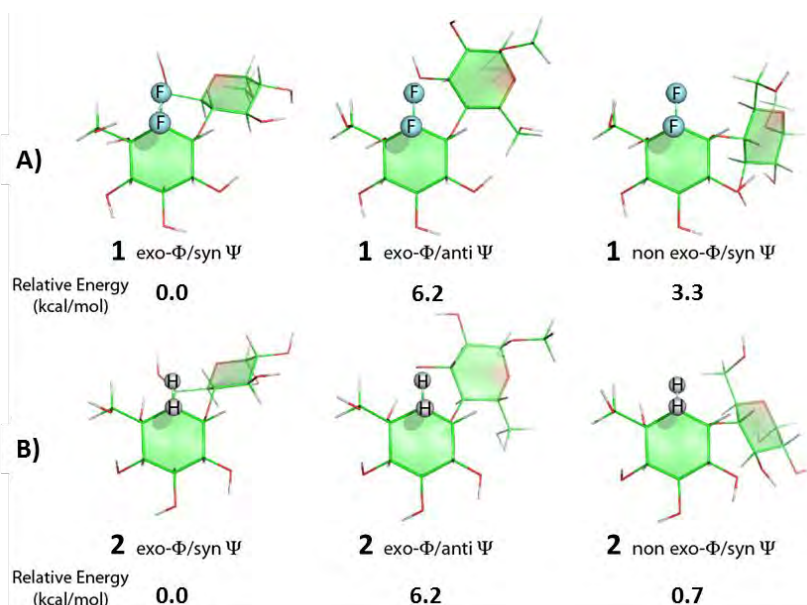
**Figure 3.3.** Molecular Orbital representation. On the left, schematic representation of bonding and anti-bonding molecular orbitals (MOs) for the C1-X bond, where X is either O5, or CF<sub>2</sub> or CH<sub>2</sub> functionalities. According to the MOT, the resulting bonding molecular orbital  $\sigma$  has a lower energy than that of the constituent atomic orbitals, while the resulting antibonding molecular orbital  $\sigma^*$  have higher energy compared to the constituent atomic orbitals. The higher energy value associated with the atomic orbital of the carbon atom of the CH<sub>2</sub> compared to that of CF<sub>2</sub> and O5, implies a higher energy value for the corresponding antibonding molecular orbital. On the right hand side a schematic representation shows the hyperconjugation of the O1 electronic lone pair to the C1-X antibonding molecular orbital, which is the origin of the *exo*-anomeric effect. This orbital overlapping is energetically more favorable for the C1-X  $\sigma^*$  MO of lower energetic value, that is for C1-O5 better than C1-CF<sub>2</sub>, better than C1-CH<sub>2</sub>. Furthermore, because of the higher electronegativity of oxygen with respect to carbon, the resulting  $\sigma$  MO for the C1-O5 bond has higher electron density close to O5 and lower close to C1. Thus, the corresponding  $\sigma^*$  MO has the opposite shape, which is more favorable for hosting the electron lone pair from O1. This effect is preserved when X is CF<sub>2</sub> because this functionality is still more electronegative than C1, while when X is CH<sub>2</sub> this effect no longer applies.

Encouraged by these preliminary results, and to experimentally validate this concept, the group of Prof. Matthieu Sollogoub also synthesized a *gem*-diF-carbadisaccharide **1** as well as its methylene carbasugar analogue, **2**, glycomimetics related to maltose. Figure 3.4.



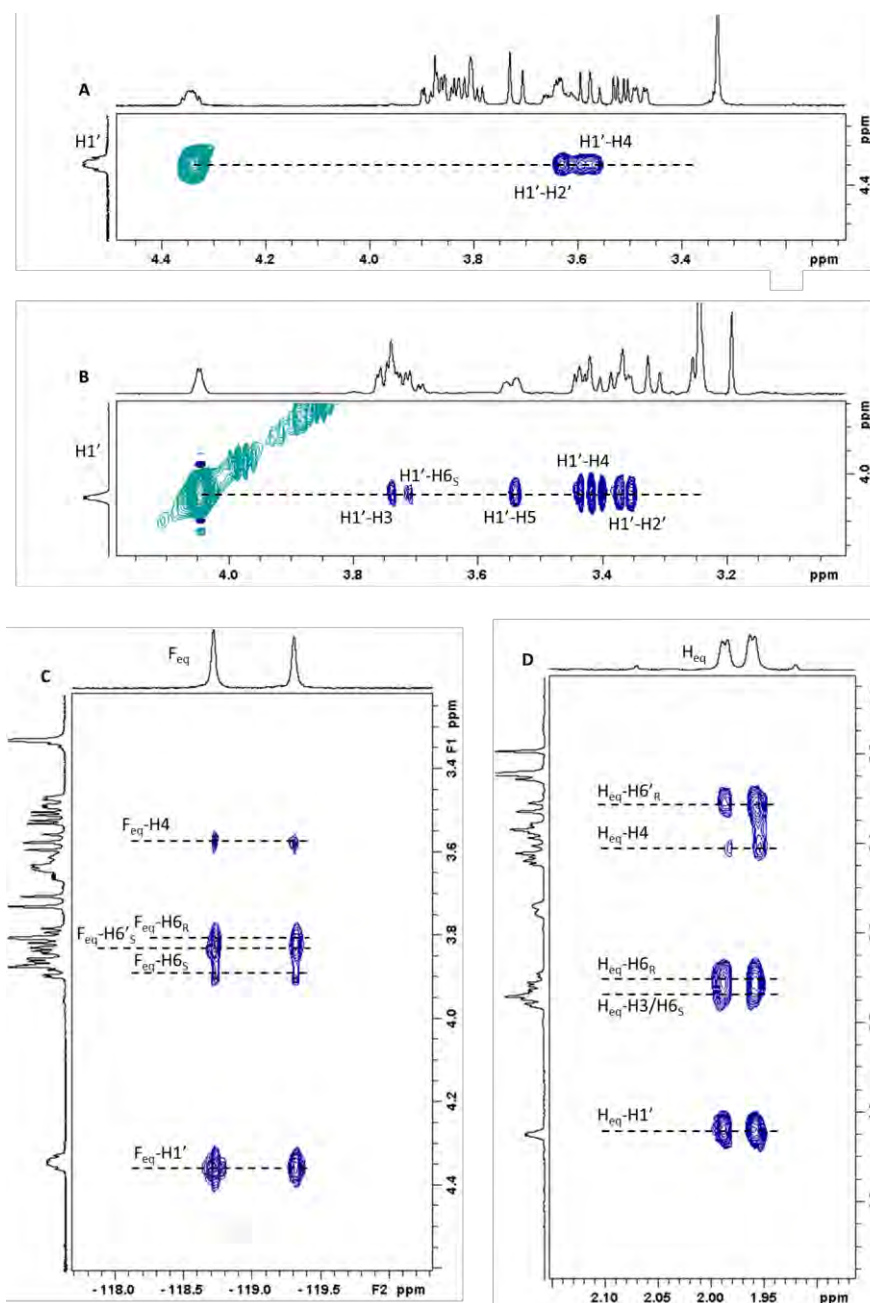
**Figure 3.4.** Schematic representation of synthetic glycomimetics of maltose **1** and **2**. The atoms labels are also shown.

**3.2.2 NMR Conformational studies.** The conformation of **1** was investigated by a combination of NMR and molecular modeling methods, and compared to the carbasugar analogue **2**, where fluorine atoms are replaced by hydrogen atoms, to probe the stereoelectronic effect of the CF<sub>2</sub> group. We first performed molecular mechanics calculations (MM3\*)<sup>[18]</sup> and found that, for both **1** and **2**, there were three stable conformations around the  $\Phi/\Psi$  glycosidic linkages. They correspond to the *exo*- $\Phi$ /*syn*- $\Psi$  ( $\Phi/\Psi$  ca. -40/-20), *exo*- $\Phi$ /*anti*- $\Psi$  ( $\Phi/\Psi$  ca. -40/180) and non-*exo*- $\Phi$ /*syn*- $\Psi$  ( $\Phi/\Psi$  ca. 40/0) conformations. The calculated conformers are represented in figure 3.5, together with the steric energy values provided by MM3\*.



**Figure 3.5.** Stick representations and relative steric energy values of the major conformers of **1**, panel A, and **2**, panel B, according to MM3\* calculations. Fluorine and hydrogen atoms of CF<sub>2</sub> and CH<sub>2</sub> respectively are highlighted as balls.

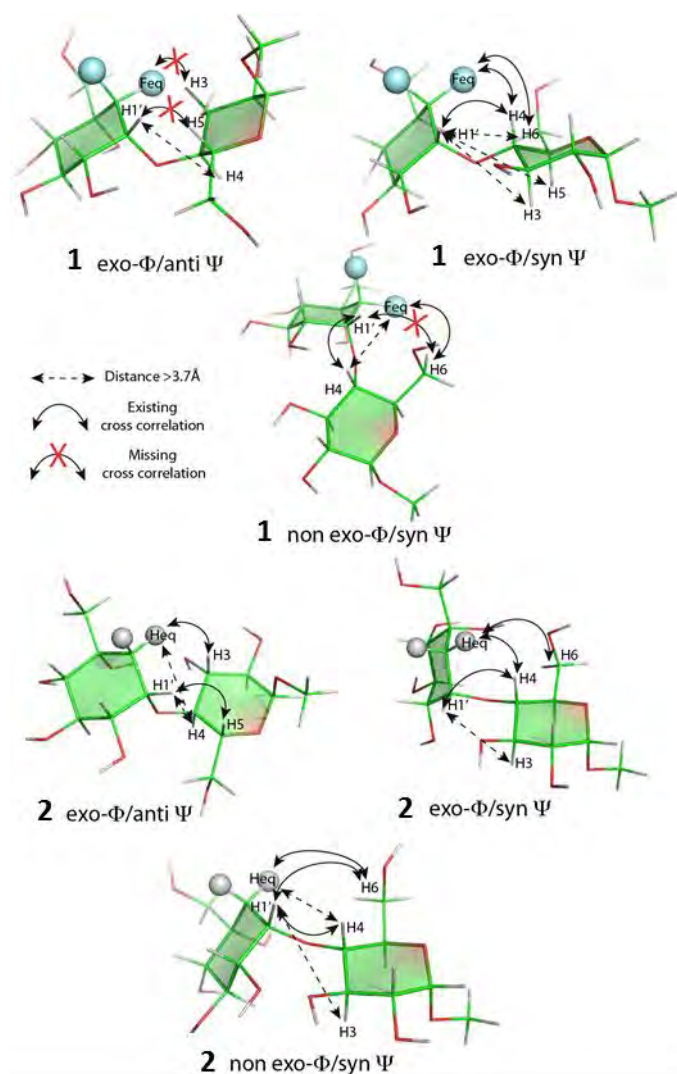
<sup>1</sup>H-<sup>1</sup>H NOESY and/or <sup>1</sup>H-<sup>19</sup>F HOESY NMR experiments were carried out on both **1** and **2**. As shown on figure 3.6 the pattern of correlations for H1' is strikingly different between the *gem*-diF-carbadiisaccharide **1** and the carbadiisaccharide **2**. While the only inter-residue correlation of H1' is with H4 for **1**, H1' correlates with H3, H4, H5 and H6 in the case of **2** (Figure 3.6A and 3.6B). Both the Feq in **1** and H5aeq (further referred to as Heq) in **2** correlate with H4 and H6 (Figure 3.6C and 3.6D).



**Figure 3.6.** (Upper) Strips of 2D NOESY spectra (700 ms mixing time) of compounds **1** (A) and **2** (B) taken at frequency of H1'. (Lower) Strips of 2D  $^1\text{H}$ - $^{19}\text{F}$  HOESY (500 ms mixing time) of **1** (C) and  $^1\text{H}$ - $^1\text{H}$  NOESY (700 ms mixing time) for **2** (D). Key NOEs are highlighted. No other interresidue NOEs are observed above the noise level.

We next estimated the interatomic distances using the calculated structures (figure 3.5) as well as the experimental proton-proton and proton-fluorine distances from the integration of the observed NOEs cross peaks using the Isolated Spin Pair Approximation (ISPA).<sup>[19]</sup> The results are summarized in the table 3.3 and illustrated on the different conformers in figure 3.7. For the

*gem*-diF-carbasugar **1**, the *exo*- $\Phi$ /*anti*- $\Psi$  conformer can be easily dismissed because no correlations are observed between H1' and H5, nor between Heq and H3. The fact that H1' only correlates with H4 cannot discriminate between the *exo*- $\Phi$ /*syn*- $\Psi$  and the non-*exo*- $\Phi$ /*syn*- $\Psi$  conformers, but the presence of a Heq-H4 correlation together with the absence of the H1'-H6 correlation allows the clear discrimination between both geometries, exclusively in favor of the *exo*- $\Phi$ /*syn*- $\Psi$  conformer as only geometry for **1**. In striking contrast, the simultaneous presence of Heq-H4, Heq-H6, H1'-H3, H1'-H4, H1'-H5, H1'-H6 correlations for carbasaccharide **2** clearly indicate a conformational equilibrium between the three conformers depicted in figure 3.7. In particular, the presence of the Heq/H4 NOE can only be explained by the *exo*- $\Phi$ /*syn*- $\Psi$  geometry. The presence of the Heq/H6S NOE and H1'/H6R NOEs can only be satisfied by the alternative non-*exo*- $\Phi$ /*syn*- $\Psi$  conformer. The observation of the H1'/H3 and H1'/H5 NOEs is only compatible with the presence of the third *exo*- $\Phi$ /*anti*- $\Psi$  conformation. (Figure 3.7) Therefore, the three conformations of **2** indeed exist in solution.



**Figure 3.7.** Schematic view of the NOE contacts and distances for each conformer.

**Table 3.3.** Relevant inter-atomic distances (Å) for the conformers of **1** and **2** from molecular mechanics calculations. For the data involving the H6 protons, the estimated distances for the gg and gt rotamers are indicated. The comparison with the key experimental NOE contacts with conformational information is also provided. The relevant intra and inter residue and ensemble average  $\langle r^{-6} \rangle^{-1/6}$  proton-proton and proton-fluorine distances (Å) were estimated from the integration of the observed NOEs cross peaks (measured in NOESY and HOESY experiments with mixing times of 300, 500, and 700 ms) using the ISPA approximation. The data for the “best” conformer of **1** and **2** are underlined.



Compound 1					
Atom Pair	Calculated distance for each conformer (Å)			Observed NOE intensity	Experimental Distance (Å) (±5%)
	exo- $\Phi$ /syn- $\Psi$	exo- $\Phi$ /anti- $\Psi$	non-exo- $\Phi$ /syn- $\Psi$		
Feq H4	<u>3.1</u>	4.2	4.5	Medium weak	3.2
Feq H6S	<u>4.0<sup>gg</sup>/2.9<sup>gt</sup></u>	>5	4.3 <sup>gg</sup> /3.0 <sup>gt</sup>	weak	3.5
Feq H6R	<u>3.1<sup>gg</sup>/2.6<sup>gt</sup></u>	>5	<u>2.6<sup>gg</sup>/4.4<sup>gt</sup></u>	medium	3.1
H1' H3*	<u>3.7</u>	2.1	<u>4.4</u>	not observed	>3.7
H1' H4	<u>2.3</u>	3.6	<u>2.4</u>	very strong	2.3
H1' H5	<u>4.4</u>	2.3	<u>3.7</u>	not observed	>3.7
H1' H6S	<u>&gt;4.5</u>	3.6 <sup>gg</sup> /3.0 <sup>gt</sup>	4.0 <sup>gg</sup> /2.8 <sup>gt</sup>	not observed	>3.7
H1' H6R*	<u>&gt;4.5</u>	3.0 <sup>gg</sup> /4.1 <sup>gt</sup>	3.0 <sup>gg</sup> /2.0 <sup>gt</sup>	not observed	>3.7

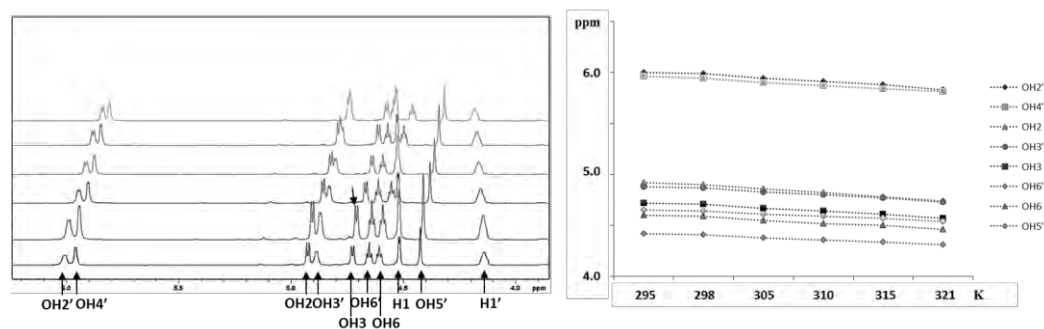
  

Compound 2					
Atom Pair	Calculated distance for each conformer (Å)			Observed NOE intensity	Experimental Distance (Å) (±5%)
	exo- $\Phi$ /syn- $\Psi$	exo- $\Phi$ /anti- $\Psi$	non-exo- $\Phi$ /syn- $\Psi$		
Heq H4	2.3	<u>4.0</u>	<u>4.0</u>	very weak	3.6
Heq H6S	<u>4.4<sup>gg</sup>/3.4<sup>gt</sup></u>	>5	<u>4.5<sup>gg</sup>/3.0<sup>gt</sup></u>	very weak	3.6
Heq H6R	4.0 <sup>gg</sup> /3.0 <sup>gt</sup>	>5	<u>2.8<sup>gg</sup>/4.6<sup>gt</sup></u>	medium strong	2.7
H1' H3*	4.8	<u>2.0</u>	<u>4.4</u>	weak	3.5
H1' H4	<u>2.5</u>	3.5	<u>2.2</u>	very strong	2.3
H1' H5	4.8	<u>2.1</u>	<u>3.8</u>	Medium weak	3.4
H1' H6R*	>4.5	<u>3.8<sup>gg</sup>/3.2<sup>gt</sup></u>	<u>4.2<sup>gg</sup>/3.0<sup>gt</sup></u>	weak	3.5
H1' H6S	>4.5	3.2 <sup>gg</sup> /4.1 <sup>gt</sup>	<u>3.3<sup>gg</sup>/2.3<sup>gt</sup></u>	Medium weak	3.3

\*overlap

**3.2.3 Intramolecular hydrogen bond.** The possibility of establishing an intra-residue OH5'...O1', figure 3.4, hydrogen bond was also scrutinized because it could stabilize one or the other conformation. According to the calculations, this possibility would exist for both *exo*- and non-*exo* conformers. However, since the experiments have been performed in water (D<sub>2</sub>O) solution, given the massive presence of water molecules, the importance of any intra-residue OH...O interaction should be minimum and not influence the conformational behavior. Nevertheless, it has been suggested that conventional<sup>[20]</sup> and nonconventional<sup>[21]</sup> inter-residues hydrogen bonds could lock the sugar conformation, even in water solution, resulting in a narrow cluster of  $\Phi/\psi$  torsion angles. Therefore the possibility of occurrence of the intra-molecular HB for **1** was also experimentally addressed. Among many others, two approaches have extensively been used to characterize secondary weak interactions as hydrogen bonds. The first one consists in monitoring the chemical shift perturbation of the hydroxyl protons as function of temperature.<sup>[22]</sup> In fact, protons directly involved in HB interactions are less sensitive to temperature changes. The magnitude of thermal motions increases with the temperature, and it results in lengthening of average hydrogen bond lengths or, finally, in the complete break of the

interaction. In a protic solvent, it is expected that all the hydrogen bond donor/acceptor groups in the solute are satisfied by their counterparts on the solvent molecules. However, in aprotic solvents, these interactions are not possible and the solute balances the dipolar interactions via intra-molecular interactions. As a result, the discussed secondary interactions, if present, are maximized in aprotic solvents. Therefore, we analysed the NMR spectra of **1** in the aprotic solvent DMSO. We assumed that, if an intramolecular HB exists, it should be easily detectable in this non-competitive media. The values of the temperature coefficients of the different hydroxyl protons of **1** as potential hydrogen bond indicators are represented in figure 3.8. The narrow range of temperature coefficients for all the hydroxyl protons demonstrates that no significant differences between them exist, even in the aprotic solvent, so further dismissing the possibility of the existence of any stabilizing intramolecular interaction in water.

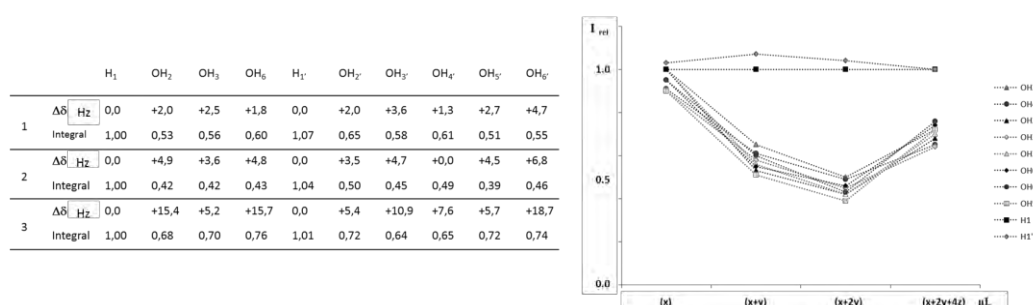


**Figure 3.8.** Temperature coefficients measured for the different hydroxyl groups of **1**, from the analysis of the  $^1\text{H}$  NMR spectra recorded in DMSO- $d_6$  between 295 (bottom) and 321 K (top). The range of temperature coefficients is very narrow for all the hydroxyl groups, being in between  $-4.2$  and  $-6.5$  ppb/ $^\circ$ . Left, proton NMR spectra of **1** at different temperature. Right, graphic representation of temperature coefficients.

However, among all the sugar hydroxyl protons, the one that showed the lowest temperature coefficient ( $-4.2$  ppb/ $^\circ$ ) was indeed OH5'. Therefore, in order to definitively discard the existence of any HB interaction, even weak, a second approach was also employed. Hydrogen/Deuterium exchange experiments have been extensively used in protein NMR structural characterization.<sup>[23]</sup> Specifically, several intra-molecular HB contribute to the stabilization of the secondary structural elements in proteins, with the protein folding defining those protein regions that are solvent exposed. Therefore, an amide proton which is part of an intramolecular HB or that is buried from the solvent will exchange slowly with respect to those that are solvent exposed or not involved in HBs. Thus, H/D exchange experiments has been largely applied to characterize protein folding,<sup>[24]</sup>

protein-protein interactions<sup>[25]</sup> or protein conformational changes.<sup>[26]</sup> In our context, differences in H/D exchange rates between hydroxyl protons of the fluorinated carbodisaccharide can be, obviously, only related to intramolecular HB interactions. The results for H/D exchange in gem-difluorocarbodisaccharide are reported in table 3.4 and plotted in the associated graphics. In particular, to a sample prepared in neat DMSO-d<sub>6</sub>, 1μL (first point) and 2μL (second point) of D<sub>2</sub>O were added and the reduction in the intensities of the hydroxyl <sup>1</sup>H NMR signals monitored immediately and after 30 minutes. (see also SI Figure 3.1). All the exchangeable protons on the sugar mimic experienced the same signal decrease, between 40-50% after the first addition, and 50-60% after the second one. In fact, OH5' experienced the faster exchange rate. The third point corresponds to the addition of 4μL of H<sub>2</sub>O. The presence of light water introduces protons into the system, with the consequent exchange process. All the hydroxyl protons recovered the 70% of the original signals intensity. These observations unequivocally demonstrate that no persistent intraresidue hydrogen bond exists for this molecule, even in non-competitive solvent. Consequently, we can ensure that the stereoelectronic component of the *exo*-anomeric effect is the only reason for the natural conformational behavior of the *gem*-diF-carbasugar.

**Table 3.4.** Exchange rate and chemical shift perturbation.



Left. The table refers to the chemical shift perturbation and relative signal intensity for all hydroxyl groups and for H1 and H1' (used as references). First row, upon addition of 1μL of D<sub>2</sub>O to compound **1** in neat DMSO-d<sub>6</sub>. Second row, upon addition of 2μL of D<sub>2</sub>O, and third row, after addition of 4μL of H<sub>2</sub>O. Right. Graphic representation of the data in the table. Where (x) is pure DMSO-d<sub>6</sub>, (y) is 1μL of D<sub>2</sub>O and (z) 1μL of H<sub>2</sub>O. The Integral values indicate the same behavior for all the hydroxyl protons, with no evidences of inter-residue hydrogen bonds.

### 3.3 Conclusions

Our theoretical calculations predicted that the *exo*-anomeric effect of maltose, which is almost completely abolished in its carbasugar analogue, is significantly restored when a CF<sub>2</sub> group is present at the endocyclic position, corresponding to O5 in the natural sugar. As an experimental demonstration, we have determined the conformational behavior of a *gem*-diF-carbasugar maltose analogue **1**, which indeed only exists in solution in the *exo*-anomeric conformation. In striking contrast, the corresponding carbasugar **2** displays a mixture of the *exo*- $\Phi$  and non-*exo*- $\Phi$  geometries in solution. It has therefore been demonstrated that it is possible to restore the *exo*-anomeric effect for an acetal when replacing one of the oxygen atoms by a CF<sub>2</sub> group. This result provides key findings in chemical sciences as it strongly suggests the importance of the stereoelectronic component for the *exo*-anomeric effect. Additionally, the obtained mimicking of the natural glycoside conformation may open new avenues for sugar-based drug design.

### 3.4 Methods

**3.4.1 Ab initio calculations:** Calculations were carried out with the Gaussian 03 suite of programs. The geometry optimization was performed utilizing Becke's hybrid three-parameter exchange functional and the nonlocal correlation functional of Lee, Yang, and Parr (B3LYP). Electron correlation energies were calculated by applying the second order Møller–Plesset (MP2) perturbation theory. The geometry of molecules (**1a**), (**2a**) and (**3a**) was optimized at the MP2/6-311<sup>++</sup>G(d,p) level of approximation. Natural bond orbital (NBO) calculations using the density by the Hartree-Fock (HF) calculations were performed with the NBO keyword included in Gaussian 03. In order to incorporate all contributions of the solvent to the molecular geometry, the solvent effects are included using the polarizable continuum model (PCM) representing water.

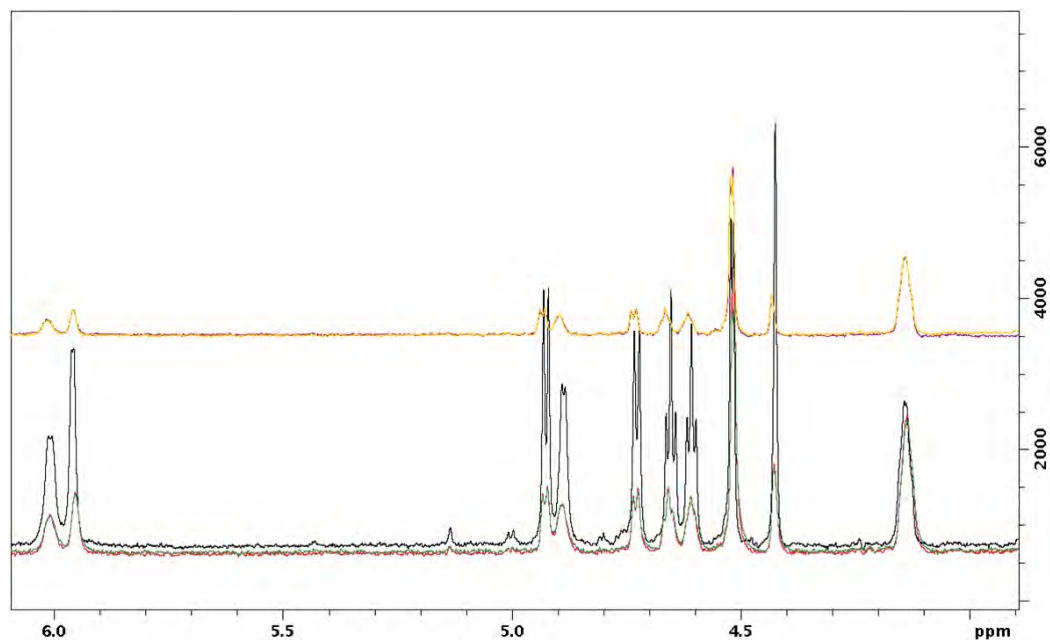
**3.4.2 NMR spectroscopy:** <sup>1</sup>H and <sup>13</sup>C NMR spectra were acquired with Bruker Avance 500 MHz and 600 MHz spectrometers equipped with a 5mm TXI probe. Experiments were carried out at 298K. <sup>1</sup>H and <sup>13</sup>C NMR chemical shift assignments were performed using standard 2D homonuclear experiments. In particular, <sup>1</sup>H,<sup>1</sup>H-COSY/TOCSY/NOESY, <sup>1</sup>H,<sup>13</sup>C-HSQC and <sup>1</sup>H,<sup>13</sup>C-HMBC. All experiments employed 256 t<sub>1</sub>-increments of 2 K points each, with a relaxation delay of 2 s. The <sup>1</sup>H dimensions were 3 ppm for **1** and 5 ppm for **2**. The <sup>13</sup>C dimensions were 50 ppm for **1** and 80 ppm for **2**. 8 scans were employed for COSY, HSQC, and TOCSY (mixing time 66 ms), while 32 scans were employed for the NOESY and 64 for the HMBC and HOESY experiments (mixing

times of 300, 500, and 700 ms). The  $^{19}\text{F}$  dimension was 22 ppm (for **1**). The  $^1\text{H}$  NMR chemical shifts are given in ppm and TSP was used as standard proton reference ( $\delta_{\text{H}} 0$  ppm).  $^{13}\text{C}$  Chemical shifts are expressed in ppm relative to internal acetone (2.225 and 31.4 ppm for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively). The  $^{19}\text{F}$  NMR experiments were performed at 470MHz on a Bruker Avance 500MHz equipped with a 5mm SEF probe. Assignment of 1D  $^{19}\text{F}$  NMR spectra were completed with the information extracted from 1D and 2D HOESY experiments. TFE, trifluoroethanol, was used as internal reference. 3mm NMR tubes were used to prepare the NMR samples. Thus, 1–2 mg of the glycomimetic was dissolved into 0.2mL of deuterated water. The purchased deuterated solvents were used without further purification. The experimental data were acquired and processed using the Topspin software (Bruker GmbH, Karlsruhe, Germany) on a PC station.  $^1\text{H}$  chemical shifts and  $^nJ_{\text{HH}}$  and  $^nJ_{\text{FH}}$  coupling constants were determined with the aid of the MestReNova spin simulation software.

**3.4.3 Molecular Mechanics calculations:** The geometry optimization was performed by using the Jaguar/Schrodinger package (version 9.1) and the MM3\* force field, with the GB/SA continuum solvent model for water. The glycosidic torsion angles were defined as  $\Phi$  (H1'-C1'-O4-C4) and  $\Psi$  (C1'-O4-C4-H4). Extended nonbonded cut-off distances (van der Waals cut-off of 8.0 Å and electrostatic cut-off of 20.0 Å) were used. The three conformers for each molecule **1** and **2** were generated employing the optimized *exo*-anomeric and non-*exo*-anomeric geometries. The extension of the aglycon moiety was generated manually and the three possible staggered rotamers around  $\Psi$  were then built and minimized. The two *syn*- $\Psi$  conformers (either positive or negative) always converged to the same geometry, while the *anti*- $\Psi$  alternative was a local minimum, as described in the text. The coordinates of the obtained local minima were employed to measure the key inter-proton distances that were then compared to those obtained experimentally by the NOESY and HOESY NMR methods.

### 3.5 Supporting Information

**Figure 3.1 SI.** H/D exchange rate in *gem*-difluorocarbadisaccharide (**1**).



No changes in the NMR signal intensities of the hydroxyl protons signals were detected with time. Bottom row. The regular <sup>1</sup>H NMR spectrum of **1** in neat DMSO-d<sub>6</sub> is shown in black. The red and green spectra correspond to the spectra recorded immediately after the addition of 2 μL of D<sub>2</sub>O to the sample (red) and after 30 minutes (green). There is no difference in the signal intensities. The yellow and violet spectra correspond to the spectra recorded immediately after the addition of 2 μL more of D<sub>2</sub>O to the sample (yellow) and after 30 minutes (violet). There is no difference in the signal intensities.

### 3.6 References

1. Vidal, P. *et al.* Conformational behaviour of glycomimetics: NMR and molecular studies of the C-glycoside analogue of the disaccharide methyl  $\beta$ -D-galactopyranosyl-(1-3)- $\beta$ -D-glucopyranoside. *Carbohydr. Res.* **342**, 1910-1917 (2007).
2. Montero, E. *et al.* The Conformational Behaviour of Non-Hydrolyzable Lactose Analogues: The Thioglycoside, Carboglycoside, and Carba-Iminoglycoside Cases. *European J. Org. Chem.* **2000**, 1945–1952 (2000).
3. Suami, T. & Ogawa, S. Chemistry of Carba-Sugars (Pseudo-Sugars) and their Derivatives (Ed. Elsevier, Academic Press) **1990**.
4. Ogawa, S., Matsunaga, N., Li, H. & Palcic, M. M. Synthesis of Ether- and Imino-Linked Octyl N-Acetyl-5a'-carba- $\beta$ -lactosaminides and -isolactosaminides: Acceptor Substrates for  $\alpha$ -(1 $\rightarrow$ 3/4)-Fucosyltransferase, and Enzymatic Synthesis of 5a'-Carbatrisaccharides. *European J. Org. Chem.* **1999**, 631–642 (1999).
5. Carpintero, M., Fernández-Mayoralas, A. & Jiménez-Barbero, J. The Conformational Behaviour of Fucosyl and Carbafucosyl Mimetics in the Free and in the Protein-Bound States. *European J. Org. Chem.* **2001**, 681–689 (2001).
6. Ogawa, S., Maruyama, A., Odagiri, T., Yuasa, H. & Hashimoto, H. Synthesis and Biological Evaluation of  $\alpha$ -L-Fucosidase Inhibitors: 5a-Carba- $\alpha$ -L-fucopyranosylamine and Related Compounds. *European J. Org. Chem.* **2001**, 967–974 (2001).
7. Carpintero, M., Bastida, A., García-Junceda, E., Jiménez-Barbero, J. & Fernández-Mayoralas, A. Synthesis of Carba- and C-Fucopyranosides and Their Evaluation as  $\alpha$ -Fucosidase Inhibitors – Analysis of an Unusual Conformation Adopted by an Amino-C-fucopyranoside. *European J. Org. Chem.* **2001**, 4127–4135 (2001).
8. Tvaroška, I. & Bleha, T. Anomeric and Exo-Anomeric Effects in Carbohydrate Chemistry (Ed. Elsevier Academic Press) **1989**.
9. Cocinero, E. J., Çarçal, P., Vaden, T. D., Simons, J. P. & Davis, B. G. Sensing the anomeric effect in a solvent-free environment. *Nature* **469**, 76–79 (2011).
10. Wang, C., Ying, F., Wu, W. & Mo, Y. Sensing or No Sensing: Can the Anomeric Effect Be Probed by a Sensing Molecule? *J. Am. Chem. Soc.* **133**, 13731–13736 (2011).
11. Mo, Y. Computational evidence that hyperconjugative interactions are not responsible for the anomeric effect. *Nat Chem* **2**, 666–671 (2010).
12. Perez-Castells, J. *et al.* The conformational behaviour and P-selectin inhibition of fluorine-containing sialyl LeX glycomimetics. *Org. Biomol. Chem.* **5**, 1087–1092 (2007).
13. Denton, R. W. *et al.* Synthesis and conformational behavior of the difluoromethylene linked C-glycoside analog of  $\beta$ -galactopyranosyl-(1 $\leftrightarrow$ 1)- $\alpha$ -mannopyranoside. *Carbohydr. Res.* **342**, 1624–1635 (2007).
14. Freitas, M. P. The anomeric effect on the basis of natural bond orbital analysis. *Org. Biomol. Chem.* **11**, 2885–2890 (2013).

15. Roohi, H., Ebrahimi, A., Habibi, S. M. & Jarahi, E. NBO and AIM analyses of the anomeric effect in fluoromethanthiol. *J. Mol. Struct. THEOCHEM* **772**, 65–73 (2006).
16. Bauerfeldt, G. F., Cardozo, T. M., Pereira, M. S. & da Silva, C. O. The anomeric effect: the dominance of exchange effects in closed-shell systems. *Org. Biomol. Chem.* **11**, 299–308 (2013).
17. Mennucci, B. *et al.* Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotations of Chiral Molecules. *J. Phys. Chem. A* **106**, 6102–6113 (2002).
18. Allinger, N. L., Yuh, Y. H. & Lii, J. H. Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **111**, 8551–8566 (1989).
19. Neuhaus, D. & Williamson, M. P. The Nuclear Overhauser Effect in Structural and Conformational Analysis (Ed. Wiley-VCH, 2<sup>nd</sup> Edition) **2000**.
20. Battistel, M. D., Azurmendi, H. F., Frank, M. & Freedberg, D. I. Uncovering Nonconventional and Conventional Hydrogen Bonds in Oligosaccharides through NMR Experiments and Molecular Modeling: Application to Sialyl Lewis-X. *J. Am. Chem. Soc.* **137**, 13444–13447 (2015).
21. Zierke, M. *et al.* Stabilization of Branched Oligosaccharides: Lewis Benefits from a Nonconventional C–H···O Hydrogen Bond. *J. Am. Chem. Soc.* **135**, 13464–13472 (2013).
22. Sicińska, W., Adams, B. & Lerner, L. A detailed <sup>1</sup>H and <sup>13</sup>C NMR study of a repeating disaccharide of hyaluronan: the effects of temperature and counterion type. *Carbohydr. Res.* **242**, 29–51 (1993).
23. Raschle, T., Rios Flores, P., Opitz, C., Müller, D. J. & Hiller, S. Monitoring Backbone Hydrogen Bond Formation in  $\beta$ -Barrel Membrane Protein Folding. *Angew. Chem Int. Ed.* **55**, 5952–5955 (2016).
24. Fazelinia, H., Xu, M., Cheng, H. & Roder, H. Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome c. *J. Am. Chem. Soc.* **136**, 733–740 (2014).
25. Wang, J. *et al.* Elucidation of the interaction loci of the human pyruvate dehydrogenase complex E2-E3BP core with pyruvate dehydrogenase kinase 1 and kinase 2 by H/D exchange mass spectrometry and nuclear magnetic resonance. *Biochemistry* **54**, 69–82 (2015).
26. Kim, J. Study of the conformational change of adsorbed proteins on biomaterial surfaces using hydrogen-deuterium exchange with mass spectroscopy. *Colloids Surfaces B Biointerfaces* **141**, 513–518 (2016).



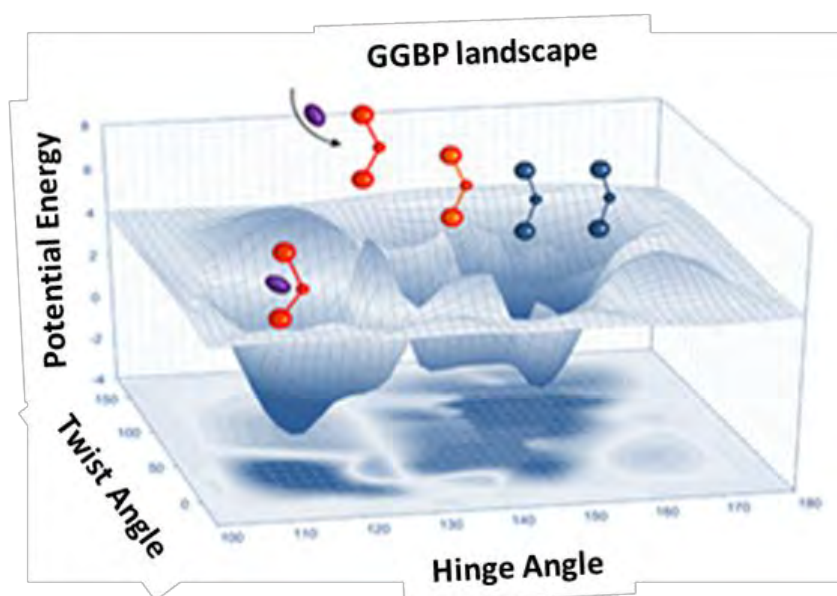


## CHAPTER IV

Sugar Receptor: unraveling the amplitude and time scale of macro-molecular motions

### Unraveling the conformational landscape of ligand binding to Glucose/Galactose-binding protein by paramagnetic NMR and MD simulations

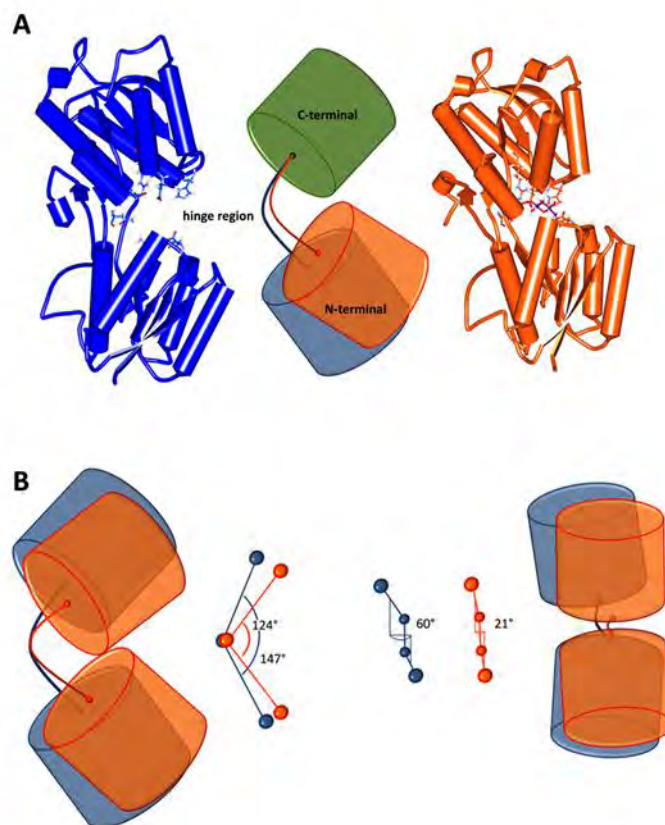
The work presented in this chapter has been performed in collaboration with the group of Dr. Oscar Millet at CICbioGUNE in Derio (Bizkaia), Spain. The expression, purification and NMR assignment of the  $^{15}\text{N}$  labeled protein was carried out by Dr. Gabriel Ortega at his lab.



## 4.1 Introduction

Protein function arises from the delicate interplay among structure, molecular recognition features and dynamics, but unraveling such contributions is often elusive. The periplasmic binding protein family (PBPs) represents a paradigm for describing functional conformational changes in flexible proteins.<sup>[1]</sup> In Gram-negative bacteria, PBPs selectively recognize and actively transport various nutrients across the inner membrane. The family is composed of about 100 members, classified according to the recognized ligand: amino acids, carbohydrates, oxyanions and vitamins.<sup>[2]</sup> Almost all of them share a common structural fold consisting of two globular Rossmann fold domains connected by three short linkers, thus suggesting interdomain flexibility.<sup>[3,4,5]</sup> This hypothesis is further supported by the different interdomain orientations found in the X-ray structures. For instance, one of the paradigmatic glycan-binding protein family, the glucose/galactose binding proteins (GGBP) from different organisms have generated a plethora of crystal structures that trap the biomolecule at distinct conformational instances: some unloaded structures are canonically open (*apo\_op*) while others are closed (*apo\_cl*) and resemble the holo-ligand-bound state (*holo\_cl*).<sup>[6,7,8,9,10]</sup> In the last years, segmental inter-domain reorientations in periplasmic binding proteins have been extensively investigated by solution NMR spectroscopy.<sup>[11]</sup> Clore and co-workers have demonstrated that a conformational selection process undergoes an open-to-closed transition in maltose binding protein (MBP)<sup>[12]</sup>, while Tjandra and co-workers have shown that an induced fit mechanism well describes open-closed transition of another PBP, glutamine-binding protein.<sup>[13]</sup> In a comparative NMR study of GGBP and the structurally homologous ribose binding protein, it is shown that the (apparent) ligand affinity can be modulated by redesigning the flexible hinge region, thus emphasizing the functional role of inter-domain dynamics.<sup>[14]</sup> However, the time scale and amplitude of these motions are experimentally ill-defined for all the investigated cases. Molecular dynamics simulations have also been widely used to characterize the conformational landscape of PBPs.<sup>[3]</sup> For instance, advanced sampling techniques have been used to study the allosteric equilibrium of the ribose-binding protein,<sup>[15]</sup> while accelerated MD simulations provided a detailed picture of the transition between the open and partially closed states in MBP.<sup>[16]</sup> Moreover, these proteins have been the target of intense studies in protein engineering<sup>[17]</sup> and the computational redesign of PBPs to build up nanobiosensors have raised great expectations.<sup>[18,19]</sup> For instance, Daunert and coworkers have proposed GGBP as a possible biosensor of glucose in blood.<sup>[20]</sup> However, ironically, the main limitation of the method is the high affinity for the substrate (nM range). Then, despite the extensive use of MD simulations in the study of PBPs, integrative approaches of protein design with experimental data are still largely unedited.

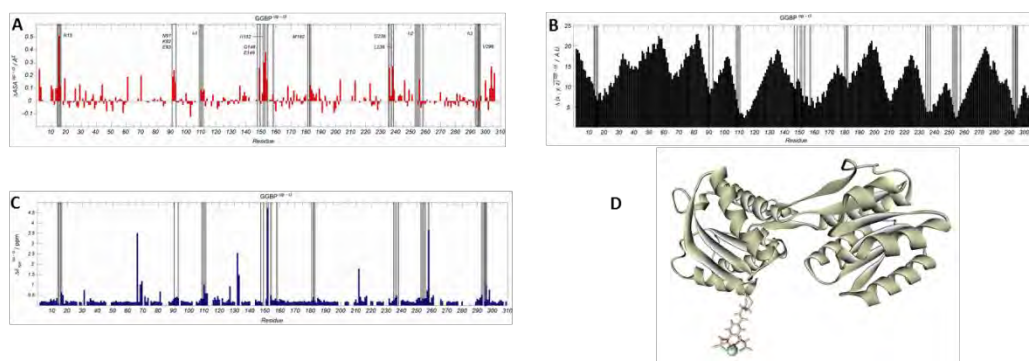
In this chapter, we propose an integrated approach by combining experimental NMR data with molecular dynamics simulations to quantitatively characterize interdomain dynamics in GGBP. First, pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) allowed disentangling the population distribution of conformers in the open-closed transition that GGBP undergoes. Next, the time scale for open-closed transition is defined by detailed molecular dynamics simulations. Finally, the energy barrier in the protein landscape has been estimated using non-equilibrium molecular dynamics calculations. Our results demonstrate that, in its *apo* state, the protein coexists between the open (68%) and closed (32%) conformations. The time scale for closed-open inter-conversion is around 25 ns. The presence of the ligand is the driving force for closing, largely through a conformational selection mechanism.



**Figure 4.1.** (A) Crystal structures of *apo\_op*GGBP (left, 2FW0) and *holo\_cl*GGBP (right, 2FVY). The bound  $\beta$  anomer of D-glucopyranose and residues Asp14, Asn91, His152, Asp154, Arg158, Asn211, Asp 236, Asn256 forming stabilizing H-bonds with the ligand are drawn as stick models (ligand in violet). Inset: schematic representation of globular domains (cylinders) and the hinge region (lines). The difference in N-terminal domain position highlights the difference in closure angle, according to X-ray structures. (B) Domain reorientation of GGBP. Left, side view illustrating hinge domain rearrangement between *apo\_op* (blue) and *holo\_cl* (orange) GGBP. Right, front view illustrating twist motion. Inset: the angle between the segments connecting the center of mass of the hinge region and those of C-terminal domain and N-terminal domain is defined as hinge angle, while the center of mass of the N-terminal domain, the base of the N-terminal domain, the C-terminal domain and the base of the C-terminal domain define three segments. The dihedral angle formed by these three segments is defined as twist angle.

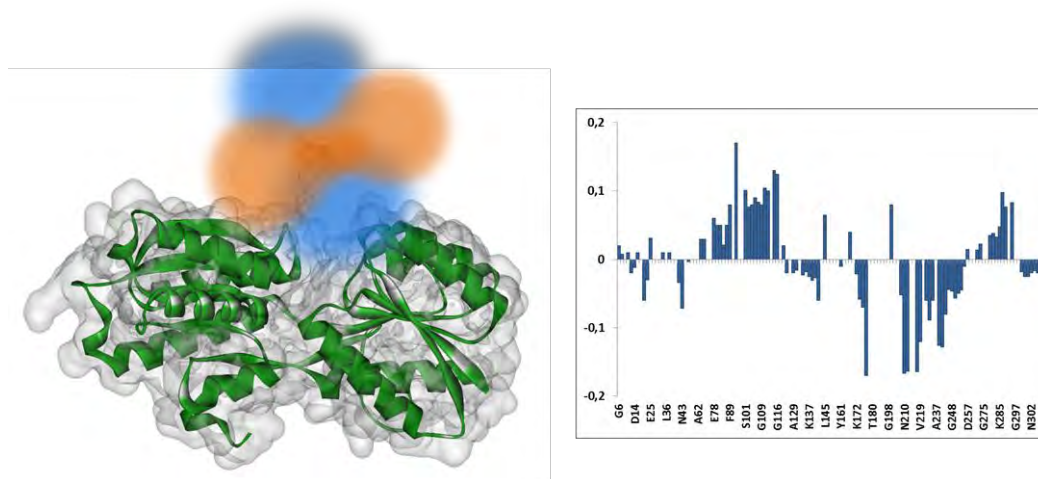
## 4.2 Results and Discussion

**4.2.1 Self-alignment with a paramagnetic Tag.** GGBP consists of two globular domains, the C-terminal domain (residues 112-254 and 297-306) and the smaller N-terminal domain (residues 3-108 and 258-291), linked by a three-strand hinge (residues 109-111, 255-258 and 292-296) (Figure 4.1, panel A). The  $\beta$  anomer of glucose binds to GGBP through an extensive network of hydrogen bonds and CH- $\pi$  interactions with high affinity and specificity ( $K_D = 290$  nM at 37 °C and pH 7.0) into the cleft near the hinge region, invoking a large conformational change from the open unbound (*apo\_op*) to closed bound (*holo\_cl*) state.<sup>[6]</sup> This segmental interdomain reorientation is well described by a rotation of 40° in the twist ( $\phi$ ) angle accompanied by a 23° rotation in the hinge ( $\theta$ ) angle<sup>[14,15]</sup> (Figure 4.1, panel B). Since closed unloaded structures are available for some PBPs, an open question is whether *apo*GGBP can make excursions to the *holo\_cl* conformation in the absence of ligand. Structural data for GGBP in solution can be integrated by anisotropic NMR parameters that are induced by self-alignment of paramagnetic molecules. Such self-alignment has been achieved by binding paramagnetic metal ions to a small molecule metal chelating tag, covalently attached to the biomolecule.<sup>[21]</sup> The alignment tensors for the tagged and non-tagged domains have been determined by a combined use of residual dipolar couplings (RDCs) and pseudocontact shifts (PCSs). Actually, owing to the large magnetic moment of the unpaired electrons on the paramagnetic lanthanide ion, the paramagnetic effects are detectable up to large distances ( $> 40$  Å).<sup>[22]</sup> In a system comprising two or more species in rapid exchange, the observed effect is a population-weighted average of the component conformers. As a result, PCSs and RDCs provide a unique way to describe complex mixtures of translational and rotational interdomain motions, simply aligning the tagged domain by the paramagnetic tag and determining the induced alignment on the other moieties.<sup>[23]</sup> To that end, the paramagnetic probe has been conjugated through nucleophilic substitution to an engineered cysteine residue, M182C, located in the C-terminal domain at the periphery of the interface between the two domains. Several factors were considered for the insertion of the tag molecule: i) the minimal bio-relevant mutation Cys instead of Met, ii) an adequate distance between the tag molecule and the target binding site, so the spin-label does not perturb the backbone structure nor the ligand-binding site and iii) surface-accessible aminoacids that experience minimum variation in chemical environment upon sugar binding (Figure 4.2).



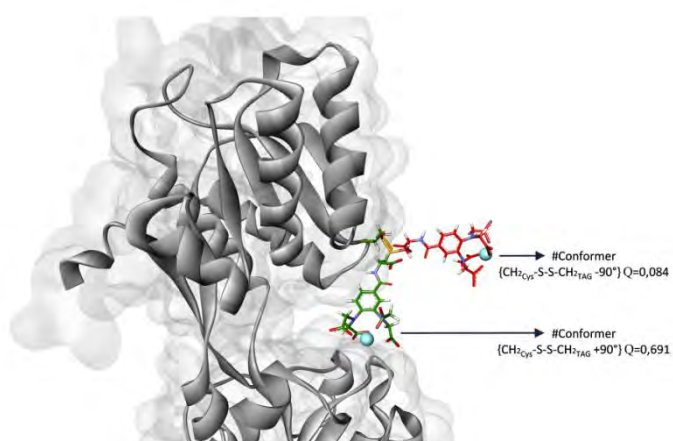
**Figure 4.2.** Single point mutation. Selection of the amino acids for effective TAG insertion (in gray lines). The inserted probe should be sensible to (A) Chemical environment change as function of open to closed transition. (B) Spatial coordinates perturbation in  $C\alpha$  as function of open to closed transition. (C) Spatial coordinates perturbation in NH as function of open to closed transition. The useful amino acids are highlighted with grey bars. (D) Schematic representation of paramagnetic Tagged protein M182C GGBP. The protein backbone is schematized according to its secondary structure, while the tag molecule is drawn as stick model. The green sphere represents the metal ion. The TAG conformation has been generated by steepest descents minimization to remove bad contacts with the protein residues.

The NMR signals for the residues within the shell around the paramagnetic center (C182 and A181) were broadened beyond detection due to paramagnetic relaxation enhancement (PRE). Nevertheless, the chemical nature of the Tag molecule (the metal is located at a distance  $>16 \text{ \AA}$  away from the protein backbone) allowed collecting 135 measurable PCSs (Figure 4.3) and 35 RDCs. Clearly, two regions of the protein orient differently with respect to the paramagnetic metal ion. One region undergoes negative chemical shift perturbation, close in space to the negative lobe of the magnetic susceptibility tensor. The other region experiences opposite changes due to its orientation towards the positive lobe of the paramagnetic metal isosurface.



**Figure 4.3.** Left, cartoon of the protein GGBP conjugated to the paramagnetic probe. On the right, Pseudo contact shifts (PCSs) obtained as the difference in the chemical shift of the protons signals in diamagnetic (lanthanum) and paramagnetic (dysprosium) conditions, for the observable  $H^N$ , N nuclei. Resonances of residues within the shell around the paramagnetic center were broadened beyond detection due to paramagnetic relaxation enhancement (PRE).

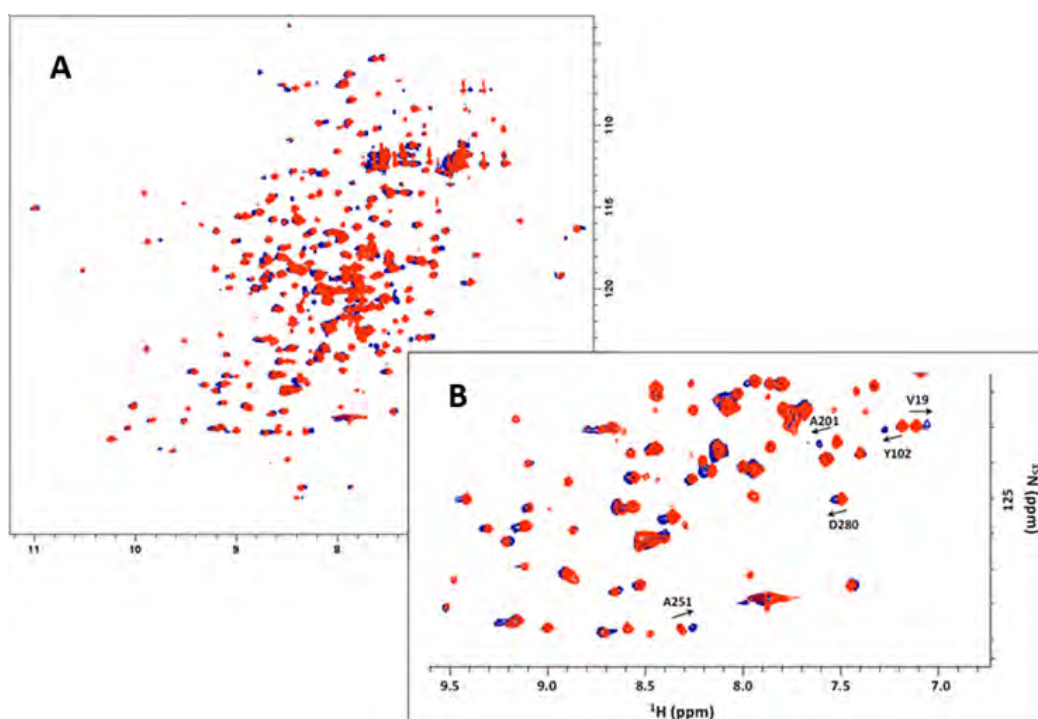
To take into account the flexibility of the linker, two main conformations with staggered dihedral angles around the di-sulfide bond have been generated ( $90^\circ$  and  $-90^\circ$ ). Only the  $-90^\circ$  conformer fits well the experimental data, confirming the inaccessibility of the  $90^\circ$  conformer due to steric clashes (Figure 4.4).



**Figure 4.4.** Geometric models of the two disulfide conformers,  $-90^\circ$  (orange) and  $90^\circ$  (green). Steric clashes on the  $-90^\circ$  conformer are easily predictable.



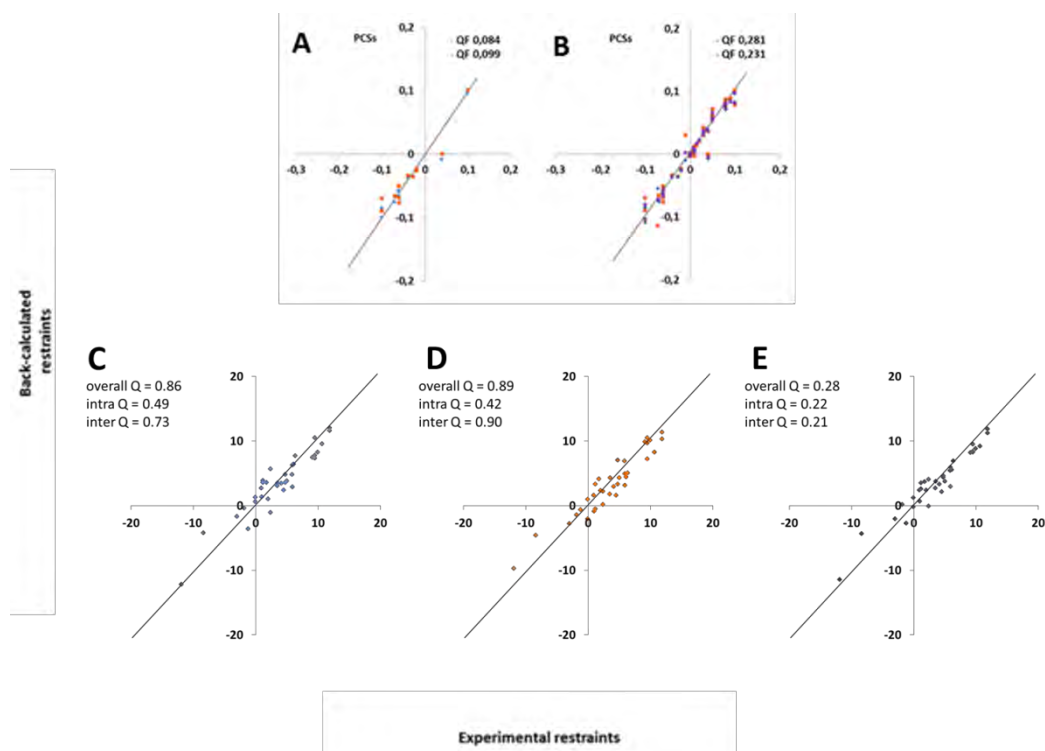
PCs of  $H^N$ , N nuclei, and RDCs of  $H^N$ -N pairs for paramagnetic lanthanide ( $Dy^{3+}$ ) were determined from  $^1H,^{15}N$  HSQC spectra. A  $La^{3+}$ -loaded complex was used as diamagnetic reference, as this metal has an ionic radius similar to the dysprosium ion. The protein shows excellent signal dispersion in both  $^1H,^{15}N$ -HSQC spectra, ( $Dy^{3+}$  and  $La^{3+}$ ), (Figure 4.5). Chemical shifts between unloaded and diamagnetic metal ion loaded molecules are virtually identical and nearly complete assignments could be obtained based on previously published data.<sup>[14]</sup> The lack of chemical shifts perturbation in GGBP backbone amide signals after metalation of the sample indicates that all the lanthanide ions are bound to the tag and not directly to the protein, consistent with the high affinity of the tag for lanthanides (in the  $10^{18} M^{-1}$  range).<sup>[24,25]</sup>



**Figure 4.5.** (A)  $^1H,^{15}N$  HSQC for GGBP loaded with diamagnetic lanthanum (orange) and paramagnetic Dysprosium (blue). Spectra were recorded at 310 K and pH 7.0 in a 600 MHz NMR spectrometer. (B) Selected spectral region of the  $^1H,^{15}N$  HSQC spectra.

Representative structures for the *apo\_op* and *apo\_cl* conformations (with  $5^\circ$  stepwise changes in the closure and twisting angles) were extracted from a molecular dynamic simulation (vide infra) and used for the alignment tensor estimation, using RDCs and PCs as experimental restraints. The structures providing the lowest quality factors (Q factor, lower value indicates best fitting to

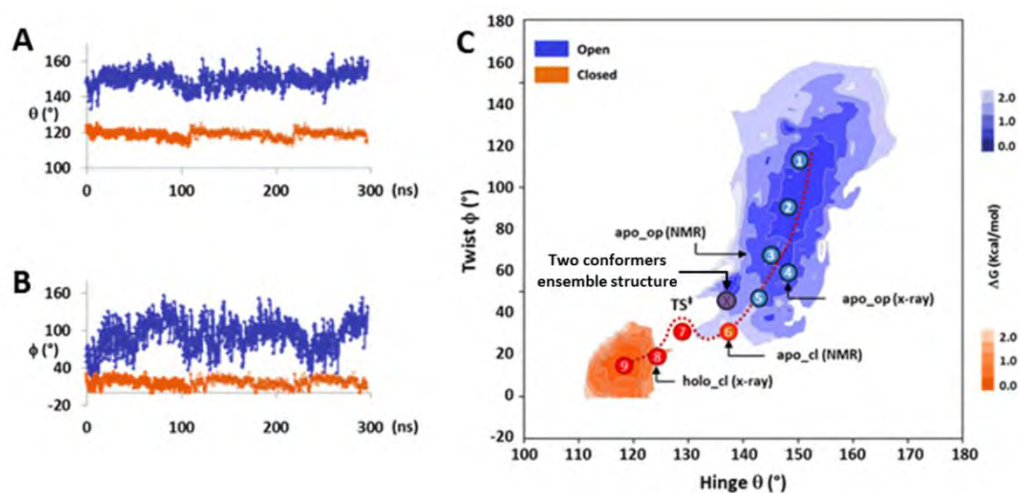
experimental data) for the tagged domain were selected and subsequently used for the prediction of the NMR parameters in the tagged-free domain. Two structures showed the best Q-factor when fitting the experimental data for tagged domain: one representative for the *apo\_op* form ( $\theta = 145^\circ$  and  $\phi = 64^\circ$ ) and another for the *apo\_cl* conformation ( $\theta = 137^\circ$  and  $\phi = 20^\circ$ ). When the data for each domain are analyzed independently, PCSs and RDCs provided excellent fits for the tagged domain for both conformers, as reflected in the range of the Q-factor values: 0.084-0.099 (PCSs) and 0.420-0.494 (RDCs). These results demonstrate that the C-terminal domain with the attached Tag molecule behaves as a rigid body (Figure 4.6, panels A and intra Q-factor in panels C-E). Interestingly, Q-factors increase drastically when both domains are included in the analysis (Figure.4.6, panel B and overall Q in panels C, D). When fitting the PCS dataset, the quality factor referred to the open structure rises up to 0.231 while, when referred to a closed-like structure, it reports a value of 0.281. These results demonstrate that a single structure is not able to explain the experimental dataset, likely because the N-terminal domain fluctuates with respect to the C-terminal. A model contemplating an average ensemble of differently populated states was tested and a combination of the two above mentioned conformations (68% for the *apo\_op* and 32% for the *holo\_cl*) provides a very good correlation with the experimental data (Figure 4.6, panel E).



**Figure 4.6.** Correlation between experimental and back-calculated restraints relative to *apo\_opGGBP* (blue), *holo\_clGGBP* (orange) and an average of 68% and 32% of both structures respectively (violet). (A) PCSs of  $H^N$ , N nuclei belonging to only the protein tagged domain. (B) PCSs of  $H^N$ , N nuclei of the entire protein. (C) RDCs of  $H^N$ -N pairs of different selected aminoacids in *apo\_opGGBP*. Label 4 in figure 4.7. (D) RDCs of  $H^N$ -N pairs of different selected aminoacids in *holo\_clGGBP*. Label 8 in figure 4.7. (E) RDCs of  $H^N$ -N pairs of different selected aminoacids in *apo\_clGGBP*. This latter structure is an ensemble average structure of 68% *apo\_op* and 32% *holo\_cl* GGBP. The values for hinge and twist angles are  $137^\circ$  and  $40^\circ$  respectively and its position in the conformational space is specified in figure 4.7.

Thus, the RDCs and PCSs experimental data reported here on *apoGGBP* fully agree with a model where the *apo* state undergoes a rapid equilibrium between a major and a minor species, the latter one occupying a region of the conformational landscape similar to the ligand bound form (Figure 4.7, panel C). These results also demonstrate the existence of the postulated dynamic equilibrium between open and partially closed *apo* states and gives credit to the hypothesis that large-scale domain rearrangements are already present in many two-domain periplasmic proteins.

**4.2.2 MD simulations.** Due to the dynamic nature of *apo*GGBP, MD simulations have been employed to investigate its conformational landscape. In a first set of calculations, the experimental RDCs were included as restraints in a conjoined rigid body-torsion angle simulated annealing followed by an MD simulation of 200 ps. (see Methods). An intermediate conformation, partially closed, complies well with the experimental restrains and can be interpreted in terms of a combination of the open (68%) and closed (32%) states. Specifically, the values for the inter-domain hinge/twist angle that satisfies the NMR restraints lies around  $135 \pm 5^\circ / 40 \pm 10^\circ$  (see Figure 4.7).



**Figure 4.7.** Trajectories of interdomain hinge (A) and twist (B) angles along the MD simulations for *apo\_op*GGBP(300ns) and *holo\_cl*GGBP (100ns) The derived 200ns of collective motion for *holo\_cl*GGBP are marked with tenuous line. (C) Free-energy landscapes of *apo\_op*GGBP (blue) and *holo\_cl*GGBP (orange) as a function of Hinge ( $\theta$ ) and Twist ( $\phi$ ) angles. Selected snapshot along the trajectory, red dashed line, are labelled with numbers in circles. From *apo\_op* MD (1-3), (5-6); (4) is the X-ray structure (2FW0); (4) and (8) have been used for derive ensemble average population from experimental NMR restraints. From *holo\_cl* MD (8-9); From SMD (7) is the high energy structure corresponding to the transition state.

The values are similar but statistically different from the consensus conformation obtained from fitting experimental diamagnetic RDCs in weakly aligned media ( $\theta$   $127^\circ$ ;  $\phi$   $32^\circ$ )<sup>[14]</sup>, thus suggesting the existence of inter-domain dynamics. To further characterize such motions, a second calculation, without including the experimentally determined NMR constraints (i. e. PCSs and RDCs) for avoiding biases in the conformational landscape, was accomplished. Two MD simulations, one starting from X-ray structures of *apo\_op* (2FW0) and other from *holo\_cl* GGBP bound to glucose (2FVY), were performed with the ff10\_Amber and GLYCAM\_06h force fields in

explicit water at 310 K to generate datasets of atomic coordinates that describe the protein ensemble. Due to the expected higher dynamic complexity, MD simulation for *apo\_op* run for 300 ns, while *holo\_cl* run for a time of 100 ns (figure 4.7). It is important to emphasize that both free-MD simulations are long enough to explore a wide conformational space and to define the periodic dynamic behavior of GGBP (*apo* and *holo*) in solution. After excluding the preparatory steps from the trajectory, MD trajectories were analyzed in terms of the inter-domain hinge and twist angles (Figure 4.7, panels A-B). For *apo\_op* GGBP, a range of twist and hinge angle are accessible, oscillating between a partially closed conformation ( $\theta$  135°;  $\phi$  40°) and a widely open conformation ( $\theta$  170°;  $\phi$  160°), consistent with previous studies.<sup>[6]</sup> Conformational interconversion in *apo\_op* GGBP involves concerted changes in both angles: the twist angle fluctuates between 30° and 160°, with equally low energy conformers, and it is always accompanied by a hinge angle oscillation between 140° and 160°. The time scale for such collective dynamics, between crest and wave of the periodic motion, is around 25 ns. As expected, MD simulation for *holo\_cl* GGBP structure is characterized by minor excursions within the conformational space. The hinge and the twist angles oscillate between  $\pm 10^\circ$  and  $\pm 20^\circ$  respectively around the starting values. The overall Q-factor value estimated for the ensemble structure between the *apo\_op*(X-ray) (label 4 in Figure 4.7) and *holo\_cl*, (label 8 in Figure 4.7) markedly improves with respect the single conformation ones (see Figure 4.6, panel C-E). The results demonstrate that the conformational behavior of GGBP in its unbound state is compatible with a dynamic process that involves partially closed conformations suggesting that ligand recognition event cannot be described by pure “induced-fit” or pure “conformational selection” models.

The free energy of the conformational landscape explored by the MD simulations has been estimated from the populations by dividing the conformational space into regular intervals of (2° x 4°) hinge/twist angles. The population for each interval has been converted into the energy difference with respect to the highest populated one according to the expression:

$$\Delta G_{B \rightarrow A} = \frac{R T \ln P_A}{P_B}$$

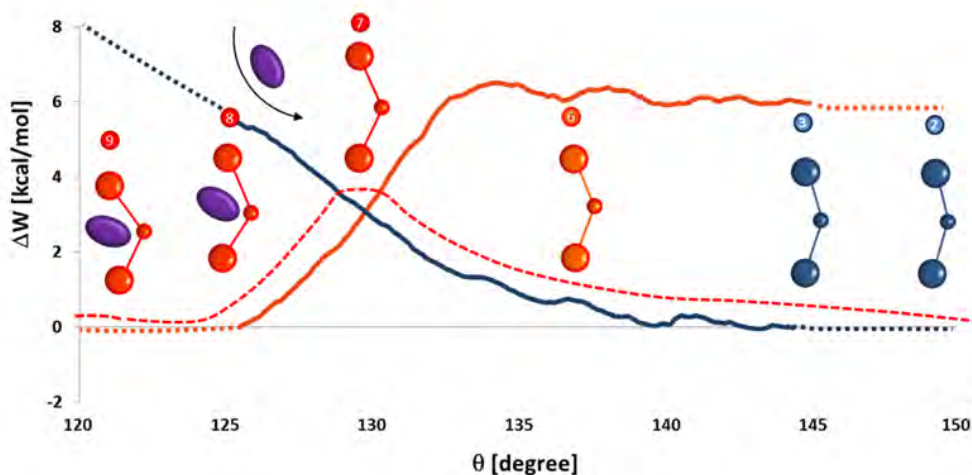
Where R is the gas constant, T the temperature, P<sub>A</sub> and P<sub>B</sub> are the populations for all the considered intervals along the conformational space.

---

The free energy landscape for *apo\_op*GGBP (Figure 4.7, panel C) shows a wide global minimum corresponding to an ensemble of highly variable structures in terms of twist/hinge angles.

Remarkably, a variety of conformations expanding the open-to-close conformational coordinate (i.e. superopened and closed-like conformations) have a free energy excess of only 1-2 kcal·mol<sup>-1</sup> as compared to the global minimum, indicating that these structures are also accessible. As inferred from the *apo\_op*GGBP MD trajectory (Figure 4.7, panels A-B), these conformations are periodically revisited, demonstrating that they are not transiently populated high energy states. On the other hand, the free energy landscape for *holo\_cl*GGBP fluctuates around a global minimum, structurally corresponding to the starting conformation of the MD calculation. Large scale conformational changes can follow multiple trajectory processes. Nevertheless, the two independent molecular dynamic simulations seem to energetically coalesce, defining the conformational coordinate for the most likely open-to-close trajectory.

In order to well define the energy profile associated with the open-to-close transition in GGBP, steered molecular dynamic simulations (SMD) have performed. This computational approach has been extensively used to calculate the free energy associated to unfolding/refolding pathways of macromolecules,<sup>[26,27]</sup> ligand-receptor binding events<sup>[28]</sup> or DNA stretching.<sup>[29]</sup> SMD employs a pulling force to cause a structural change so that different conformations may be sampled along a given pathway. To gain access to timescales that would be otherwise computationally too demanding, SMD simulations accelerate (or force) the dynamic process, thus providing the required energy associated to the induced change: while the force is executed and the motion occurs along a given coordinate, the potential energy of the system is calculated and the potential of mean force (PMF) is related to the free energy profile of the process.<sup>[30]</sup> Even if we cannot exclude the existence of other trajectories for the open-to-close event, the method ensures that the work done on the system is a function of the activation energy associated to the process. Two SMD simulations have been carried out. In the first one the value of the critical hinge angle was fixed to 124°, which corresponds to the X-ray structure determined for the *holo\_cl* conformation, and a pulling force was applied to reach the *holo\_op* conformation. In the second one the starting geometry corresponds to the X-ray structure of the unbound open GGBP (*apo\_op*), with a hinge angle of 145°, in order to achieve the closed unbound (*apo\_cl*) structure. (Figure 4.8, orange line) plots the resistance applied by the system in opening *holo\_cl*GGBP as a function of the hinge angle. According to the SMD simulations, the transition state corresponds to a hinge angle of 132° and its height is only ≈3.5 kcal·mol<sup>-1</sup> higher than the free energy of the open structure.



**Figure 4.8.** The potential of mean force as a function of the hinge angle  $\theta$  ( $^{\circ}$ ) in protein/ligand binding. The calculated energy profile relative to *holo\_cl* structure (orange) indicates that the free energy minimum for closed conformation is lower than that of the open state by roughly  $6 \text{ kcal}\cdot\text{mol}^{-1}$ . Superimposed is the energy profile for *apo\_op* structure (blue). The result indicates that the energy of the open structure increases with the closing of the hinge angle. On top, structural change associated with inter-domain closure and ligand binding. The structures correspond to snap-shots along the trajectory.

These data allow concluding that the energetic barrier for open-to-closed interconversion can be easily crossed via thermal fluctuations or with the help of water molecules that attack interdomain hydrogen bonds. Not surprisingly, giving value to the function of the protein, the closed structure is significantly stabilized by intermolecular interactions with the ligand (hydrogen bond and  $\text{CH}/\pi$  interactions) that contribute with  $6 \text{ kcal}\cdot\text{mol}^{-1}$  to the binding energy of the complex. The thick hydrogen bond network between the polar face of the sugar and the polar aminoacids on both domains of GGBP and the two aromatic residues Trp181 and Phe16 on the C-terminal and N-terminal domain, respectively, stacked against the less polar faces of the sugar provide the necessary enthalpy contribution to stabilize the *holo\_cl*GGBP structure. The second SMD simulation starting in the *apo\_op*GGBP is also shown in (Figure 4.8, blue line). In here, a quasi-linear energy dependence is observed in the transition from the open to the closed states of GGBP, with an associated Gibbs free energy that is inversely proportional to the closing angle. The stability of the protein decreases at a rate of about  $300 \text{ cal}\cdot\text{mol}^{-1}\cdot\text{deg}^{-1}$ . Considering a range of  $3 \text{ kcal}\cdot\text{mol}^{-1}$  accessible by thermal fluctuations at 310 K, domain closure of  $10^{\circ}$  to  $20^{\circ}$  is expected. Then, the *apo\_cl* conformation of the protein, which is almost  $6 \text{ kcal}\cdot\text{mol}^{-1}$  higher in energy than the *apo\_op*, would be inaccessible by thermal fluctuations, consistent with previous results reported for MBP.<sup>[11]</sup> The intersection between the two independent SMD define the

transition state, where only the incoming ligand provides the necessary energy stabilization to shift the equilibrium in favor of the most stable ligand-protein complex.

#### **4.2.3 The synergy between structure and dynamics is essential for proteins function.**

Segmental dynamics in proteins are often functional being at the basis of protein allostery. Yet, they are often very loosely characterized due to the lack of experimental tools available. Here, we demonstrate that a combination of NMR spectroscopy and MD simulations successfully unravels the energy landscape for the functional conformational coordinate of GGBP.

A single structural model is not enough to describe the vast conformational space covered by *apo\_op*GGBP in solution. According to unrestricted MD simulations, the conformational dynamics are well characterized by large amplitude motions between hemi-closed and open structures in the nanosecond time scale. Actually, the first reported X-ray structure of *apo*\_GGBP is partially closed.<sup>[6]</sup> The authors noticed the presence of the citrate ion in the binding cleft and hypothesized that they may be thermodynamically relevant for crystal growth. Our results agree, demonstrating that the ion coordinates at the hinge position, thus stabilizing the hemi-closed structures in *apo*\_GGBP. Importantly, these conformational interconversions are experimentally validated: while the combined use of PCSs and RDCs induced by self-alignment of paramagnetic metal provides a fine method to unveil the existence of functional closed conformations that are consistent with a 32% of the total population of *apo*\_GGBP. Albeit the limited range of hinge angles for the deposited X-ray structure of *apo\_op*GGBP that spans between 147° and 149°, our data suggest that excursions toward more open states are permitted (150° and 170°, “superopen” structures). Such conformations have been also predicted and observed for other bi-domain proteins like uroporphyrinogen III synthase<sup>[31]</sup> and it is probably dependent on the hinge region structure and composition since it reflects the maximum torque force that this region allows. On the other hand, closed like states, between 132° and 147°, are also represented supporting the existence of closed like structures for *apo*GGBP as observed via X-ray crystallography.<sup>[32]</sup> The inter-domain excursions found for *apo*\_GGBP are largely quenched in *holo*\_GGBP but, interestingly, they are not completely abrogated (Figure 4.7). Actually, when bound to glucose, *holo\_c*GGBP still holds some degree of flexibility, with inter-domain librations of up to 10-15° in the closure and hinge angles. According to MD simulations, such motions also fall in the nano-second timescale and they are consistent with the increased dynamics previously observed in the order parameter of *holo*\_GGBP,<sup>[14]</sup> that likely interfered on the spectral density function motional parameters. The functional role of these dynamics naturally emerges when the energetics are introduced and the integrative



analysis of MD simulations and NMR analysis comply largely with a mainly conformational selection mechanism for which the presence of the ligand is indispensable for the open-to-closed transition. MD simulations reveal a gap in the conformational landscape between *apo\_op*GGBP and *holo\_cl*GGBP (125-140° hinge and 30-40° twist, Figure 4.7, panel C). This high energetic point corresponds to the transition state with an activation barrier of  $\approx 3,5 \text{ kcal}\cdot\text{mol}^{-1}$  as estimated by SMD simulations (Figure 4.8). Thus, ligand binding provides the energy to overcome such barrier and shift the conformational ensemble towards *holo\_cl*GGBP in a second step that agrees well with an induced fit mechanism. Remarkably, *apo\_op*GGBP is predicted to unfold at low values of the closing angle due to the increase in nonpolar character in the solvent accessible area located on the hinge region on the opposite site respect to the ligand binding pocket. This negative term, if not balanced by the enthalpic surplus of the binding event, will result in the protein unfolding. This mechanism is equivalent to the that experimentally found for maltose binding protein, where the analysis of hinge mutants with different closure angles shows protein unfolding at low closure angles, within a similar free energy range.<sup>[11]</sup>

### 4.3 Conclusions

In conclusion, the amplitude and time scale of GGBP inter-domain dynamics have been unveiled by NMR spectroscopy and detailed MD simulations. The population contribution of bio-functional relevant conformers has been determined by PCSs and RDCs induced by paramagnetic metal ion attached to the protein. Subsequently, the energetic barrier in open-to-closed transition has been defined by non-equilibrium MD simulations. We conclude that the population of *apo\_cl* like conformations is essential in activate the transition to *holo\_cl* form, according to a conformational selection mechanism coupled to a final rearrangement that obeys an induced-fit kinetics process.

From a general perspective, the protocol exemplified herein can be extended to the study a variety of molecular recognition processes in which significant molecular rearrangements take place, thus expanding the limits of the application of NMR methods when exploring binding events.

## 4.4 Methods

**4.4.1 Sample preparation for Paramagnetic studies.** Protein samples for NMR studies were prepared at a final protein concentration of 0.5 mM in 20mM Tris (pH 7.0), 150mM NaCl, 10 M CaCl<sub>2</sub>, 1mM NaN<sub>3</sub> with 10% D<sub>2</sub>O. In order to conjugate the paramagnetic tag to the C-terminal domain cysteine, the protein was titrated with 10mM solution of lanthanide (Ln) chelating tag previously loaded with the lanthanide (Ln = La<sup>3+</sup>, Dy<sup>3+</sup>). The titrations were performed by monitoring the changes in the chemical shift in a <sup>1</sup>H, <sup>15</sup>N HSQC spectrum. The nucleophilic substitution reaction is instantaneous and the excess of chelating tag molecule was removed by filtration.

**4.4.2 NMR Spectroscopy.** All the NMR experiments were carried out on a 600 MHz AVANCE-III Bruker spectrometer. Spectra were acquired at 310 K. All NMR spectra were processed with the software TopSpin. The Program CARR was used for the analysis of the 2D spectra. Lanthanum was used as diamagnetic reference as it has an ionic radius similar to the paramagnetic dysprosium. PCSs were measured as the difference between the chemical shift of the corresponding nuclei in the paramagnetic and diamagnetic samples. Residual dipolar couplings <sup>1</sup>D<sub>HN</sub> were measured as the <sup>1</sup>H-doublet splitting of the paramagnetic sample minus the equivalent splitting difference in the diamagnetic sample.

PCSs and RDCs analyses were performed using MSpin software. In order to assess inter-domain dynamics, we used selected conformations from molecular dynamics simulation, including some that resemble the experimentally obtained x-ray open and closed GGBP conformations. The optimal ensemble of conformations was found after evaluating the tensor from the tag domain using both PCSs and RDCs independently. The PCSs and RDCs for the other domain were back calculated for different protein coordinates and the ensemble of structures that better fits the experimental values was selected on the basis of best quality factor.

**4.4.3Molecular Dynamic simulations.** The coordinates from the X-ray structure of *apo\_op* GGBP (PDB code 2FW0) and *holo\_cl* GGBP (PDB code 2FVY) were used as starting points to generate intermediate models by molecular dynamics (MD) simulations at 310 K applying the ff10 and GLYCAM\_06h<sup>[33]</sup> amber force fields. Missing hydrogens were added to the starting PDB structures using the program LEAP. The N- and C-terminal residues were acetylated and amidated according to the AMBER standard database. The latter structure was solvated in a cubic TIP3P water box[ and eight sodium ions were added to neutralize the system. In order to fulfill all the proteins cavities by water molecules, a previous minimization for only solvent and ions was made. Moreover, to reach a reasonable starting structure, the entire system was

minimized with a higher number of cycles, using the accurate steepest descent algorithm. The system was subjected to two short molecular dynamic simulations of 20ps and 100 ps respectively, before start the real dynamic simulation of 270 ns for *apo\_op* and 100 ns for *holo\_cl*GGBP. During these two preparatory steps, the structure was slowly heated from 0 to 310 K. Fifty-thousand additional steps were performed to switch from constant volume to constant pressure. A relaxation time of 2 ps was used in order to equilibrate the entire system in each step. The final simulations of 270 and 100 ns were performed starting from equilibrated structures. Coordinates and energy values were recorded every 2 ps for a total of 135000 MD frames for *apo\_op* and 50000 for *holo\_cl*GGBP. For the SMD (Steered Molecular Dynamics) simulations, the starting structures, with a hinge value of 124° and 147° have been extracted by unrestricted MD simulations of *holo\_cl* and *apo\_op*GGBP respectively, so the entire system was already equilibrated. The center of mass of the N-terminal domain together with the center of mass of the hinge segment has been fixed, while the center of mass of the C-terminal domain has been pulled with a constant force  $K = 500 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^2$ . The total time for the molecular dynamics simulation was of 10ns with an angle opening or closing of 2 degree per ns. Atomic coordinates were saved every 2ps and the energy information extracted.

Constrained MD simulations were initiated from the X-ray structures *apo\_op*(2FW0) and *holo\_cl*(2FVY), pointed as blue and orange circles, numbered 4 and 8 respectively, in figure 4.7. Conjoined rigid body-torsion angle simulated annealing was performed as previously described<sup>[34,14]</sup>. The hinge region for the GGBP is defined by residues 109-111, 253-256 and 293-296. The starting structure (*apo\_op*, F2W0) was heated to 600 K for 3 ps with *tautp* equal to 0.4. Then, the system was cooled to 100 K for 297 ps (*tautp* = 4.0). The final cooling to 0 K was carried out for 100 ps with *tautp* varying from 1.0 to 0.1. The  $\chi$  tensor anisotropy parameters, calculated with MSpin software, were extracted from the structures with lowest Q-Factors. MD simulations were performed with ff10\_Amber and GLYCAM\_06h force fields incorporating the experimentally derived NMR restraints. The restrained MD calculations were performed in explicit water solvent and using a simulated annealing approach.

**4.4.4 Analysis of the trajectories.** Root mean-square deviation (RMSD) and thermodynamic data were monitored throughout the whole trajectories to confirm that all simulations evolved along a stable plateau. For the analysis of the collective motions, the closure (hinge), twisting and bending coordinate system were used.<sup>[35]</sup> Hinge and twist angles, were obtained from clusters of one in every 100 models/frames (1350 in total) for *apo\_op* and one in every 50 models/frames for the closed bound GGBP conformation (*holo\_cl*GGBP). For all structures the values of hinge  $\theta$  and twist  $\phi$  angles were calculated using an in house scrip run in Matlab to adequately represent

the conformational landscape in terms of the hinge and twist angles. For the evaluation of the tag-domain data, the structures were aligned with respect to the backbone of residues 112-254. The Tag molecule was accommodated for all the structures and a rapid minimization on the tag region was made. The structures were visualized and evaluated by using the programs VMD and Discovery studio.

#### 4.5 References

1. Shilton, B. H., Flocco, M. M., Nilsson, M. & Mowbray, S. L. Conformational Changes of Three Periplasmic Receptors for Bacterial Chemotaxis and Transport: The Maltose-, Glucose/Galactose- and Ribose-binding Proteins. *J. Mol. Biol.* **264**, 350–363 (1996).
2. Tam, R. & Saier, M. H. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **57**, 320–346 (1993).
3. Loeffler, H. H. & Kitao, A. Collective Dynamics of Periplasmic Glutamine Binding Protein upon Domain Closure. *Biophys. J.* **97**, 2541–2549 (2009).
4. Stockner, T., Vogel, H. J. & Tieleman, D. P. A Salt-Bridge Motif Involved in Ligand Binding and Large-Scale Domain Motions of the Maltose-Binding Protein. *Biophys. J.* **89**, 3362–3371 (2005).
5. Vyas, N. K., Vyas, M. N. & Quioco, F. A. Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and Functional Similarity. *J. Biol. Chem.* **266**, 5226–5237 (1991).
6. 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.* **16**, 1032–1041 (2007).
7. Borrok, M. J., Zhu, Y., Forest, K. T. & Kiessling, L. L. Structure-based design of a periplasmic binding protein antagonist that prevents domain closure. *ACS Chem. Biol.* **4**, 447–456 (2009).
8. Cuneo, M. J., Changela, A., Warren, J. J., Beese, L. S. & Hellinga, H. W. The crystal structure of a thermophilic glucose binding protein reveals adaptations that interconvert mono and di-saccharide binding sites. *J. Mol. Biol.* **362**, 259–270 (2006).
9. 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-beta-D-galactopyranoside. *FEBS J.* **276**, 2116–2124 (2009).
10. Tian, Y. *et al.* Structure-based design of robust glucose biosensors using a *Thermotoga maritima* periplasmic glucose-binding protein. *Protein Sci.* **16**, 2240–2250 (2007).
11. Wemmer, D. E. The energetics of structural change in maltose-binding protein. *Proc. Natl. Acad. Sci.* **100**, 12529–12530 (2003).

12. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449**, 1078–1082 (2007).
13. Bermejo, G. A., Strub, M.-P., Ho, C. & Tjandra, N. Ligand-Free Open–Closed Transitions of Periplasmic Binding Proteins: the Case of Glutamine-Binding Protein. *Biochemistry* **49**, 1893–1902 (2010).
14. Ortega, G., Castano, D., Diercks, T. & Millet, O. Carbohydrate affinity for the glucose-galactose binding protein is regulated by allosteric domain motions. *J. Am. Chem. Soc.* **134**, 19869–19876 (2012).
15. Ravindranathan, K. P., Gallicchio, E. & Levy, R. M. Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *J. Mol. Biol.* **353**, 196–210 (2005).
16. Wereszczynski, J. & McCammon, J. A. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* **45**, 1–25 (2012).
17. Dwyer, M. A. & Hellinga, H. W. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr. Opin. Struct. Biol.* **14**, 495–504 (2004).
18. Jeffery, C. J. Engineering periplasmic ligand binding proteins as glucose nanosensors. *Nano Rev.* **2**, (2011).
19. Vercillo, N. C., Herald, K. J., Fox, J. M., Der, B. S. & Dattelbaum, J. D. Analysis of ligand binding to a ribose biosensor using site-directed mutagenesis and fluorescence spectroscopy. *Protein Sci.* **16**, 362–368 (2007).
20. Salins, L. L., Ware, R. A., Ensor, C. M. & Daunert, S. A novel reagentless sensing system for measuring glucose based on the galactose/glucose-binding protein. *Anal. Biochem.* **294**, 19–26 (2001).
21. Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S. & Griesinger, C. Interdomain Dynamics Explored by Paramagnetic NMR. *J. Am. Chem. Soc.* **135**, 17111–17120 (2013).
22. Canales, A. *et al.* Breaking pseudo-symmetry in multiantennary complex N-glycans using lanthanide-binding tags and NMR pseudo-contact shifts. *Angew. Chem. Int. Ed. Engl.* **52**, 13789–13793 (2013).
23. Bertini, I. *et al.* Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6841–6846 (2004).
24. Haberz, P. *et al.* Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution. *Org. Lett.* **8**, 1275–1278 (2006).
25. Tjandra, N., Kuboniwa, H., Ren, H. & Bax, A. Rotational dynamics of calcium-free calmodulin studied by <sup>15</sup>N-NMR relaxation measurements. *Eur. J. Biochem.* **230**, 1014–1024 (1995).
26. Gao, M., Wilmanns, M. & Schulten, K. Steered molecular dynamics studies of titin I1 domain unfolding. *Biophys. J.* **83**, 3435–3445 (2002).
27. Krammer, A., Lu, H., Isralewitz, B., Schulten, K. & Vogel, V. Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proc. Natl.*

- Acad. Sci. U. S. A.* **96**, 1351–1356 (1999).
28. Gu, J., Li, H. & Wang, X. A Self-Adaptive Steered Molecular Dynamics Method Based on Minimization of Stretching Force Reveals the Binding Affinity of Protein–Ligand Complexes. *Molecules* **20**, (2015).
  29. Harris, S. A., Sands, Z. A. & Laughton, C. A. Molecular Dynamics Simulations of Duplex Stretching Reveal the Importance of Entropy in Determining the Biomechanical Properties of DNA. *Biophys. J.* **88**, 1684–1691 (2005).
  30. Perišić, O. & Lu, H. On the Improvement of Free-Energy Calculation from Steered Molecular Dynamics Simulations Using Adaptive Stochastic Perturbation Protocols. *PLoS One* **9**, e101810 (2014).
  31. Schubert, H. L., Phillips, J. D., Heroux, A. & Hill, C. P. Structure and Mechanistic Implications of a Uroporphyrinogen III Synthase–Product Complex,. *Biochemistry* **47**, 8648–8655 (2008).
  32. Flocco, M. M. & Mowbray, S. L. The 1.9 Å x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Biol. Chem.* **269**, 8931–8936 (1994).
  33. Kirschner, K. N. *et al.* GLYCAM06: A generalizable biomolecular force field. carbohydrates. *J. Comput. Chem.* **29**, 622–655 (2008).
  34. Clore, G. M. & Bewley, C. A. Using conjoined rigid body/torsion angle simulated annealing to determine the relative orientation of covalently linked protein domains from dipolar couplings. *J. Magn. Reson.* **154**, 329–335 (2002).
  35. Skrynnikov, N. R. *et al.* Orienting domains in proteins using dipolar couplings measured by liquid-state NMR: differences in solution and crystal forms of maltodextrin binding protein loaded with beta-cyclodextrin. *J. Mol. Biol.* **295**, 1265–1273 (2000).

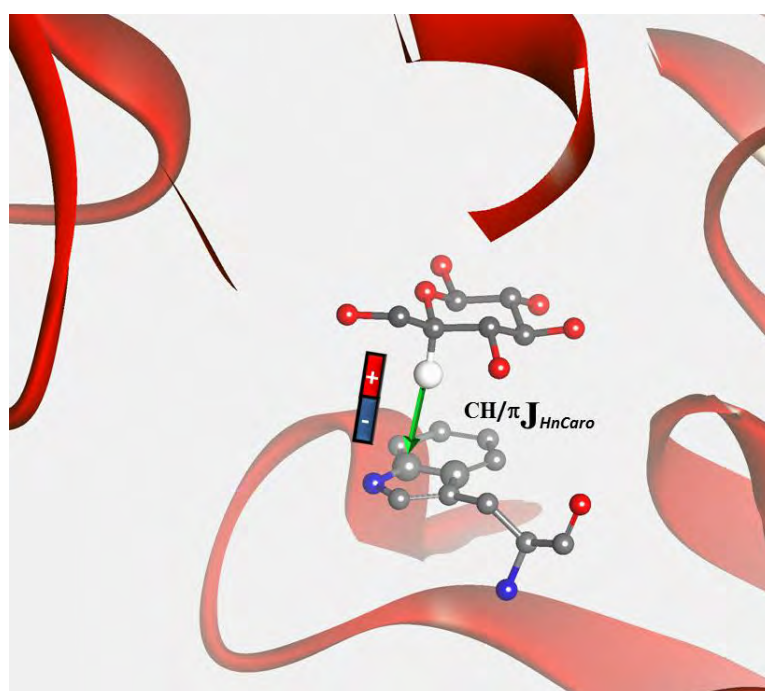


## CHAPTER V

*CH/π* interactions: elucidating fundamental weak intermolecular forces

### Towards the direct detection of *CH/π* interactions in sugar/protein complexes

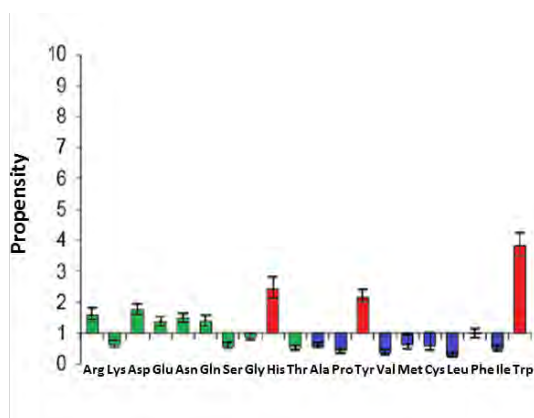
The work presented in this chapter has been performed in collaboration with the group of Dr. Niels Reichardt, CIC biomaGUNE, at San Sebastian, Spain. The synthesis of the deuterated sugar discussed herein was performed as part of my secondment therein. Expression and purification of the triple labeled [<sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N] protein was carried out in the laboratory of Dr. Oscar Millet, CIC bioGUNE, Derio, Spain, by Dr. Sivanandam Veeramuthu.





## 5.1 Introduction

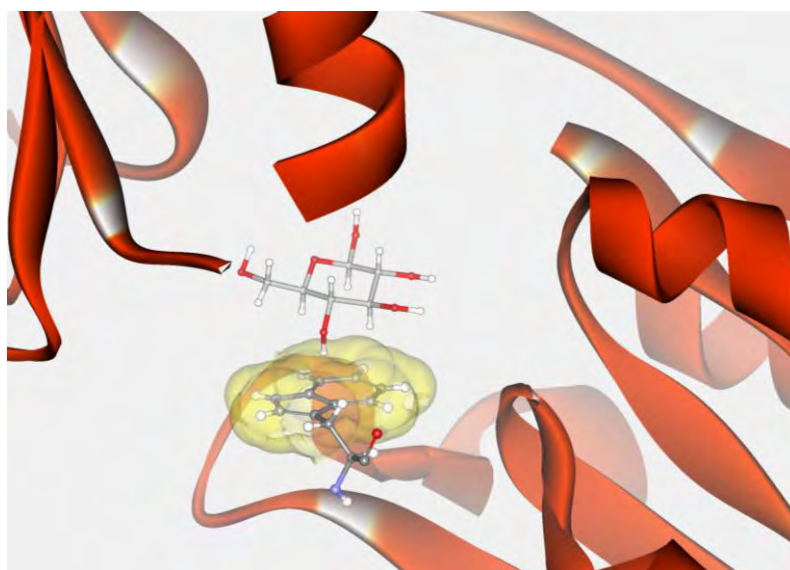
The characterization of protein-carbohydrate interactions is experiencing a growing interest due to their implications in diverse biological and biomedical processes. In parallel, the control and manipulation of these interactions would provide clues to dose and predict biological responses, as consequence of the ligand-protein recognition process.<sup>[1]</sup> Therefore, the achievement of robust SAR (Structure Activity Relationships) in this context is a key objective, following the paradigm developed by the pharmaceutical industry.<sup>[2,3,4]</sup> However, it is obvious that this is not an easy task, specially due to the incomplete quantitative understanding of the physical-chemical laws behind weak intermolecular interactions. The importance of hydrogen bonds between the sugar hydroxyl groups and the polar moieties of amino acids at the protein binding pocket has been well recognized. A major advance in the study of the importance of such hydrogen bonds has come with the direct evidence that the intermolecular scalar (J) coupling between the acceptor and donor nuclei can be measured using high-resolution nuclear magnetic resonance (NMR) spectroscopy experiments.<sup>[5,6]</sup> This approach has allowed the unambiguous identification of the involved donor and acceptor groups and the study of the relevance of the individual hydrogen bonds. Besides hydrogen bonds, CH- $\pi$  interactions are also of paramount importance to stabilize sugar/protein complexes,<sup>[7]</sup> as well as other supramolecular structures.<sup>[8,9,10]</sup> In the glycosciences field, their importance is assessed by the wide presence of aromatic amino acids in the binding pocket of saccharide binding proteins. Figure 5.1.



**Figure 5.1.** Amino acids proximal to carbohydrates in the X-ray diffraction structures of protein-glycan complexes, as compared to the distributions of amino acids across all PDB structures.<sup>[11]</sup>

CH- $\pi$  interactions are usually described as weak *polar* interactions in which the delocalized system of  $sp^2$ -hybridized covalent bonds can act as an acceptor group, while the hydrogen atom

on the *polarized* C-H moiety acts as donor. Despite their initial classification as non-conventional hydrogen-bond-like interactions, the dual dispersive/electrostatic nature of these CH- $\pi$  forces has been a matter of debate for years.<sup>[12,13,14,15,16,17]</sup> Theoretical analyses have revealed that, in the gas phase, the stability of CH- $\pi$  complexes largely arises from dispersion forces.<sup>[18,19,20]</sup> However, detailed investigations of different simple systems, including benzene and water or ammonium or methane and benzene have revealed the existence of intermolecular hydrogen-bond-like interactions in which the  $\pi$ -electrons of benzene act as an acceptor group. The interaction energies oscillate between 1 and 4 kcalmol<sup>-1</sup> when the donor is methane, ammonium or water respectively, demonstrating the importance of the polarity of the donor group.<sup>[21]</sup> Specifically, in sugars, the CH donors groups are polarized by the geminal hydroxyl groups, making the CH vector a suitable moiety for establishing significant interactions.<sup>[22]</sup> From the point of view of the acceptor, a direct correlation between the type of aromatic amino acid and their electrostatic potential surface (ESP) has been found. Actually, the ranking of amino acids involved in CH- $\pi$  interactions is Trp > Tyr > Phe > His. This order reflects the ESPs of these side chains and implies that electron-rich aromatic systems are the most likely to be engaged in CH- $\pi$  interactions.<sup>[11]</sup> In 2015, Asensio and co-workers provided convincing evidences that polarization of the CH moieties in water largely stabilizes the CH- $\pi$  contacts, which are further exacerbated by the hydrophobic “solvent cage” effect. Thus, electrostatics and charge transfer forces are remarkably relevant.<sup>[17]</sup> As a consequence, it appears plausible that NMR could provide direct evidences for the existence of a weak polarized and directional electron density between the aromatic  $\pi$  system, the acceptor at the protein site, and the C-H donor at the ligand. In spectroscopic terms, the existence of this non-covalent bond could be encoded by a weak scalar intermolecular (J) coupling between the hydrogen nucleus on the sugar and the sp<sup>2</sup> carbon nucleus on the aromatic moiety of the amino acid. However, the direct detection of extremely weak intermolecular (J) coupling constants is not exempt of severe methodological problems. Sugar-protein interactions are usually rather weak, with K<sub>D</sub> in mM- $\mu$ M range. From the experimental point of view, this means that the life time for the complex may be too short to be detected within the different NMR time scales. Under this premise, an unusually tighter glycan binder is necessary. As anticipated in chapter IV of this thesis, the  $\beta$  anomer of glucose binds to GGBP, the glucose/galactose binding protein, with high affinity and specificity, K<sub>D</sub> = 290 nM at 37° C and pH 7.0.<sup>[23]</sup> The binding site of GGBP presents a tryptophan amino acid that is stacked against the less polar  $\alpha$  face of the sugar, engaging protons 1, 3 and 5 in a well-defined CH- $\pi$  interaction as shown in Figure 5.2. Therefore, we guessed that the glucose/GGBP system could be suitable for direct detection of the intermolecular (J) coupling.

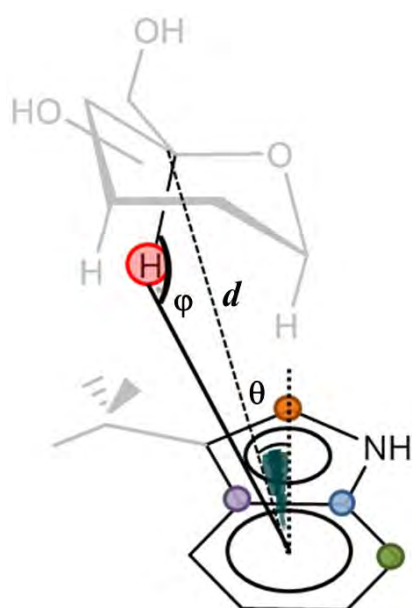


**Figure 5.2.** The GGBP binding pocket. The Trp183 surface is highlighted. The CH-1, -3 and -5 vectors of the sugar point towards the  $sp^2$ -hybridized electron cloud of the aromatic ring.

First, we tested our hypothesis *in silico*. Previous works have shown that hybrid density functional theory (DFT) calculations may predict the existence of weak J couplings between the nuclei involved in Me- $\pi$  intramolecular interactions in well-structured proteins domains.<sup>[24]</sup> We believed that with the proper experimental set up and with the aid of an optimized isotope-labelling scheme, both on the ligand and protein partners, the system could be suitable for our aims.

## 5.2 Results and Discussion

**5.2.1 DFT prediction of J couplings between the nuclei involved in CH- $\pi$  intermolecular interactions.** A series of DFT calculations were first performed to assess whether measurable J couplings indeed exist between the donor and acceptor nuclei involved in CH- $\pi$  interactions. A model system consisting of  $\beta$ -D-glucose (Glc) and tryptophan (Trp) was used to explore the variation of  $J_{Hn-CarO}$  couplings within the parameters describing the geometry of the CH- $\pi$  interaction. Consistent with previous published data,<sup>[24]</sup> three geometric parameters were used to identify the putative CH- $\pi$  interactions:  $d$ , the distance between the donor carbon atom (CH on the sugar) and the centre of the acceptor ring;  $\theta$ , the angle between the ring normal and a vector connecting the CH carbon atom and the centre of the ring, and  $\varphi$ , the angle between the C-H and the ring centre-H vectors. Figure 5.3.



**Figure 5.3.** Schematic view of the three parameters ( $d$ ,  $\theta$  and  $\varphi$ ) used to describe CH/ $\pi$  interactions. The atoms that share J coupling constants are labelled as circles. The colour coding corresponds to that used in table 5.1.

The results are gathered in table 5.1. Due to the employed sugar isotope labelling scheme (vide infra), the only CH moiety considered in table 5.1 is C(5)-H; however, the complete JHn-Caro analysis is reported in the Supporting Information (SI table 5.1).

Model	$d$ (Å)	$\theta$ (°)	$\varphi$ (°)	$J_{H5Caro}$ (Hz)
1	4.64	19.73	148.41	0.2 <span style="color:red">●</span> <span style="color:blue">●</span>
2	4.40	18.71	157.01	0.3 <span style="color:red">●</span> <span style="color:blue">●</span>
3	4.30	18.40	159.46	0.3 <span style="color:red">●</span> <span style="color:green">●</span>
4	4.31	21.53	151.45	0.3 <span style="color:red">●</span> <span style="color:purple">●</span>
5	5.14	35.20	140.83	0.1 <span style="color:red">●</span> <span style="color:orange">●</span>

**Table 5.1.** Variation of  $J_{H5-Caro}$  as a function of the  $d$ ,  $\theta$  and  $\varphi$  geometric parameters. The aromatic carbon atom that is involved in J coupling with the sugar proton H5 is marked as coloured circle, as represented in figure 5.3.

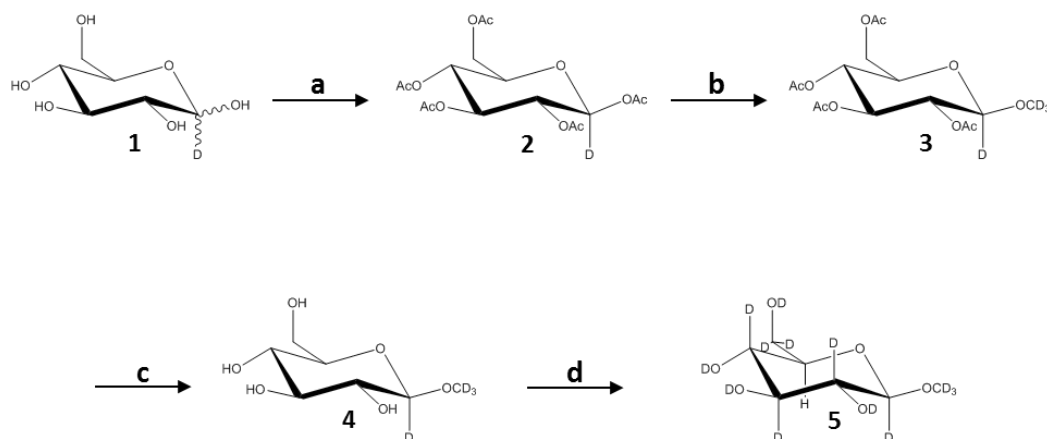
The results in table 5.1 show that the size of  $J_{H5-Caro}$  depends on the distance between the sugar carbon atom and the aromatic plane, with optimal values for  $d < 4.40$  Å. At a distance

greater than 5.0 Å, the coupling is significantly reduced. In model 5, the sugar CH vector points at the pyrrol moiety of the tryptophan, and the J coupling drastically diminishes. For models 2, 3, and 4, JH5-CarO is almost constant, even when coupled to the different sp<sup>2</sup>-hybridized carbon nuclei, depending on the different orientation of the sugar ring with respect to the Trp moiety. In fact, the location of the donor CH group above the aromatic ring greatly affects the size of the coupling. The maximum coupling is observed for model 4 (J = 0.32 Hz), which displays the CH vector directly above the CD2 aromatic carbon ( $\theta = 21.5^\circ$ ). In the X-ray structure of GGBP complexed to  $\beta$ -D-Glc, pdb code 2FVY, the geometric parameters are very similar to those defined in model 4,  $d = 4.43 \text{ \AA}$ ;  $\theta = 20.86^\circ$ ;  $\varphi = 145,21^\circ$ , suggesting that a measurable value of JH5-CarO could reasonably be expected. Nevertheless, the estimated magnitude of the JH5-CarO coupling is very weak to be detected even using long-range HCC and HMQC experiments.<sup>[24,25,26]</sup> To our purpose, the sensitivity of the NMR experiments should be improved by complete deuteration of all but the C5-H hydrogen atom on the sugar and by employing a uniformly [<sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N]-labelled protein sample. This scheme would ensure a maximal experimental sensitivity during the long transfer periods required to evolve the JH5-CarO coupling. It is well known that fractional or complete deuteration of proteins improves the resolution and sensitivity of NMR experiments.<sup>[27]</sup> The origin of this effect is that the substitution of <sup>1</sup>H for <sup>2</sup>H strongly reduces the rate of dipole-dipole relaxation of the observed nuclei, since the gyromagnetic ratio of deuterium is 6.5 times smaller than that of proton (Equation 5.1). In uniformly deuterated proteins, <sup>13</sup>C carbon and <sup>15</sup>N nitrogen atoms relax slowly, thus allowing more magnetization to be transferred between the J-coupled nuclei and thus, providing better signal/noise ratio. From the sugar point of view, the absence of J<sub>HH</sub> scalar coupling constants in the molecule also implies the presence of a sharper proton signal. Thus, the presence of a J<sub>HC</sub> coupling constant for C5H could only be originated by its long range coupling with the labelled protein carbon atoms.

$$R_{1(DD)} = \frac{1}{T_{1(DD)}} = \gamma_X^2 \cdot \gamma_H^2 \left(\frac{h}{2\pi}\right)^2 \tau_c \frac{1}{r_{HX}^6} \quad \text{Eq.5.1.}$$

The equation governing Dipole-Dipole (DD) relaxation of an X nucleus by nearby protons. Since the proton has the highest magnetic dipole of common nuclei, it is the most effective nucleus for causing DD relaxation. DD relaxation is the principal relaxation pathway for protons in regular molecules, and for nuclei with directly attached protons.  $\tau_c$  is the rotational motion correlation time, the average time for the H-X vector to rotate 1 radian.  $\gamma_X$ ,  $\gamma_H$  are the gyromagnetic ratios and  $r_{HX}$  is the distance between H and X nucleus being relaxed.

**5.2.2 Sugar deuteration.** In 2012, Sajiki and coworkers described a stereo- and region (chemo)selective deuterium-labelling method of various sugars under mild Ru/C-H<sub>2</sub>-D<sub>2</sub>O conditions.<sup>[28]</sup> The authors suggested that deuterium-labelled sugars can be utilized as powerful tools for the structural analyses of high-sugar-containing molecules as nucleic acids or glycoproteins, due to their spectroscopic advantages. To our purpose, site-specific isotope labelling on the sugar should benefit both spectral simplicity and relaxation properties, with the consequent sensitivity improvement, line sharpening and slow relaxation. The direct H-D exchange reaction selectively proceeds on carbons adjacent to free hydroxyl groups, whereas it never proceeds at the  $\alpha$ -position of protected-hydroxyl groups. In other terms, for D-Glc, the reaction should provide a selectively C(5)-H D-Glc(1,2,3,4,6,6-D6). However, the authors also showed that deuteration of D-Glc causes a complete decomposition of the starting material, presumably due to the hydrolysis of the hemi-acetal moiety and the subsequent Ru/C-catalyzed hydrogenation. Differently, the corresponding methyl glucoside underwent the H-D exchange reaction with the deuterium atoms incorporated at the 2, 3, 4 and 6 positions to give the multi-deuterated methyl glycoside. Our strategy for the synthesis of fully deuterated  $\beta$ -D-methylglucoside but C(5)-H is thus outlined in scheme 5.1.



**Scheme 5.1.** Reagents and conditions: a) Anhydrous Sodium Acetate (1.76 eq.), Acetic anhydride (9.5 eq.), 90 °C, 2h. b) CD<sub>3</sub>OD (1.1 eq.), SnCl<sub>4</sub> (1.2 eq.), 40 °C, 3h. c) MeOH/(CH<sub>3</sub>CH<sub>2</sub>)<sub>3</sub>N/H<sub>2</sub>O (2mL), RT, 24h. d) Ru/C (5mol%), D<sub>2</sub>O, 80 °C, 72h, H<sub>2</sub> atmosphere.

The synthesis started from commercially available D-[1-<sup>2</sup>H]Glc (**1**), from Cortecnet, which was peracetylated to give compound (**2**), in order to protect the hydroxyl group for the successive glycosylation reaction with deuterated methanol.<sup>[29,30,31]</sup> The participation of the acetyl group at position two, as anchimeric assistance, ensures that the main product of the reaction is the protected  $\beta$ -D-methyl glucoside (**3**), which was easily purified from the mixture. As final step, the

full-protected sugar was deprotected under basic conditions to provide  $\beta$ -D-methyl( $^2\text{H}$ -3)glucoside( $1\text{-}^2\text{H}$ ) (**4**), employed as starting material for the H-D exchange reaction. According to the authors, the H-D exchange reaction of the sugar is initiated by the formation of the  $\text{H}_2$ - and  $\text{D}_2\text{O}$ -activated Ru catalyst and it is driven by the hydroxyl group of the sugars as a directing group.<sup>[28]</sup> The subsequent oxidative addition at the C-H bond adjacent to the hydroxyl group afforded the tetra-coordinate intermediate. The intramolecular H-D exchange reaction and reductive elimination gives the regio- and stereo-selectively deuterated sugar (**5**). The details for chemical synthesis are presented in the methods section, while the complete NMR spectra are provided in the SI section.

### 5.3 Partial Conclusions

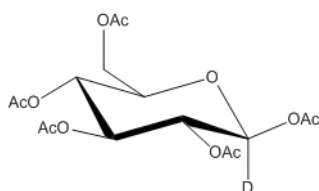
DFT calculations predict that intermolecular J coupling constants indeed exist in a simple model constituted by a glucose unit and a tryptophan amino acid. The values for the direct J coupling indicate that this spectroscopic parameter could be detected with the aid of an optimized isotope-labelling scheme. We have designed a synthetic procedure to obtain a single proton labelled glucose. The fully deuterated  $\beta$ -D-methylglucoside but C(5)-H has been synthesized successfully. As upcoming objective the triple labelled [ $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ] protein should be expressed and purified in good yields and the experiments set up properly.

### 5.4 Methods

**5.4.1 DFT calculations.** All DFT calculations of indirect nuclear spin-spin (J) coupling constants were carried out using the Gaussian 03 suite of programs.<sup>[32]</sup> Calculations of the intermolecular J couplings were performed using the hybrid B3LYP functional and the 6-311++G\*\* basis set on all atoms. In order to take into account the role of solvent due to the hydrophobic “solvent cage” effect, we used the polarizable continuum model (PCM) representing water.<sup>[33]</sup> The NMR isotropic shielding constants were calculated using the standard Gauge Independent Atomic Orbital (GIAO) approach.

## 5.4.2 Synthesis.

### 5.4.3 Acetyl 2,3,4,6-tetra-O-acetyl- $\beta$ -D-[1-<sup>2</sup>H]glucopyranoside, compound 2.



D-[1-<sup>2</sup>H]glucose (2.77 mmol) and anhydrous sodium acetate (4.88 mmol, 1.76 eq) were dissolved in acetic anhydride, 2.47 mL (9.50 eq.). The reaction mixture was attached to a reflux condenser and heated at 90 °C for 2 hours. The reaction was cooled to room temperature and neutralized with NaHCO<sub>3</sub> saturated solution. The aqueous phase was extracted with ethyl acetate and the organic extracts were washed with fresh water. Drying (MgSO<sub>4</sub>) and evaporation gave a white solid, which was recrystallized from hot methanol (2mL), giving 1.04 mmol, 38%, of pure  $\beta$  anomer of the per-O-acetylated glucoside.

**Rf:** 0.68 (Hexane / EtOAc, 1:1)

$[\alpha]_D^{20} = +6.6$  (c = 1.0, CHCl<sub>3</sub>)

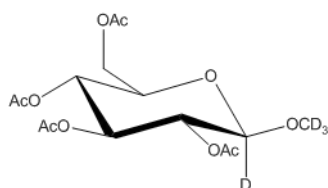
**<sup>1</sup>H NMR** (400 MHz, CDCl<sub>3</sub>)  $\delta$  5.18 (t, J = 9.7 Hz, 1H, H-3), 5.07 (d, J = 9.7 Hz, 1H, H-2), 5.06 (t, J = 9.7 Hz, 1H, H-4), 4.22 (dd, J = 4.8, 12.5 Hz, 1H, H-6R), 4.04 (dd, J = 2.4, 12.5 Hz, 1H, H-6S), 3.77 (m, J = 9.7, 4.8, 2.4 Hz, 1H, H-5), 2.05 (s, 3H, H-OAc), 2.02 (s, 3H, H-OAc), 1.97 (s, 6H, H-OAc), 1.95 (s, 3H, H-OAc).

**<sup>13</sup>C NMR** (101 MHz, CDCl<sub>3</sub>)  $\delta$  91.3 (t, J<sub>CD</sub> = 25.2 Hz, 1C, C-1), 72.8 (C-3), 72.7 (C-5), 70.2 (C-2), 67.7 (C-4), 61.5 (C-6), 20.6-20.9 (5C, C-OAc).

**MS** m/z (%): calcd for C<sub>16</sub>H<sub>21</sub>DO<sub>11</sub>Na [M+Na]<sup>+</sup> 414.1123, found 414.076 [M+Na]<sup>+</sup> (90); calcd for C<sub>16</sub>H<sub>22</sub>O<sub>11</sub>Na [M+Na]<sup>+</sup> 413.1100, found 413.066 [M+Na]<sup>+</sup> (100).



#### 5.4.4 [1,1,1-<sup>2</sup>H]Methyl 2,3,4,6-tetra-O-acetyl-β-D-[1-<sup>2</sup>H]glucopyranoside, compound 3



The peracetylated product, acetyl 2,3,4,6-tetra-O-acetyl-β-D-[1-<sup>2</sup>H]glucopyranoside (0.77 mmol) and molecular sieves (300mg) were suspended in anhydrous dichloromethane (6mL) at room temperature for 1-2 hours. 1.1 eq. of deuterated methanol, previously dissolved in 500 μl of dichloromethane, and 1.2 eq. of tin(IV) chloride were added. The reaction was immediately warmed at 40 °C and followed via NMR. The starting material was completely wearied out after 3h. The reaction was cooled to room temperature and neutralized with a NaHCO<sub>3</sub> saturated solution. The aqueous phase was extracted with dichloromethane and the organic layers were washed with fresh water. Drying (MgSO<sub>4</sub>) and evaporation gave a yellow oil that was triturated with ether to give a pure crystalline product. 0.56 mmol, 73%, of the β anomer of per-O-acetylated methyl glucoside were obtained.

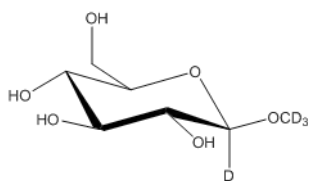
$$[\alpha]_D^{20} = -1.4 \text{ (c = 1.0, CHCl}_3\text{)}$$

<sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>) δ 5.21 (t, J = 9.7 Hz, 1H, H-3), 5.10 (t, J = 9.7 Hz, 1H, H-4), 4.99 (d, J = 9.7 Hz, 1H, H-2), 4.28 (dd, J = 4.8, 12.5 Hz, 1H, H-6R), 4.15 (dd, J = 2.4, 12.5 Hz, 1H, H-6S), 3.70 (m, J = 9.7, 4.8, 2.4 Hz, 1H, H-5), 2.09 (s, 3H, H-OAc), 2.05 (s, 3H, H-OAc), 2.03 (s, 3H, H-OAc), 2.00 (s, 3H, H-OAc).

<sup>13</sup>C NMR (101 MHz, CDCl<sub>3</sub>) δ 102.0 (C-1), 72.8 (C-3), 72.7 (C-5), 71.1 (C-2), 68.4 (C-4), 61.8 (C-6), 20.6-20.9 (4C, C-OAc).

MS m/z (%): calcd for C<sub>15</sub>H<sub>18</sub>D<sub>4</sub>O<sub>10</sub>Na [M+Na]<sup>+</sup> 389.1362, found 389.079 [M+Na]<sup>+</sup> (100).

#### 5.4.5 [1,1,1-<sup>2</sup>H]Methyl-β-D-[1-<sup>2</sup>H]glucopyranoside, compound 4.



The product from the glycosylation reaction (0.409 mmol) was deacetylated by treatment with 2 mL of methanol:triethylamine:water (2:1:1) at room temperature for 24 hours. The deprotection reaction was stopped by filtration on Amberlite IR120 hydrogen form. The methanol was evaporated and the product in water was freeze-dried. The product still showed a yellow color, because of the presence of triethylamine. So, the deacetylated glucoside was dissolved in water and flowed through a C18 chromatography column. The residual salts and the triethylamine were eluted in water while the monosaccharide was eluted in water/methanol (1:1). Again, the methanol was evaporated and the product in water was freeze-dried. 0.360 mmol, 88% of pure β anomer of [1,1,1-<sup>2</sup>H]methyl-β-D-[1-<sup>2</sup>H]glucopyranoside were obtained.

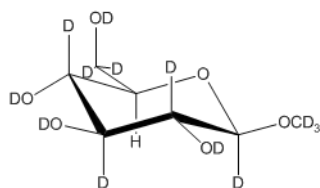
$$[\alpha]^{20}_{\text{D}} = -25.0 \text{ (c = 0.1, H}_2\text{O)}$$

<sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O) δ 3.84 (d, J = 2.4, 12.5 Hz, 1H, H-6S), 3.63 (d, J = 4.8, 12.5 Hz, 1H, H-6R), 3.40 (t, J = 9.7 Hz, 1H, H-3), 3.37 (m, J = 9.7, 4.8, 2.4 Hz, 1H, H-5), 3.29 (t, J = 9.7 Hz, 1H, H-4), 3.17 (d, J = 9.7 Hz, 1H, H-2).

<sup>13</sup>C NMR (101 MHz, D<sub>2</sub>O) δ 102.7 (C-1), 75.9 (C-5), 75.7 (C-3), 72.6 (C-2), 69.3 (C-4), 60.2 (C-6), 56.3 (C-Me).

MS m/z (%): calcd for C<sub>7</sub>H<sub>10</sub>D<sub>4</sub>O<sub>6</sub>Na [M+Na]<sup>+</sup> 221.0939, found 221.00966 [M+Na]<sup>+</sup> (100).

#### 5.4.6 Selective deuterium-labelling, compound 5.



A suspension of [1,1,1-<sup>2</sup>H]methyl-β-D-[1-<sup>2</sup>H]glucopyranoside (60.5 μmol) and 5% Ru/C (5 mol%) in D<sub>2</sub>O (1mL) was stirred at 80 °C in a test tube under a hydrogen atmosphere (balloon). The reaction was followed via NMR. After 72 hours, a full conversion of the partially protonated sugar into the fully deuterated β-D-methylglucoside, but C(5)-H, was reached. The mixture was cooled to room temperature and filtered by a membrane filter (Milipore, Millex-LH, 0.45 μm) to remove the 5% Ru/C catalyst. The filtrate was freeze-dried. 43.6 μmol, 72%, of the β anomer of [1,1,1-<sup>2</sup>H]methyl-β-D-[1,2,3,4,6,6-<sup>2</sup>H<sub>9</sub>]glucopyranoside were obtained.

$$[\alpha]_{\text{D}}^{20} = -29.0 \text{ (c = 0.1, H}_2\text{O)}$$

<sup>1</sup>H NMR (800 MHz, D<sub>2</sub>O) δ 3.36 (s, 1H, H-5).

<sup>2</sup>H NMR (123 MHz, H<sub>2</sub>O) δ 4.23 (m, 1D, D-1), 3.78 (m, 1D, D-6S), 3.57 (m, 1D, D-6R), 3.40 (s, 3D, D-Me), 3.34 (m, 1D, D-3), 3.25 (m, 1D, D-4), 3.12 (m, 1D, D-2).

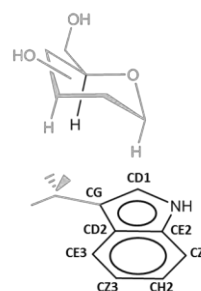
<sup>13</sup>C NMR (200 MHz, H<sub>2</sub>O) δ 102.7 (C-1), 76.0 (C-5), 75.3 (C-3), 72.6 (C-2), 69.3 (C-4), 60.2 (C-4), 56.3 (C-Me).

MS m/z (%): calcd for C<sub>7</sub>H<sub>5</sub>D<sub>9</sub>O<sub>6</sub>Na [M+Na]<sup>+</sup> 226.1253, found 226.1283 [M+Na]<sup>+</sup> (100).

## 5.5 Supporting Information

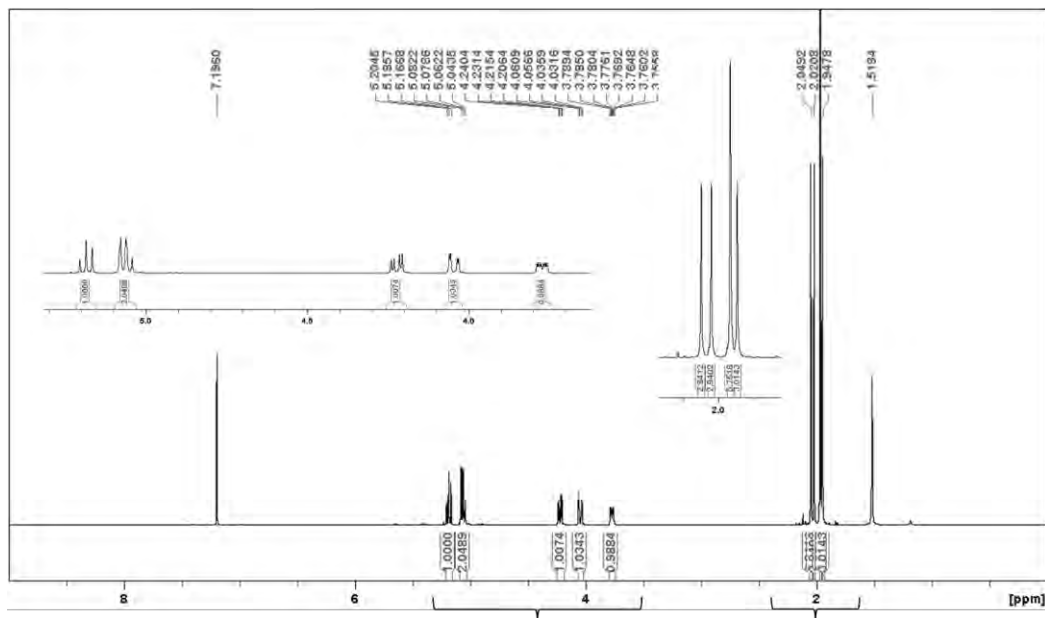
**Table 5.1 SI.** DFT analysis of JHn-Caro couplings in CH/ $\pi$  interactions in a model system consisting of tryptophan and  $\beta$ -D-glucose. The models differ in the orientation of the sugar ring with respect to the aromatic amino acid plane. A threshold value of 0.07 Hz for JHn-Caro was considered. Atom labels are shown in the figure at the right side.

Model	CH1 JH1Caro (Hz)								
	CD1	CG	CD2	CE3	CZ3	CH2	CZ2	CE2	NH
1						0.27			
2						0.26			
3					0.25	0.23			
4						0.33	0.35		
5			0.25					0.19	
Model	CH3 JH3Caro (Hz)								
	CD1	CG	CD2	CE3	CZ3	CH2	CZ2	CE2	NH
1				0.28					
2				0.24					
3			0.26	0.26					
4				0.32	0.35				
5	0.23	0.27							
Model	CH5 JH5Caro (Hz)								
	CD1	CG	CD2	CE3	CZ3	CH2	CZ2	CE2	NH
1								0.19	
2								0.26	
3							0.30	0.29	
4			0.32					0.26	
5	0.14								0.08

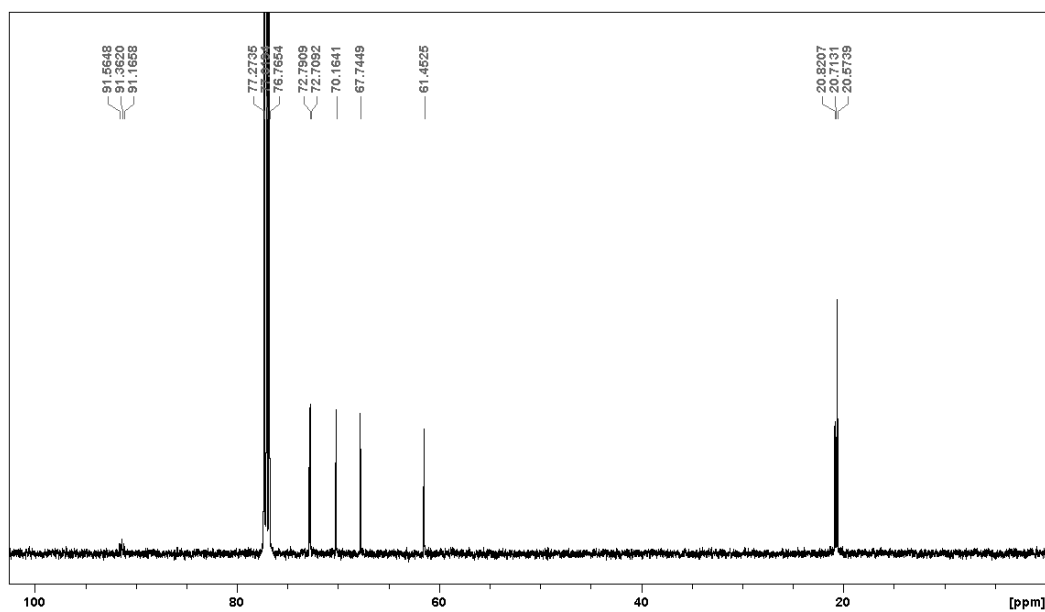


**Figure 5.2 SI.** Selected 1D and 2D NMR spectra of compound **2** recorded in a 500 MHz Bruker spectrometer (CDCl<sub>3</sub>, 298 K). a) <sup>1</sup>H-NMR spectrum; b) <sup>13</sup>C-NMR spectrum; c) <sup>1</sup>H-<sup>13</sup>C-HSQC; d) <sup>1</sup>H-<sup>1</sup>H-COSY.

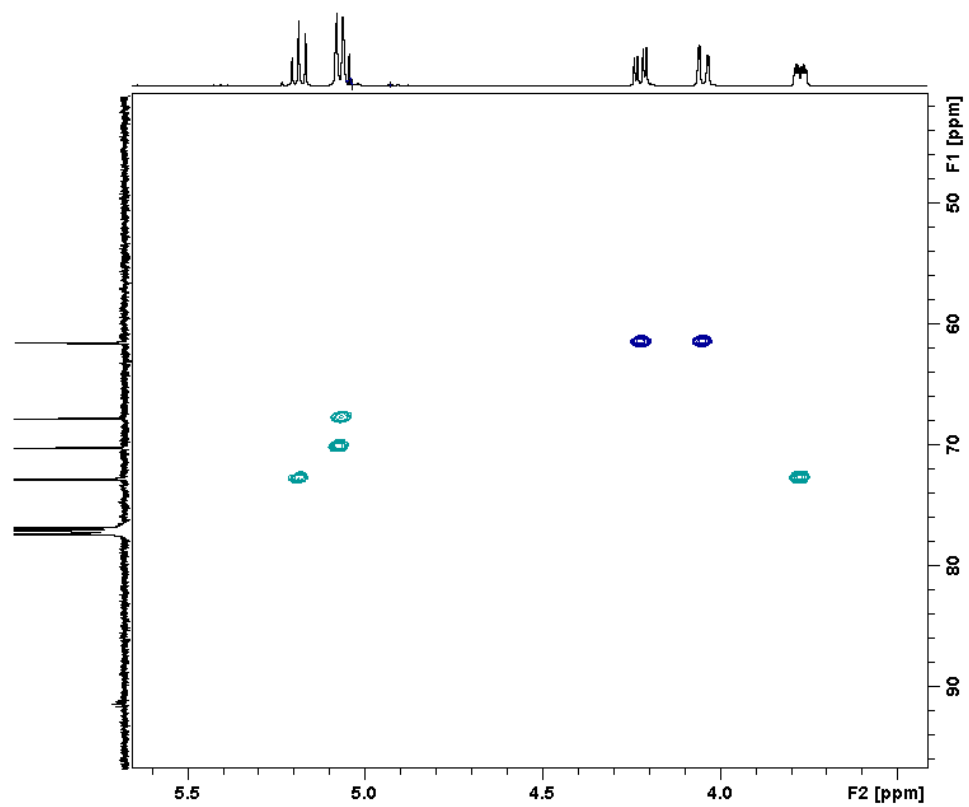
a)



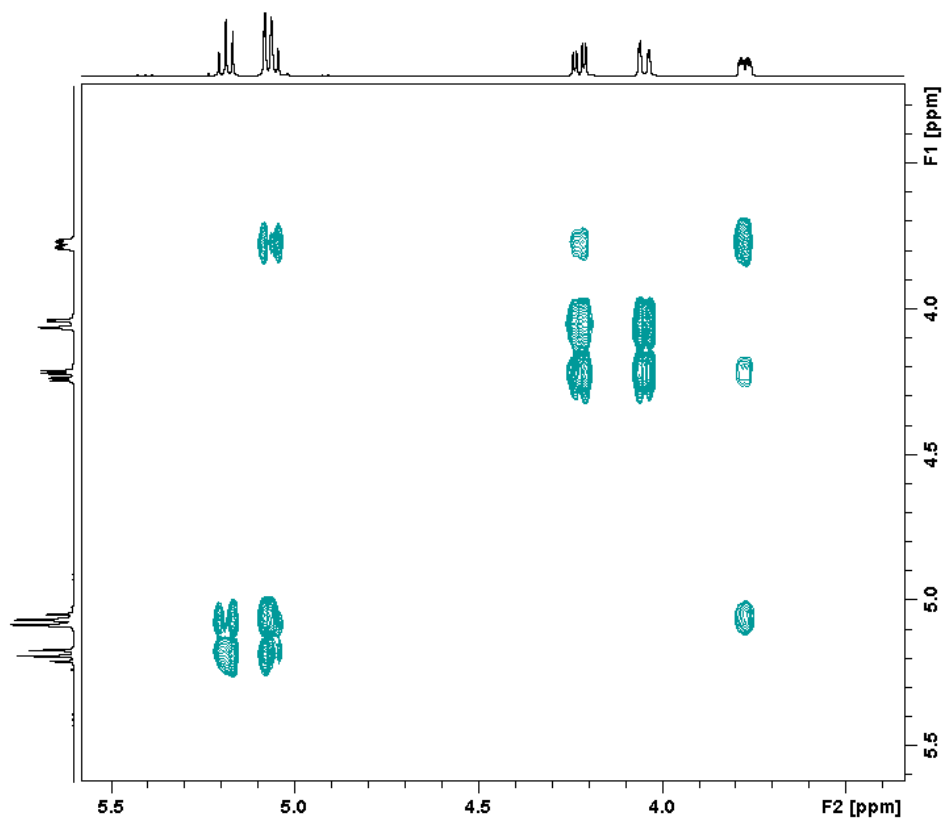
b)



c)

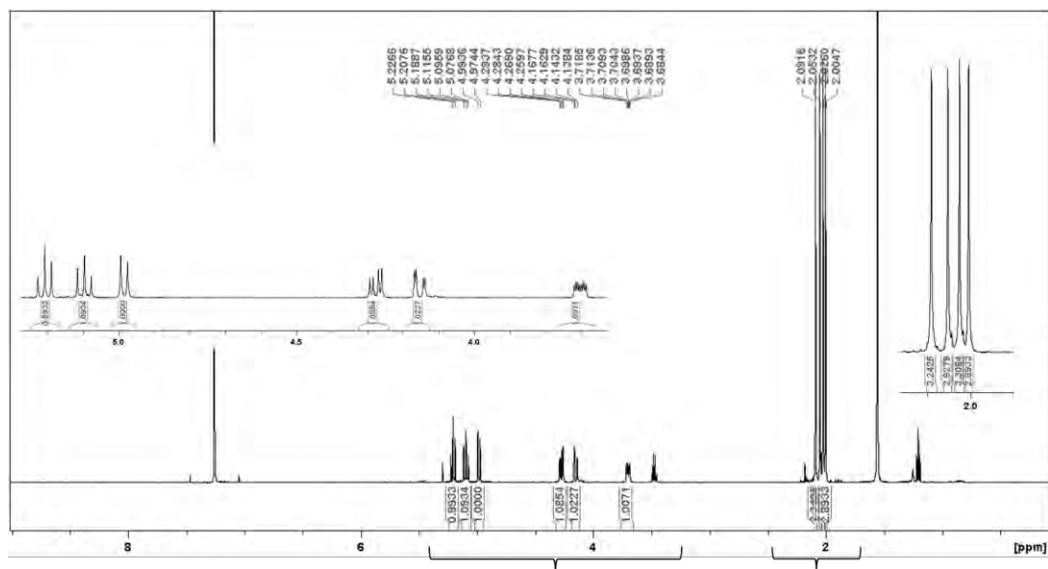


d)

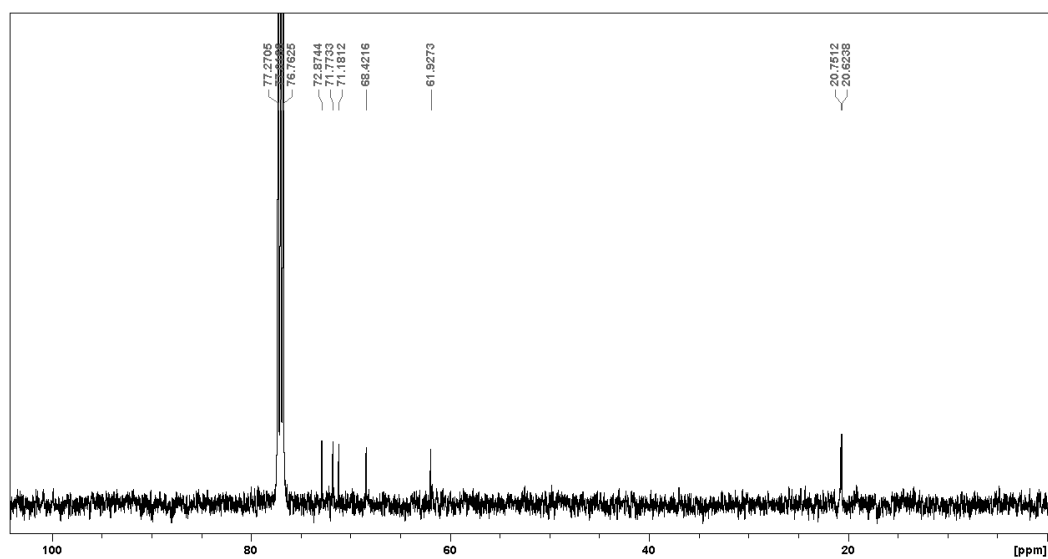


**Figure 5.3 SI.** Selected 1D and 2D NMR spectra of compound **3** recorded in a 500 MHz Bruker spectrometer (CDCl<sub>3</sub>, 298 K). a) <sup>1</sup>H-NMR spectrum; b) <sup>13</sup>C-NMR spectrum; c) <sup>1</sup>H-<sup>13</sup>C-HSQC; d) <sup>1</sup>H-<sup>1</sup>H-COSY.

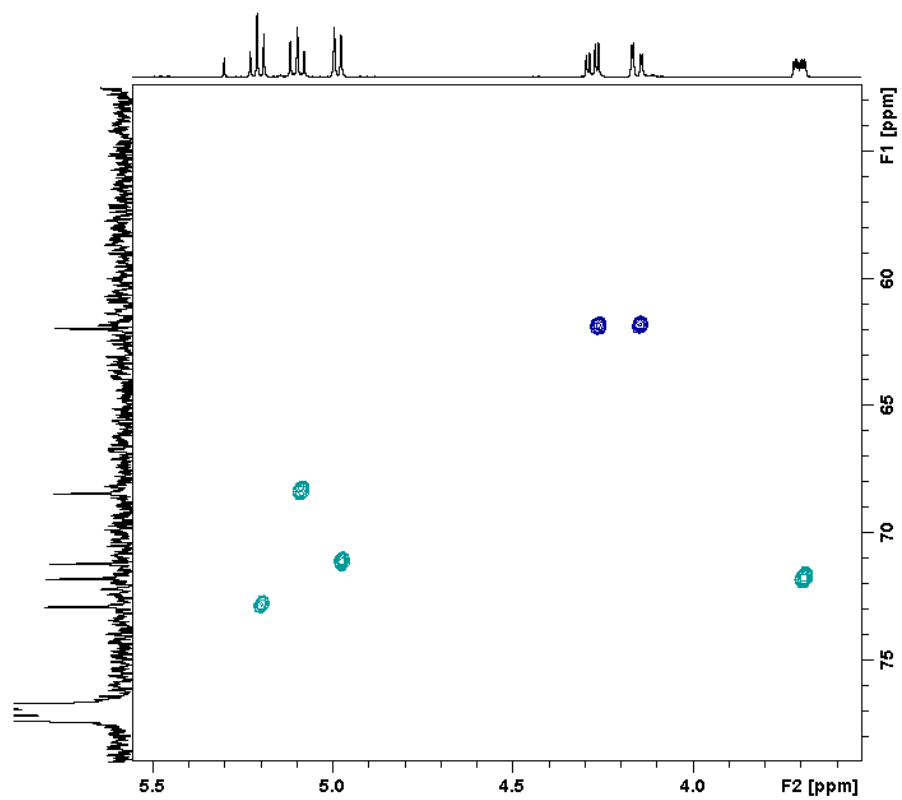
a)



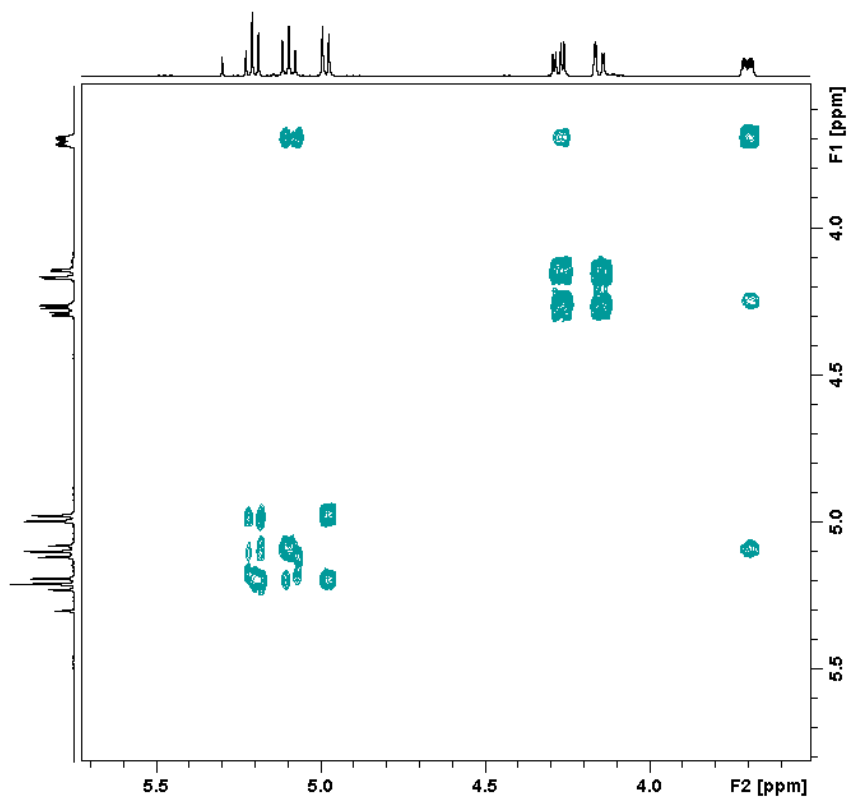
b)



c)



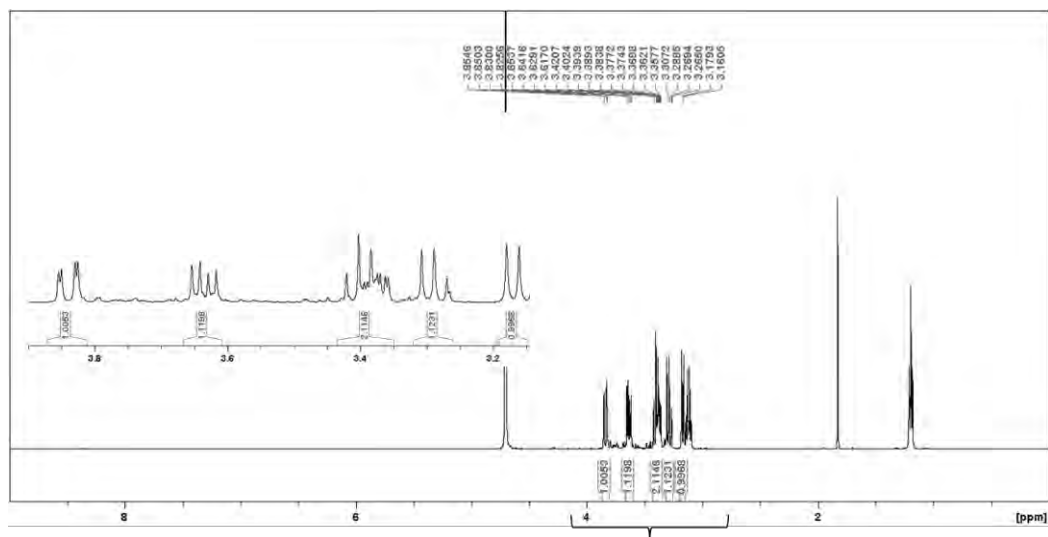
d)



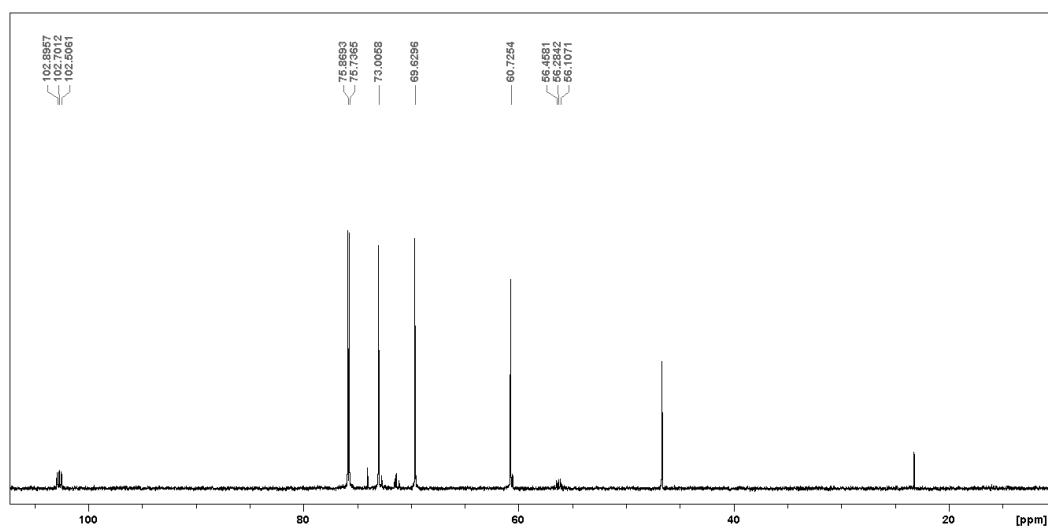


**Figure 5.4 SI.** Selected 1D and 2D NMR spectra of compound **4** recorded in a 500 MHz Bruker spectrometer (D<sub>2</sub>O, 298 K). a) <sup>1</sup>H-NMR spectrum; b) <sup>13</sup>C-NMR spectrum; c) <sup>1</sup>H-<sup>13</sup>C-HSQC; d) <sup>1</sup>H-<sup>1</sup>H-COSY.

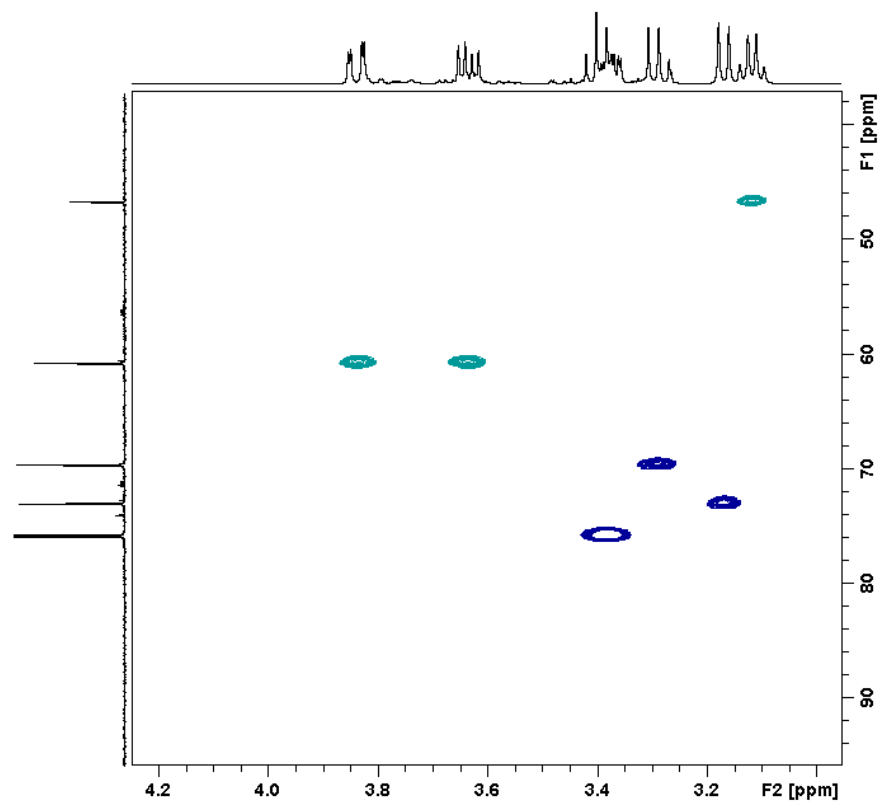
a)



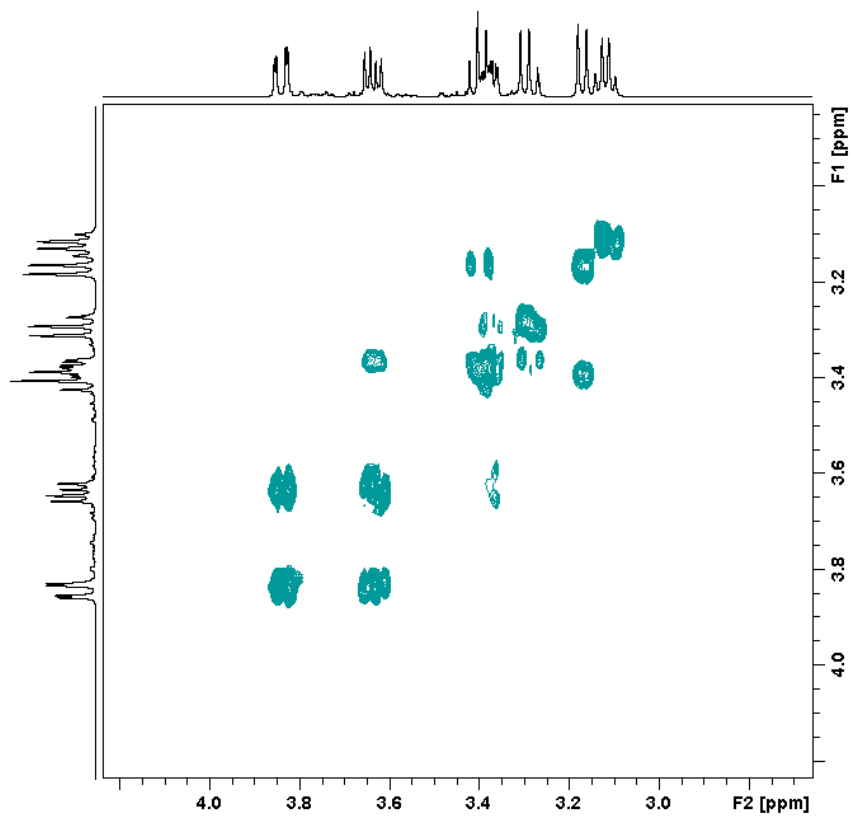
b)



c)

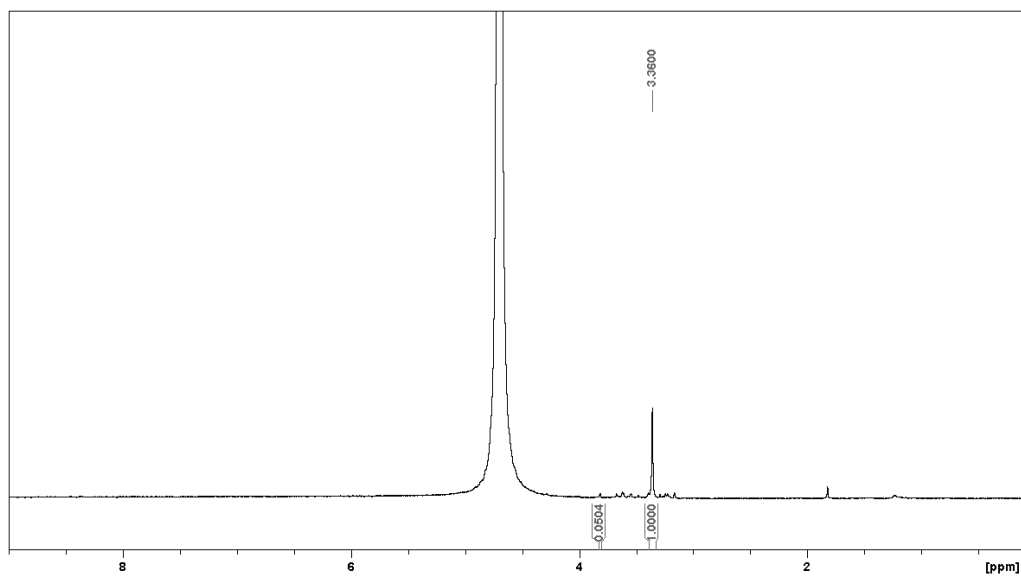


d)

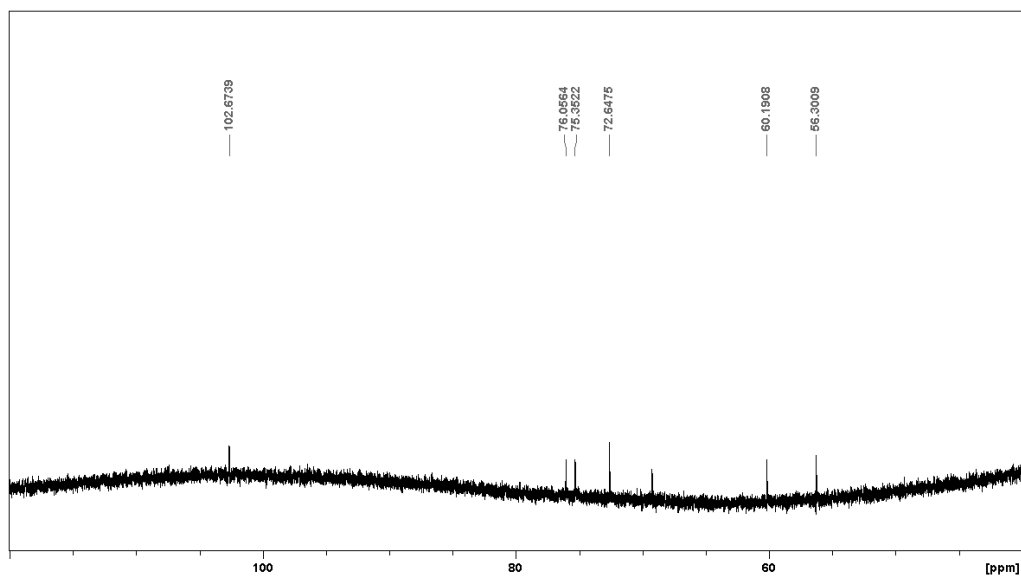


**Figure 5.5 SI.** Selected 1D NMR spectra of compound **5** recorded in 500 or 800 MHz Bruker spectrometers (D<sub>2</sub>O, 298 K). a) <sup>1</sup>H-NMR spectrum; b) <sup>13</sup>C-NMR spectrum; c) <sup>2</sup>H-NMR spectrum.

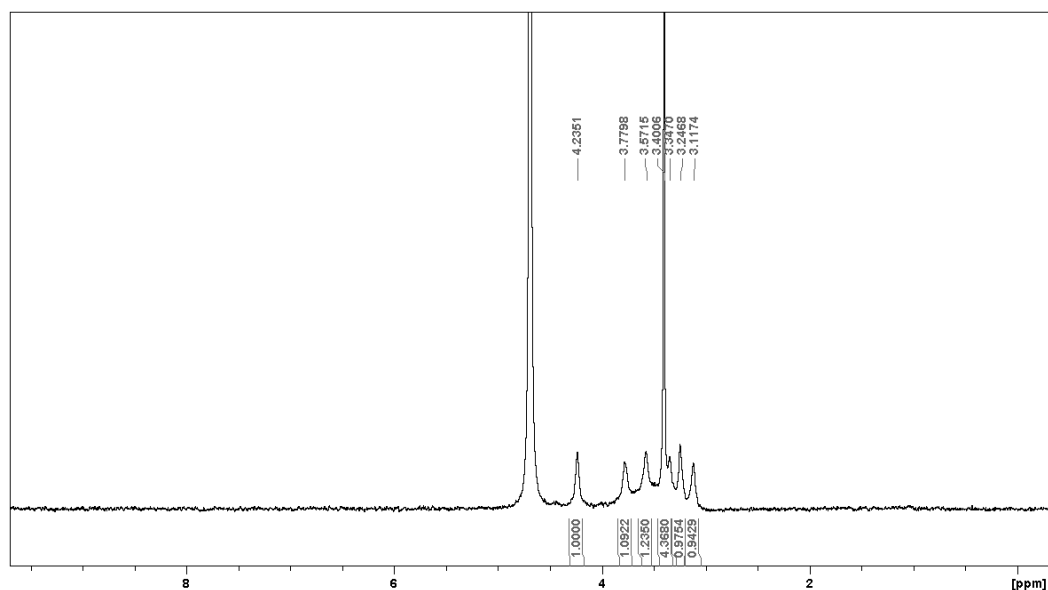
a)



b)



c)



## 5.6 References

1. National Research Council. Transforming Glycoscience. (The National Academies Press) **2012**.
2. Khamwan, K. *et al.* Pharmacokinetic modeling of [(18)F]fluorodeoxyglucose (FDG) for premature infants, and newborns through 5-year-olds. *EJNMMI Res.* **6**, 28 (2016).
3. Ernst, B. & Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nat Rev Drug Discov* **8**, 661–677 (2009).
4. Kiessling, L. L. & Splain, R. A. Chemical approaches to glycobiology. *Annu. Rev. Biochem.* **79**, 619–653 (2010).
5. Pervushin, K. *et al.* NMR scalar couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14147–14151 (1998).
6. Cordier, F. & Grzesiek, S. Direct Observation of Hydrogen Bonds in Proteins by Interresidue <sup>3</sup>HJNC' Scalar Couplings. *J. Am. Chem. Soc.* **121**, 1601–1602 (1999).
7. Asensio, J. L., Ardá, A., Cañada, F. J., Jiménez-Barbero, J. Carbohydrate-Aromatic Interactions. *Accounts of Chemical Research*, **46**, 946-954 (2013).
8. Ferrand, Y., Crump, M. P. & Davis, A. P. A Synthetic Lectin Analog for Biomimetic Disaccharide Recognition. *Science (80-. )*. **318**, 619–622 (2007).
9. Laughrey, Z. R., Kiehna, S. E., Riemen, A. J. & Waters, M. L. Carbohydrate- $\pi$  Interactions: What Are They Worth? *J. Am. Chem. Soc.* **130**, 14625–14633 (2008).

10. Ardá, A. *et al.* Molecular Recognition of Complex-Type Biantennary N-Glycans by Protein Receptors: a Three-Dimensional View on Epitope Selection by NMR. *J. Am. Chem. Soc.* **135**, 2667–2675 (2013).
11. Hudson, K. L. *et al.* Carbohydrate-Aromatic Interactions in Proteins. *J. Am. Chem. Soc.* **137**, 15152–15160 (2015).
12. Barwell, N. P. & Davis, A. P. Substituent Effects in Synthetic Lectins - Exploring the Role of CH- $\pi$  Interactions in Carbohydrate Recognition. *J. Org. Chem.* **76**, 6548–6557 (2011).
13. Tsuzuki, S., Uchimaru, T. & Mikami, M. Magnitude and Nature of Carbohydrate-Aromatic Interactions: Ab Initio Calculations of Fucose-Benzene Complex. *J. Phys. Chem. B* **113**, 5617–5621 (2009).
14. Chen, W. *et al.* Structural and Energetic Basis of Carbohydrate-Aromatic Packing Interactions in Proteins. *J. Am. Chem. Soc.* **135**, 9877–9884 (2013).
15. Lucas, R. *et al.* Effects of Sugar Functional Groups, Hydrophobicity, and Fluorination on Carbohydrate-DNA Stacking Interactions in Water. *J. Org. Chem.* **79**, 2419–2429 (2014).
16. Santana, A. G. *et al.* A Dynamic Combinatorial Approach for the Analysis of Weak Carbohydrate/Aromatic Complexes: Dissecting Facial Selectivity in CH/ $\pi$  Stacking Interactions. *J. Am. Chem. Soc.* **135**, 3347–3350 (2013).
17. Jiménez-Moreno, E. *et al.* Modulating weak interactions for molecular recognition: A dynamic combinatorial analysis for assessing the contribution of electrostatics to the stability of CH- $\pi$  bonds in water. *Angew. Chemie - Int. Ed.* **54**, 4344–4348 (2015).
18. Nishio, M. The CH/ $\pi$  hydrogen bond in chemistry. Conformation, supramolecules, optical resolution and interactions involving carbohydrates. *Phys. Chem. Chem. Phys.* **13**, 13873–13900 (2011).
19. Nishio, M., Umezawa, Y., Fantini, J., Weiss, M. S. & Chakrabarti, P. CH- $\pi$  hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys.* **16**, 12648–12683 (2014).
20. Fujii, A. *et al.* Experimental and theoretical determination of the accurate interaction energies in benzene-halomethane: the unique nature of the activated CH/ $\pi$  interaction of haloalkanes. *Phys. Chem. Chem. Phys.* **10**, 2836–2843 (2008).
21. Tsuzuki, S., Honda, K., Uchimaru, T., Mikami, M. & Tanabe, K. Origin of the Attraction and Directionality of the NH/ $\pi$  Interaction: Comparison with OH/ $\pi$  and CH/ $\pi$  Interactions. *J. Am. Chem. Soc.* **122**, 11450–11458 (2000).
22. Fernández-Alonso, M. D. C., Cañada, F. J., Jiménez-Barbero, J. & Cuevas, G. Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J. Am. Chem. Soc.* **127**, 7379–7386 (2005).
23. 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.* **16**, 1032–1041 (2007).
24. Plevin, M. J., Bryce, D. L. & Boisbouvier, J. Direct detection of CH/ $\pi$  interactions in proteins. *Nat. Chem.* **2**, 466–471 (2010).
25. Zierke, M., Smies, M., Rabbani, S., Aeschbacher, T. & Cutting, B. Stabilization of Branched Oligosaccharides : Lewis<sup>x</sup> Bene fi ts from a Nonconventional C-H $\cdots$ O Hydrogen Bond. *J. am. Chem. Soc.* **135**, 13464-13472 (2013).
26. Battistel, M. D., Azurmendi, H. F., Frank, M. & Freedberg, D. I. Uncovering Nonconventional and

- Conventional Hydrogen Bonds in Oligosaccharides through NMR Experiments and Molecular Modeling: Application to Sialyl Lewis-X. *J. Am. Chem. Soc.* **137**, 13444–13447 (2015).
27. Atreya, H. S. *Isotope labeling in Biomolecular NMR*. (Ed. Springer) **2012**.
  28. Sawama, Y. *et al.* Stereo- and regioselective direct multi-deuterium-labeling methods for sugars. *Chem. - A Eur. J.* **18**, 16436–16442 (2012).
  29. Lemiéux, R. U. & Shyluk, W. R. A New synthesis of  $\beta$ -glucopyranosides. *Can. J. Chem.*, **31**, 528-535 (1955)
  30. Lemiéux, R. U. & Huber, G. The alpha and beta 1,3,4,6-tetraacetyl-D-glucopyranoses and their chloroacetyl derivatives. *Can. J. Chem.*, **31**, 1040-1046 (1953).
  31. Konstantinovic, S. *et al.* SnCl<sub>4</sub> induced formation of C7–C16-alkyl D-glucopyranosides. *J. Serb. Chem. Soc.* **66**, 65–71 (2001).
  32. Frisch, M. J. *et al.* Gaussian03, Revision C.02. Gaussian, Inc., Wallingford CT, **2004**.
  33. Mennucci, B. *et al.* Polarizable Continuum Model (PCM) Calculations of Solvent Effects on Optical Rotations of Chiral Molecules. *J. Phys. Chem. A* **106**, 6102–6113 (2002).



## Conclusions

The dynamic behavior of different sugars and mimetics and of their receptors has been studied at atomic scale using Nuclear Magnetic Resonance spectroscopy and Computational methods.

The topic of structural flexibility in carbohydrates has been addressed at different level of structural complexity.

1.-For the analysis of the conformational equilibria in monosaccharides, a new generation of fluorine-containing glycomimetics has been presented.

The importance of introducing fluorine atoms in these glycomimetics has also been highlighted. Only the use of  $^{19}\text{F}$  NMR experiments has permitted to detect a dynamic process of paramount significance that would have been otherwise remained unobserved.

2.-Additionally, only in the presence of fluorine atoms, within a  $\text{CF}_2$  moiety that replaces the endocyclic oxygen atom, the Ido-like six-membered ring recovers its required flexibility, absent in regular  $\text{CH}_2$ -Ido-carbasugars.

The combination of NMR experiments and computational methods has permitted to show that these idose-like analogues resemble the conformational plasticity of the natural parent molecules that is required for key molecular recognition process and ultimately for biological activity.

3.-For the analysis of glycosidic linkage conformations in disaccharides, the theoretical calculations predicted that the *exo*-anomeric effect of maltose, which is almost completely abolished in carbasugar analogue, is significantly restored when a  $\text{CF}_2$  group is present at the endocyclic position.

4.-We have experimentally determined the conformational behavior of a *gem*-diF-carbasugar maltose analogue, which indeed only exists in solution in the *exo*-anomeric conformation, strikingly different from the methylene counterpart,  $\text{CH}_2$ -carbasugar, which shows a marked dynamic behaviour with the population of *exo*- and non *exo*-anomeric conformations. It has therefore been experimentally demonstrated that it is possible to restore the *exo*-anomeric effect for an acetal when replacing one of the oxygen atoms by a  $\text{CF}_2$  group. This result is of fundamental interest as it strongly suggests the importance of the stereoelectronic component for the *exo*-anomeric effect.



5.-Large scale domain rearrangements in glycan receptor have been studied theoretically and experimentally. The amplitude and time scale of GGBP inter-domain dynamics have been unveiled by NMR spectroscopy and detailed MD simulations.

GGBP protein has been modified through single point mutation and chemical conjugation with a substituent able to chelate lanthanides cations. The modified protein has been labeled with dysprosium, as paramagnetic ion, and lanthanum, as diamagnetic ion.

6.-The population contribution of the bio-functional relevant conformers that contribute to the equilibrium has been determined through the analysis of proteins chemical shift perturbation, pseudocontact chemical shift (PCSs) and residual dipolar couplings (RDCs) induced by paramagnetic metal ion.

7.-Moreover, the energy barrier for the open-to-closed transition process has been defined by non-equilibrium MD simulations.

8.-All the data point out that the population of *apo\_cl* like conformations is essential for activating the transition to *holo\_cl* form, according to a conformational selection mechanism coupled to a final rearrangement that obeys an induced-fit kinetics process.

From a general perspective, the protocol exemplified herein can be extended to the study a variety of molecular recognition processes in which significant molecular rearrangements take place, thus expanding the limits of the application of NMR methods to explore binding events.

9.-Following with intermolecular interactions in sugar-proteins complexes, specifically with CH/ $\pi$  interactions, DFT calculations have predicted that intermolecular J coupling constants indeed exist in a simple model constituted by a glucose unit and a tryptophan amino acid. The theoretical values for the direct J coupling indicate that this spectroscopic parameter could be detected with the aid of an optimized isotope-labelling scheme that minimizes the intramolecular coupling constants, which could hide the intermolecular couplings.

10.-We have designed a synthetic procedure to obtain a single proton labelled glucose through the substitution of all hydrogen atoms by deuterium, save H5 that participate in the intermolecular coupling. Thus, the fully deuterated  $\beta$ -D-methylglucoside, except for C(5)-H, has been synthesized.



## Conclusiones

El comportamiento dinámico de distintos azúcares, glicomiméticos y sus receptores se ha estudiado a escala atómica usando técnicas de espectroscopia de Resonancia Magnética Nuclear y métodos computacionales. El tema de la flexibilidad ha sido abordado en diferentes niveles de complejidad estructural.

1.- Para el estudio de equilibrio conformacional en monosacáridos, se ha presentado una nueva generación de glicomiméticos fluorados. Se ha demostrado la importancia de introducir átomos de flúor en estos glicomiméticos. Solamente mediante el uso de experimentos de RMN de  $^{19}\text{F}$  se ha podido detectar y describir un proceso dinámico de suma importancia que, de otra manera, no se hubiera podido observar.

2.-Además, la presencia de átomos de flúor dentro de un sistema  $\text{CF}_2$  en lugar del átomo de oxígeno endocíclico, ha facilitado que el anillo de idosa recupere su propia flexibilidad, ausente en el análogo  $\text{CH}_2$ -Ido-carbaazúcar.

La combinación de experimentos de RMN y métodos computacionales ha permitido demostrar que estos análogos de idosa mantienen la plasticidad conformacional de los análogos naturales, que es clave para el proceso de reconocimiento molecular y en última instancia para su actividad biológica.

3.-Para el análisis conformacional del enlace glicosídico en disacáridos, los cálculos teóricos predicen que el efecto *exo*-anomérico en maltosa, que no existe en los análogos de tipo carbaazúcar, se restaura significativamente cuando un grupo  $\text{CF}_2$  está presente en la posición endocíclica.

4.-Se ha determinado experimentalmente el comportamiento conformacional de un *gem*-diF-carbaazúcar análogo de maltosa. Este muestra una conformación exclusiva de tipo *exo*-anomérica. Esta observación es diferente a la que se da con el correspondiente  $\text{CH}_2$ -carbaazúcar, que muestra un comportamiento dinámico, con la presencia de conformaciones de tipo *exo*- y no *exo*-anoméricas. Por lo tanto, se ha demostrado que es posible restablecer el efecto *exo*-anomérico de un acetal cuando se sustituye uno de los dos átomos de oxígeno por un grupo  $\text{CF}_2$ . Este resultado es de interés fundamental dado que demuestra la importancia de la componente estereoelectrónica del efecto *exo*-anomérico.

5.-Se han estudiado, teórica y experimentalmente, los movimientos intramoleculares interdominios en receptores de glicanos. En concreto, se ha determinado la amplitud y la escala de tiempo de los movimientos inter-dominio del receptor de glucosa y galactosa GGBP mediante espectroscopia de RMN y simulaciones de dinámica molecular. El receptor GGBP se ha modificado mediante técnicas de mutación dirigida y marcaje químico covalente con un sustituyente capaz de enlazar cationes lantánidos. La proteína así modificada ha sido etiquetada con disprosio, catión paramagnético, y con lantano, catión diamagnético.

6.-La distribución de poblaciones de los diferentes conformeros bio-funcionales ha sido determinada por medidas de desplazamientos químicos de pseudocontacto ("pseudocontact chemical shifts", PCSs) y de constantes residuales dipolares ("residual dipolar constants", RDCs).

7.-Se ha determinado la barrera energética de la transición entre la estructura abierta y cerrada mediante simulaciones de dinámica molecular de no equilibrio.

8.-Todos los datos obtenidos han permitido demostrar que la población de conformaciones tipo *apo\_cl* es esencial por la activación de la transición hacia la forma *holo\_cl*, de acuerdo con un mecanismo de selección conformacional seguido de una reorganización entre los dos dominios, que obedece a un proceso de ajuste inducido. Desde una perspectiva general, el protocolo usado en este trabajo puede extenderse al estudio de diferentes procesos de reconocimiento molecular, ampliando de esta manera los límites de los métodos de RMN para explorar acontecimientos de unión.

9.-Siguiendo con las interacciones intermoleculares en complejos entre proteínas y azúcares y, en concreto, interacciones CH- $\pi$ , el uso de cálculos DFT ha permitido estimar los valores de constantes de acoplamiento (J) intermoleculares en un modelo simple, constituido por una unidad de glucosa y un aminoácido de triptófano.

10.-Los valores teóricos calculados para la constante de acoplamiento (J) indican que, para poder determinar este parámetro, es necesario diseñar un procedimiento donde se minimicen otras posibles J presentes en el sistema. Se ha diseñado una estrategia basada en un etiquetado isotópico optimizado mediante la sustitución de todos los hidrógenos por deuterio excepto aquel participante en el acoplamiento intermolecular objeto de estudio. Así, se ha diseñado, y usado con éxito, un procedimiento de síntesis para obtener el  $\beta$ -metil glucopiranosido

selectivamente protonado en posición 5, sustituyendo el resto de hidrógenos de la molécula por átomos de deuterio.

