



UNIVERSIDAD AUTÓNOMA DE
MADRID
Facultad de Ciencias



Molecular Recognition of Glycomimetics by Proteins
A 3D view by using NMR

MEMORIA

para optar al grado de

Doctor en Ciencias Químicas

presenta

Silvia Mari

Centro de Investigaciones Biológicas
(C.S.I.C.)

Madrid 2006

Index

| | |
|---|------|
| <u>Introduction and General Background</u> | p.1 |
| <u>Carbohydrates as cell recognition points</u> | p.3 |
| Carbohydrate sequences | p.3 |
| Principles of Protein-Carbohydrate Recognition. Glycomimetics | p.4 |
| Plant Lectins: General Overview | p.5 |
| Animal Lectins: General Overview | p.7 |
| Protein-Carbohydrate Interactions: CH- π , polar interactions, hydrogen bonding, and the role of water. | p.9 |
| <i>CH-π Interactions</i> | p.11 |
| <i>Polar Interactions: Recognition of Charged Groups on Sugars and the Role of Divalent Cations</i> | p.12 |
| <i>Hydrogen Bonding and Water-Mediated Hydrogen Bonds</i> | p.14 |
| References | p.16 |
| <u>The three dimensional structures of oligosaccharides and glycomimetics.</u> | p.22 |
| <u>Factors and facts</u> | |
| The shape of the six-membered rings | p.23 |
| Glycosidic and aglyconic linkages | p.29 |
| The anomeric effects | p.30 |
| Symbol Representation for Glycans | p.31 |
| <i>Symbol Representations</i> | p.32 |
| References | p.35 |
| <u>Investigation of sugar conformations: the Free State</u> | p.37 |
| X-Ray Crystallography | p.37 |
| NMR Methods for Structure Determination | p.38 |
| <i>Assignment Methods</i> | p.38 |
| <i>Scalar Coupling Constants</i> | p.42 |
| <i>The Nuclear Overhauser Effect</i> | p.43 |
| <i>Residual Dipolar Couplings: Theory and General Concepts</i> | p.46 |

| | |
|---|------|
| NMR and molecular modeling | p.52 |
| Molecular Modeling and the Nuclear Overhauser Enhancement | p.55 |
| <i>Isolated Spin Pair Approximation (ISPA)</i> | p.56 |
| <i>Direct Comparison of cross and diagonal peaks</i> | p.57 |
| <i>Reference structure based iterative methods</i> | p.57 |
| Simulating the nOe: the NOEPROM Software | p.57 |
| References | p.59 |

The bound state

| | |
|--|------|
| The Fast Exchange Approximation | p.65 |
| Ligand-Based versus Receptor-Based Screening | p.66 |
| Ligand-Based NMR Methods | p.67 |
| Transferred NOE | p.67 |
| Saturation Transfer Difference | p.69 |
| Applications of RDCs to the study of multiple fragments or complexes | p.72 |
| References | p.75 |

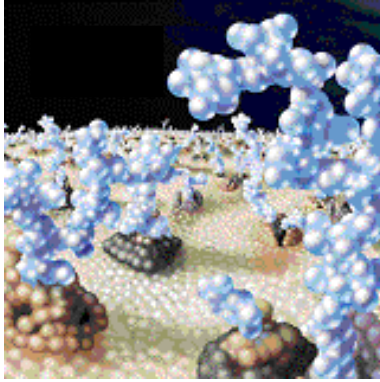
Results and Discussion

Objective

The Cholera Toxin B₅ pentamer and its ligands

| | |
|---|-------|
| The membrane receptor: GM1 ganglioside | p.81 |
| The CT:GM1 complex | p.82 |
| Development of CT artificial ligands | p.82 |
| References | p.83 |
| 1st Paper: Mimics of ganglioside GM1 as cholera toxin ligands: replacement of the GalNAc residue | p.84 |
| 2nd Paper: Intramolecular Carbohydrate–Aromatic Interactions and Intermolecular van der Waals Interactions Enhance the Molecular Recognition Ability of GM1 Glycomimetics for Cholera Toxin | p.92 |
| 3rd Paper: Synthesis and conformational analysis of galactose-derived bicyclic scaffolds | p.104 |

| | |
|---|-------|
| <u>The DC-SIGN receptor</u> | p.118 |
| References | p.120 |
| 4th Paper: 1D-STD NMR Experiments On Living Cells. The DC-SIGN/oligomannose interaction | p.122 |
| <u>Glycosidases and Glycomimetics</u> | p.125 |
| Glycosidases: mechanism of reaction | p.125 |
| Glycosidase inhibitors | p.126 |
| References | p.128 |
| 5th Paper: The conformational behavior of novel glycosidase inhibitors with substituted azepan structures. A NMR and Modeling Study | p.130 |
| 6th Paper: Synthesis, Conformational Studies, and Mannosidase stability of a Mimic of 1,2-Mannobioside | p.141 |
| <u>Conformer selection of phenyl-lactic GM1 glycomimetic by three different lectins</u> | p.148 |
| NMR studies of a phenyl-lactic mimic with VAA and CG14 | p.148 |
| <u>Contributions to Symposia</u> | p.154 |
| 1st poster | p.155 |
| 2nd poster | p.157 |
| 3rd poster | p.159 |
| <u>Acknowledgments</u> | p.161 |



Introduction and General Background

In this remarkable age of genomics, proteomics, and functional proteomics, I am often asked by my colleagues why glycobiology has apparently lagged so far behind the other fields. The simple answer is that glycoconjugates are much more complex, variegated, and difficult to study than proteins or nucleic acids.

Saul Roseman in "Reflections on Glycobiology"
J. Biol. Chem. (2001) **276**, 41527-41542

Carbohydrates as cell recognition points

Carbohydrates in nature form an important family of biomolecules, as simple or complex entities, either alone or covalently linked to proteins or lipids. Years ago, sugar molecules were nearly exclusively assigned as building blocks of protective cell wall constituents (for example, cellulose and chitin) or as biochemical fuels in energy metabolism (*born to be cleaved*, Manuel Rico, personal communication). Only recently it has been demonstrated that they also play numerous and different biological functions (Dwek 1996; Gabius et al. 2004; Science 2001; Solis et al. 2001; Varki 1993). Glycobiology has thus emerged as a new and challenging research area at the interface of biology and chemistry. Nowadays it is clear that sugars are the key point in molecular recognition events, with lectins, toxins, enzymes and antibodies, (Akama et al. 2002; Drickamer et al. 1993; Guo et al. 2004; Lis et al. 1998; Rudd et al. 1994; Weis et al. 1991) and that play a key role in important biological phenomena involving cell-cell interaction such as fertilization, bacterial infection, inflammatory processes, cell growth, etc... The glycoconjugates, *i.e.*, glycolipids, glycoproteins, and glycosaminoglycans present at the cell surface, display diverse glycosylation patterns between species, which appear to be driven by evolutionary selection pressures (Gagneux et al. 1999), but which also can occur between cell types in the same organism. Modifications of cell glycosylation also occur during cell activation, inflammation, and cancer. Elucidation of the three-dimensional structures and the dynamical properties of oligosaccharides is a prerequisite for a better understanding of the recognition processes and for the rational design of carbohydrate-derived drugs.

The last two decades have witnessed tremendous advances in the elucidation and understanding of the conformations and dynamics of oligosaccharides thanks to the technical developments in nuclear magnetic resonance (NMR), X-ray diffraction, and computer modeling methods.

Carbohydrate sequences

To answer to the question of why carbohydrates have remained nearly unnoticed for so long it should be underlined that glycoconjugates are much more complex, diverse, and difficult to study than proteins or nucleic acids (Roseman 2001).

Carbohydrates have potential information content that is several orders of magnitude higher than any other biological macromolecule. Considering nucleotides and amino acids as prime biochemical hardware in transferring information, (Laine 1994) comparative calculations can be performed. At the level of trimer, 64 sequence permutations (4^3) for nucleotides using the four types of purine and pyrimidine bases may be formed, while 8000 isomers (20^3) for peptides and the standard collection of 20 amino acids. To appraise the respective abilities of sugars, similar calculation can be performed leading to a number of nine million carbohydrate isomers (Laine 1994).

Introduction of one substituent, e.g. a sulfate group, to a free hydroxyl group yields a 10-fold increase of this number. Because the presence of more than one linkage point per building block makes it possible to devise branched oligomers, this new feature should be reflected in the pool size of

oligomers larger than trimers. For instance, the number of hexamers for peptides with 64 millions looks impressive compared to 4096 hexanucleodides. However, it is surpassed by the isomer quantity of hexasaccharides, i.e. 1.44×10^{15} . Thus, even a small number of monosaccharide units can provide a large number of different oligosaccharides.

This structural diversity allows carbohydrate to encode information for specific molecular recognition and to serve as determinant of protein folding, stability and pharmacokinetics. And because glycosylation is one of the most ubiquitous forms of post-translational modification, the unexpected small number genes identified in the analysis of the human genome sequence provide even more impetus for understanding the biological role of carbohydrates.

Principles of Protein-Carbohydrate Recognition: toward the concept of glycomimetics

We have long lived with the paradigm of the “lock and-key” model of protein-carbohydrate interactions. In surveying the results of recent molecular modelling calculations, there are hints that we should at least consider that carbohydrates occupy diverse glycosidic linkage conformations. An alternative metaphor to the “lock-and-key” model is tentatively suggested, labeled “the bunch of keys”. The carbohydrate moves in solution through a bunch of shapes each of which may be selected by a receptor. The receptor may choose a large key or a small key, and so the consideration of densities of conformations, which include minor conformations, is a promising and sensible method of description. The investigation of this idea surely provides an exciting journey of exploration ahead.

Barry J. Hardy in “The glycosidic linkage flexibility and time-scale similarity hypotheses”, J. Mol. Struct. (1997) 395-396, 187-200.

As implied by referring to the existence of a glycode in the previous section, information stored, as sequence and shape will have to be translated and transmitted it into the intended responses. Thus, in addition to serving physicochemical roles to control folding, oligomerization, and access of proteolytic enzymes (Drickamer et al. 1998; Gabius et al. 1997; Gagneux et al. 1999; Reuter et al. 1999; Varki 1996), oligosaccharides in glycan chains can be compared to the postal code in an address for conveying distinct messages read by the appropriate receptors. These carbohydrate-binding proteins are classified into enzymes responsible for assembling, modifying, and degrading sugar structures, immunoglobulins homing in on carbohydrates as antigens, and, finally, lectins. The third of these classes evidently encompasses all carbohydrate-binding proteins, which are neither antibodies nor enzymes which couple ligand recognition with catalytic activity to process the target (Barondes 1988; Gabius 1994). Lectins are found in most organisms, ranging from viruses and bacteria to plants and animals. They are readily obtainable in purified form, mostly by affinity chromatography on the immobilized ligand, and more recently also by recombinant DNA techniques. They represent a heterogeneous group of oligomeric proteins that vary widely in size, structure, molecular organization, as well as in the constitution of their combining sites. Nonetheless, many of them belong to distinct protein families with similar sequences and structural features. In fact, sequence similarity with known lectins provides a novel guideline for the detection and identification of new ones.

Although lectins were first described at the turn to 20th century, their study started to gain momentum only in the 1960s (Liener et al. 1986; Sharon et al. 1987; Sharon et al. 1989). Lectins became invaluable tools for the structural and functional investigation of complex carbohydrates, especially glycoproteins, and for the examination of changes that occur on cell surfaces during physiological and pathological processes, from cell differentiation to cancer (Goldstein et al. 1997; Lis et al. 1986). At present, they are the focus of intense attention because of the realization that they act as recognition determinants in diverse biological processes (Sharon et al. 1993; Sharon et al. 1989). These include clearance of glycoproteins from the circulatory system, control of intracellular traffic of glycoproteins, adhesion of infectious agents to host cells, recruitment of leukocytes to inflammatory sites, as well as cell interactions in the immune system, in malignancy and metastasis. Investigation of lectins and their role in cell recognition, as well as the application of these proteins for the study of carbohydrates in solution and on cell surfaces, are making marked contributions to the advancement of glycobiology (Dwek 1996). Developments in lectin research and glycobiology, are now moving ahead hand in hand and mainly this was made possible by the refinement of old techniques and development of new ones. In particular, high resolution X-ray crystallography of lectins in complex with their ligands allowed the identification of the chemical groups on the protein and on the carbohydrate that interact with each other and of the types of bond formed. Further information on the contribution of individual amino acids to the activity of a lectin has been obtained by site-directed mutagenesis experiments and also by molecular modeling.

Plant Lectins: General Overview

The members of the diverse group of plant lectins that are studied and used most frequently are listed in Figure 1. The leading position is held by concanavalin A, the "classical" Man/Glc-binding lectin from *Jack beans*. Its properties explain why this lectin has attained its status as a reliable and popular workhorse, especially for carbohydrate chemists looking for a lectin to use in an attempt to prove the ligand properties of a sugar compound attached to a new synthetic scaffold. The other lectins listed in Figure 1 are capable of following the role model concanavalin A, although they are less prominently used in research. These compounds form a panel of probes for isolation and structural characterization of glycoconjugates (glycoproteins, glycolipids, or polysaccharides), as well as use in various assays in cell biology, histochemistry, and the medical sciences (Cummings 1997; Damjanov 1987; Danguy et al. 1994; Gabius 2000; Lis et al. 1986; Osawa et al. 1987; Peumans et al. 1998; Rudiger et al. 2001).

Lectins frequently display mitogenic activity, and the fact that plant lectins can affect lymphocyte activity and proliferation has led to suggestions that the laboratory tools could be introduced as immunomodulatory therapeutic agents in clinical applications. The example of the galactoside-specific mistletoe lectin (VAA, formerly ML-1), a constituent of proprietary extracts used in Austria, Germany, and Switzerland, shows that immune functions such as secretion of proinflammatory cytokines or priming of granulocytes/ activity of NK cells can indeed be stimulated at nontoxic doses of lectin (VAA concentration needed to elicit in vivo effects: 1 ± 2 ng kg⁻¹ body weight, given subcutaneously) (Hajto et al. 1990; Hajto et al. 1989). However, this immunomodulatory capacity is unlikely to have a

clinical perspective, because it is becoming evident that these immune factors can also trigger growth responses in malignant cells (Chouaib et al. 1997; Coussens et al. 2002; Gabius 2001; Sogn 1998; Yu et al. 2003).

| Plant species and name/ abbreviation of lectin | Family | Mono- or disaccharide specificity | Comments |
|---|----------------|--------------------------------------|---|
| <i>Canavalia ensiformis</i> (concanavalin A, ConA) | Leguminosae | Man/Glc | cheapest and most popular lectin; first lectin isolated by crystallization and demonstrated to interact with carbohydrate (see text and Table 2 for details) |
| <i>Ricinus communis</i> (ricin) | Euphorbiaceae | Gal | ribosome-inactivating protein, type II (RIP II), used for generating immunotoxins; biohazard |
| <i>Triticum vulgare</i> (WGA) | Gramineae | (GlcNAc) ₁₋₃ , Neu5Ac | potential function in plant defence mechanisms |
| <i>Phaseolus vulgaris</i> (PHA) | Leguminosae | no simple carbohydrate known | isolectin L ₄ is a strong mitogen for T-lymphocytes, isolectin E ₄ is a strong erythrocyte agglutinin (see Figure 1 B for chromatographic isolectin separation); distinguish between bisected and nonbisected <i>N</i> -glycans; cause of severe gastrointestinal irritation when ingested in insufficiently cooked beans |
| <i>Glycine max</i> (SBA) | Leguminosae | GalNAc/Gal | cell sorting, bone marrow purging |
| <i>Pisum sativum</i> (PSA) | Leguminosae | Man/Glc | binding of <i>N</i> -glycans enhanced by core fucosylation |
| <i>Viscum album</i> (VAA, viscumin) | Viscaceae | Gal | RIP II used for generating immunotoxins, constituent of proprietary mistletoe extracts (immunomodulatory and growth stimulatory for tumor cells in vitro and in vivo at low doses; see text for details) |
| <i>Arachis hypogaea</i> (PNA) | Leguminosae | Gal, Galβ3GalNAcα (TF-antigen) | very popular in histochemistry; separates immature from mature thymocytes |
| <i>Lens culinaris</i> (LCA) | Leguminosae | Man/Glc | binding of <i>N</i> -glycans enhanced by core fucosylation; lymphocyte mitogen |
| <i>Dolichos biflorus</i> (DBA) | Leguminosae | GalNAcα3GalNAc, GalNAc | cell sorting, agglutinates blood group A erythrocytes |
| <i>Griffonia simplicifolia</i> (GSA-I) | Leguminosae | Gal/GalNAc | isolectin GSA-I-A ₄ agglutinates blood group A erythrocytes, isolectin GSA-I-B ₄ blood group B erythrocytes |
| <i>Griffonia simplicifolia</i> (GSA-II) | Leguminosae | (GlcNAc) _n | insecticidal activity, potential defence role |
| <i>Artocarpus integrifolia</i> (jacalin) | Moraceae | Gal (Man, TF-antigen) | used for isolation of IgA ₁ and mucins, mitogenic for CD4 ⁺ T-cells |
| <i>Solanum tuberosum</i> (STA) | Solanaceae | (GlcNAc) _n | potential function in plant defence mechanisms |
| <i>Galanthus nivalis</i> (GNA) | Amaryllidaceae | Man | does not bind Glc as the Leguminosae lectins do, application for insect and nematode defence in transgenic crop plants tested, antiretroviral activity in vitro, selective agglutination of rabbit but not human erythrocytes |
| <i>Ulex europaeus</i> (isolectin UEA-I) | Leguminosae | L-Fuc | agglutinates blood group O(H) erythrocytes; selective marker for endothelial cells of primates |
| <i>Erythrina corallodendron</i> (ECA) | Leguminosae | Galβ4GlcNAc, Gal, GalNAc | mitogen for human lymphocytes |
| <i>Vicia faba</i> (VFA) | Leguminosae | Man/Glc | binding of <i>N</i> -glycans enhanced by core fucosylation |
| <i>Sambucus nigra</i> (SNA) | Caprifoliaceae | Neu5Acα6Gal/GalNAc, (Gal/GalNAc) | probe for sialylated glycoconjugates, e.g. in thymocyte differentiation |
| <i>Abrus precatorius</i> | Leguminosae | Gal | RIP II used for generating immunotoxins |
| <i>Lotus tetragonolobus</i> (LTA) | Leguminosae | L-Fuc | agglutinates red cells of blood group O(H), instrumental to the definition of α-L-fucose as a crucial O(H) epitope (see Table 2) |
| <i>Lycopersicon esculentum</i> | Solanaceae | (GlcNAc) _n | potential function in plant defence mechanisms; marker of endothelium of small vessels in rats |
| <i>Phaseolus lunatus limensis</i> | Leguminosae | GalNAcα3[Fucα2]Gal, GalNAc | agglutinates blood group A erythrocytes |
| <i>Datura stramonium</i> (DSA) | Solanaceae | (GlcNAc) _n | potential function in plant defence mechanisms |
| <i>Maackia amurensis</i> (MAA) | Leguminosae | Neu5Acα3Gal/GalNAc | probe for sialylated glycoconjugates |
| <i>Phytolacca americana</i> (PWM) | Phytolaccaceae | GlcNAc | known as pokeweed mitogen; detected in 1969 in the course of investigating a fatality associated with ingestion of pokeweed berries |
| <i>Bauhinia purpurea</i> (BPA) | Leguminosae | GalNAcβ3GalNAc, GalNAc | enrichment of B lymphocytes, isolation of T cells producing IL-2 |
| <i>Urtica dioica</i> (UDA) | Urticaceae | (GlcNAc) _n | antifungal activity |
| <i>Hevea brasiliensis</i> (hevein) | Euphorbiaceae | (GlcNAc) _n | antifungal activity; allergen in rubber products of poor quality |
| <i>Maclura pomifera</i> (MPA) | Moraceae | T-antigen > Tn-antigen | mitogen for lymphocytes |

Figure 1. Examples of plant lectins and their inter- and intrafamily diversities (Gabius et al. 2004)

A recent review on the controversial issue of the clinical use of *Viscum album* extracts in cancer treatment concluded that “mistletoe therapy has the potential to harm cancer patients”(Ernst et al. 2003). By the way, it is difficult to produce a succinct compendium of lectin functions *in situ*. In principle, each lectin might have distinct functions at the site of expression and through interplay with binding partners in the cell and the extracellular environment, an idea also valid for animal lectins. One particular protein can thus take care of several tasks. Powerful techniques used to regulate lectin presence on the level of gene expression *in vitro* and *in vivo* that were a boon for the elucidation of lectin functions in animals are starting to be exploited in plants,(Brill et al. 2001; Esteban et al. 2002; Yong et al. 2003) so progress in refining and extending current knowledge of the functions of plant lectins will not be long in coming.

Animal Lectins: General Overview

A minimum of five lectin families has been solidly defined: C-, I-, and P type lectins, pentraxins, and galectins (Gabius 1997; Kilpatrick 2002). As presented in Figure 2, all these lectins bind distinct glycan determinants, providing a general mechanism of recognition and communication (Roth 1973), and offer enormous potential for applied research in chemical biology.

Endocytic receptors of the C-type lectin family, for instances those that allow the uptake into the hepatocytes cell, are ideal as targets for synthetically tailored drug carriers, due to their fixed geometry of binding sites (Davis et al. 2002; Lee 1992; Meijer et al. 1995; Rice 1997; Rogers et al. 1971). Antiviral drugs can thus be delivered to hepatocytes, for example by using triantennary N-glycans with GalNAc in the terminal position as a post code. Fixed topological presentation of binding sites is also a prerequisite for blocking access to bacterial/viral lectins, a new concept for interfering with the adhesion step of infections and the binding of AB5-toxins (Arosio et al. 2004; Mammen et al. 1998; Schengrund 2003). The fivefold symmetry of the presentation of the binding sites in these toxins provides the potential for extremely tight binding by a suitably designed pentavalent ligand.

Glycodendrimers have indeed been shown to impair binding of galectins both in solid-phase assays and in cell-binding studies (Andre et al. 2000; Andre et al. 2004; Andre et al. 2003; Andre et al. 1999; Andre et al. 2001; Vrasidas et al. 2003). The fine specificity differences between these homologous endogenous lectins are being delineated to enhance probe selectivity, another challenge for chemical biology (Hirabayashi et al. 2002; Wu et al. 2002). A theory is forming that the structure of the ligand and the spatial mode of its presentation modulate binding in markedly different ways for individual lectins of a family. The detection of these differences lends credit to the assumption that intrafamily diversification is accompanied by quantitative alterations of the ligand profile. Systematic chemical mapping with ligand derivatives and screening of arrays/libraries to discover potent ligand mimetics are likely eventually to allow molecules to be devised that fit hand-in-glove into a particular galectin (or any other lectin of clinical interest) (Barkley et al. 2001; Marcaurelle et al. 2002; Ramstrom et al. 2002; Rudiger et al. 2000; Solis et al. 2001).

| Activity | Example of lectin |
|--|--|
| ligand-selective molecular chaperones in endoplasmic reticulum | calnexin, calreticulin |
| intracellular routing of glycoproteins and vesicles | ERGIC-53 and VIP-36 (probably also ERGL and VIPL), P-type lectins, comitin |
| intracellular transport and extracellular assembly inducer of membrane superimposition and zippering (formation of Birbeck granules) | nonintegrin 67-kDa elastin/laminin-binding protein langerin (CD207) |
| cell-type-specific endocytosis | hepatic and macrophage asialoglycoprotein receptors, dendritic cell and macrophage C-type lectins (mannose receptor family members of the tandem-repeat type and single CRD lectins such as langerin/CD207), cysteine-rich domain of the dimeric form of the mannose receptor for GalNAc-4-SO ₄ -bearing glycoprotein hormones in hepatic endothelial cells, P-type lectins |
| recognition of foreign glycans(β1,3-glucans, LPS) | CR3 (CD11b/CD18), dectin-1, <i>Limulus</i> coagulation factors C and G, earthworm CCF |
| recognition of foreign or aberrant glycosignatures on cells (including endocytosis or initiation of opsonization or complement activation) | collectins, L-ficolin, C-type macrophage and dendritic cell receptors, α/θ-defensins, pentraxins (CRP, limulin), tachylectins |
| targeting of enzymatic activity in multimodular proteins | acrosin, laforin, <i>Limulus</i> coagulation factor C |
| intra- and intermolecular modulation of enzyme activities in vitro | porcine pancreatic α-amylase, galectin-1/α2-6-sialyltransferase |
| bridging of molecules | homodimeric and tandem-repeat-type galectins, cytokines (e.g. IL-2:IL-2R and CD3 of T-cell receptors), cerebellar soluble lectin |
| induction or suppression of effector release (H ₂ O ₂ , cytokines, etc.) | galectins, selectins, and other C-type lectins such as CD23, BDCA-2, and dectin-1 |
| cell growth control and induction of apoptosis/apoptosis | galectins, C-type lectins, amphoterin-like protein, hyaluronic-acid-binding proteins, cerebellar soluble lectin |
| cell migration and routing | selectins and other C-type lectins, I-type lectins, galectins, hyaluronic-acid-binding proteins (RHAMM, CD44, hyalectans/lecticans) |
| cell–cell interactions | selectins and other C-type lectins (e.g. DC-SIGN), galectins, I-type lectins (e.g. siglecs, N-CAM, P ₀ , or L1) |
| cell–matrix interactions | galectins, heparin- and hyaluronic-acid-binding lectins such as hyalectans/lecticans, calreticulin |
| matrix network assembly | proteoglycan core proteins (C-type CRD and G1 domain of hyalectans/lecticans), galectins (e.g. galectin-3/hensin), nonintegrin 67-kDa elastin/laminin-binding protein |

Figure 2. Functions of animal lectins (Gabijs et al. 2004)

Protein-Carbohydrate Interactions: CH- π , polar interactions, hydrogen bonding, and the role of water.

When oligosaccharides bind to proteins, it is usually with dissociation constants in the millimolar range and only occasionally, micromolar. Yet it is generally accepted that recognition phenomena require at least nanomolar dissociation constants because of the low effective concentrations of receptors. How, then, can oligosaccharides act as signals? Examination of the thermodynamics of oligosaccharide binding reveals the phenomenon of "entropy/enthalpy compensation"... Molecular models for the basis of compensation, suggest some answers to the question of how oligosaccharides can act as signals.

Jeremy P. Carver "Oligosaccharide: How can flexible molecules act as signals?" Pure & Appl. Chem. (1993) 65, 763-770

The typical contributions to the Gibbs' free energy of ligand binding originate basically from hydrogen bonding, van der Waals forces, and consequences of the hydrophobic effect. Factors that are important for predicting the affinity of a ligand also include any alterations in the geometry and motion dynamics of the receptor, ligand, and/or solvent molecules. As experimentally accessible parameters by calorimetric techniques, determination of the reaction enthalpy and entropy delineates the global driving force towards complex formation. These parameters have been measured, for example, for an array of mono- and disaccharides in the cases of a plant and an animal lectin sharing specificity to D-galactose (Bharadwaj et al. 1999), and the plot of the data (Fig. 3; according to the equation: $-\Delta H = -\Delta G - T\Delta S$) reveals a slope near unity and intercepts of -16.45 kJ/mol (plant lectin) and -23.12 kJ/mol (animal lectin). This conveys a fundamental message on the relationship between enthalpic and entropic factors attributed to the participation of weak intermolecular forces. An increase in enthalpy for ligand binding is inherently balanced by an entropic penalty (or vice versa), as a common enthalpy-entropy compensation effect (Dunitz 1995; Lumry et al. 1970).

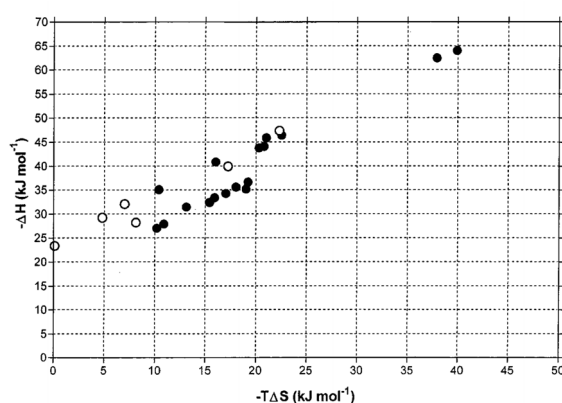


Fig. 3. Enthalpy-entropy compensation plot for the binding of a panel of mono- and disaccharides to the galactoside-specific mistletoe lectin (filled circles) and the galectin from adult chicken liver (open circles)

There are three prevailing models for the molecular basis of entropy/enthalpy compensation in the binding of oligosaccharides to proteins (Vyas 1991). The Lemieux group first reported compensation in the binding of oligosaccharides to proteins (Hindsgaul et al. 1985). Initially they attributed the observed compensation to the losses in protein degrees of freedom. More recently, Lemieux (Lemieux

1989) suggested that differential solvation effects are responsible for the compensation, a hypothesis closer to that of Lumry & Rajender (Lumry et al. 1970). Carver suggested a third model where the loss of conformational entropy by the ligand on binding provides the major source of the entropic barrier, at least for oligosaccharide ligands (Carver et al. 1989). All three mechanisms are undoubtedly operative to some extent. The question remains, however, as to which effect is predominant; and a complete answer could possibly derive by only an accurate analysis of the behavior of each participant at the molecular recognition process.

Carbohydrate display many hydroxyl groups. This supports the view that hydrogen bonds might dominate binding forces. When the spacing between two hydroxyl groups or the axial 4'-hydroxyl group and the ring oxygen atom matches that of an amino acid side chain (amide or carboxylate), two neighboring sites on the ligand can well be engaged in bidentate hydrogen bonds. The necessity for topological complementarity to yield the intricate network, schematically shown in Fig. 4, may be a source not only for enthalpy but also for selectivity, distinguishing anomers such as D-Gal vs. D-Man/D-Glc. It can thus be expected that the axial 4'-position for recognition of D-Gal and the equatorial 3',4'-positions for binding D-Man/D-Glc play decisive roles. This is strikingly verified by X-ray crystallography and in solution by chemically engineered ligand derivatives (Lis et al. 1998; Loris et al. 1998; Rini 1995; Rudiger et al. 2000; Solis et al. 1996; Weis et al. 1996). The same principle also applies to the characteristic formation of two coordination bonds with the central Ca^{2+} ion in the C-type lectins (Gabius 1997; Lis et al. 1998; Loukas et al. 1999).

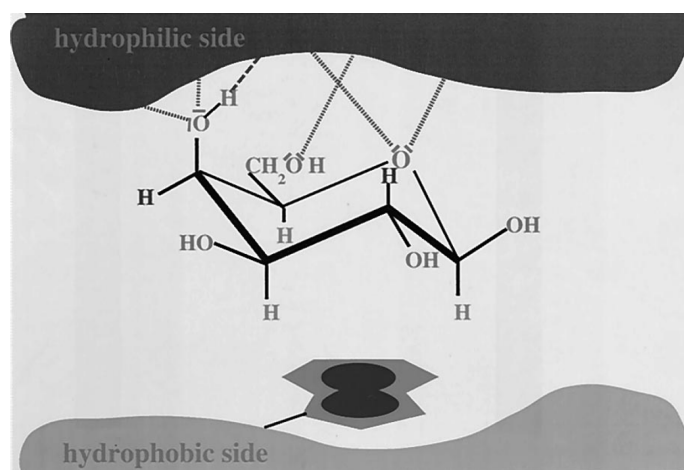


Fig. 4. The potential of D-galactose for establishing interactions with constituents of the binding pocket of a sugar receptor. While the rather polar upper side can be engaged in frequent hydrogen bonds exploiting lone electron pairs of oxygen atoms as acceptors and the protons of appropriately positioned hydroxyl groups as donors (and also coordination bonds with a Ca^{2+} ion in the case of C-type lectins), C-H/ π -electron interactions and entropically favorable stacking can be facilitated by an intimate contact of an aromatic (here: indolyl) amino acid side chain and the sugar's less polar bottom section.

Figure 4 also indicates another important feature driving ligand binding. While the upper side of Dgal is rather polar, the B-face exhibits a hydrophobic character. Stacking to the bulky aromatic amino acid side chain in the binding pocket removes both nonpolar surfaces from solvent accessibility,

although the two rings may not be aligned perfectly parallel. This alignment still contributes to complex stability and also to ligand selection despite its lower degree of directionality relative to hydrogen bonds. The shielding of the indolyl side chain by the ligand is reflected for galectins in molecular dynamics calculations and in differential UV, fluorescence, and laser photochemically induced dynamic nuclear polarization spectra (Asensio et al. 2000; Bernardi et al. 2004; Levi et al. 1981; Siebert et al. 1997). In addition to this impact on solvent molecules by reducing the apolar surface area, the π -electron cloud of the aromatic ring is likely to interact with the aliphatic D-Gal protons, which exhibit a net positive charge (Dougherty 1996; Nishio et al. 1995; Weis et al. 1996). Hydrophobic effect and van der Waals interactions may not be underestimated in affecting the overall Gibbs free energy gain. In fact the analysis of dominant forces in tight ligand binding for a variety of cases demonstrate that these factors can even surpass by far the contribution of hydrogen bonds (Kuntz et al. 1999).

CH- π Interactions

Sugars have significant nonpolar patches formed by the aliphatic protons and carbons at the various epimeric centers which extend out to the exocyclic 6 position of hexoses (Figure 5) and the glycerol moiety of sialic acids. This patch is usually packed against the face of one or more aromatic side chains of the protein. Removal of the apolar patch on the sugar, as well as the aromatic amino acid side chain from bulk solvent is believed to provide significant binding energy. In all lectin-Gal complex structures, the apolar patch of the Gal B face (Figure 5) packs against the face of aromatic residues (see Fig. 6 as example). This arrangement is frequently described as stacking, which implies that the ring of Gal is parallel to the plane of the aromatic ring (Fernandez-Alonso et al. 2005). In general sugar ring is always canted off somewhat from a parallel position. Superposition of the bound Gal seen in the various structures reveals that the aromatic rings form an angle between 17° and 52° with the least-squares plane through the pyranose ring and exocyclic carbon, with an average angle of 32° (Kolatkhar et al. 1996). In addition to the B-face of the sugar ring, the methyl group of the acetamido moiety of GlcNAc, GalNAc, and NeuNAc often interacts with an aromatic ring in lectins that specifically recognize this group, including WGA (Asensio et al. 1995; Asensio et al. 2000; Wright 1990) and HA (Weis et al. 1988). The carbon backbone of the glycerol moiety of NeuNAc can also present an apolar surface to the protein. For example, the glycerol moiety packs against the face of a tyrosine in WGA (Wright 1990). The infrequent use of aliphatic side chains in nonpolar interactions with sugars suggests that interaction with the delocalized electron cloud of the aromatic ring is energetically significant, beyond simply providing a geometrically complementary apolar surface. Most likely, the interaction is driven by the proximity of the aliphatic protons of the sugar ring, which carry a net positive partial charge, and the π -electron cloud of the aromatic ring. While Gal has thus far always been found packed against an aromatic side chain, this situation with Man is infrequent. These observations may reflect the fact that the axial disposition of the Gal 4-OH, as opposed to the equatorial 4-OH of Man, creates a more extensive and continuous nonpolar surface (Figure 5). No aromatic interactions are found in MBPs or GNA.

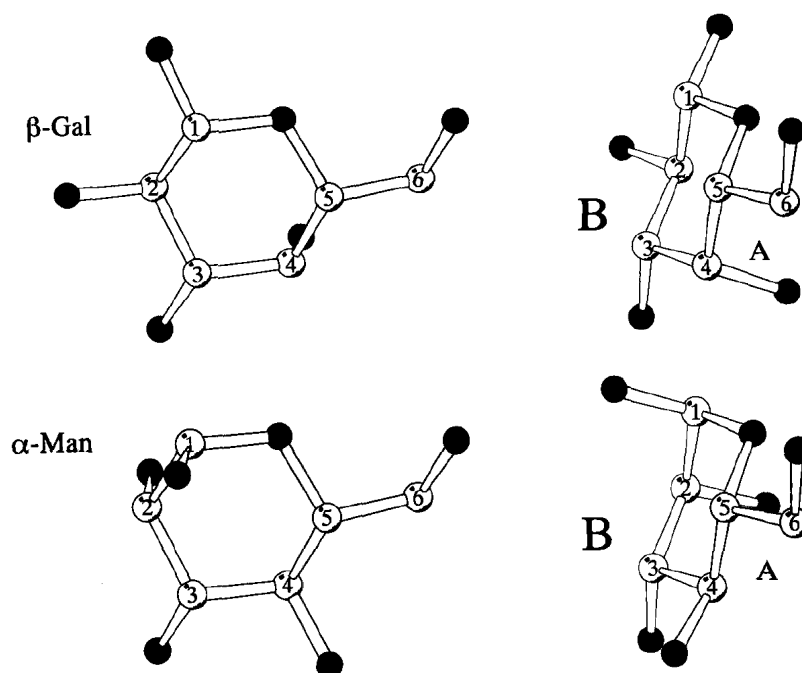


Figure 5 Comparison of nonpolar areas on Man and Gal. The B face, defined as that side of the ring that gives clockwise numbering along the ring (Rose et al. 1980), is shown in two views for each sugar. The axial 4-OH of Gal gives a more continuous aliphatic patch formed by the 3, 4, 5 and 6-carbons relative to Man.

Polar Interactions: Recognition of Charged Groups on Sugars and the Role of Divalent Cations

The most extensive structural data available for recognition of charged sugars by lectins are for sialic-acid binding lectins, namely Cholera Toxin (CT) (Merritt et al. 1994), influenza HA (Sauter et al. 1992; Watowich et al. 1994; Weis et al. 1988; Wright 1990; Wright 1984), and WGA (Wright 1990; Wright 1984). Thus far, only complexes with the most common sialic acid, N -acetylneuraminic acid (NeuNAc), and some of its derivatives (Sauter et al. 1992; Watowich et al. 1994) have been studied. Although salt-bridge formation between the NeuNAc carboxylate and positively charged amino acids might be expected to be important to recognition, the data show that the carboxylate moiety interacts with main-chain amide groups, polar side chains (especially serine), and ordered water molecules, rather than fully charged side chains (see Fig. 6 as example) in what are essentially hydrogen-bond interactions. In contrast, formation of salt bridges with the acid appears to be a common feature of neuraminidases (Burmeister et al. 1992; Crennell et al. 1994; Varghese et al. 1983). Charge-charge interactions do appear to have roles in the interaction between polyoma virus and NeuNAc-containing ligands. A low-resolution (3.6Å) structure of polyoma virus complexed with sialyl-lactose indicates that the NeuNAc carboxylate interacts with an arginine side chain (Stehle et al. 1994). Several classes of lectins require divalent cations for function. Of those with known three-dimensional structure, the legume lectins use Ca^{2+} and Mn^{2+} to stabilize the binding site and fix the positions of amino acids that interact with sugar ligands.

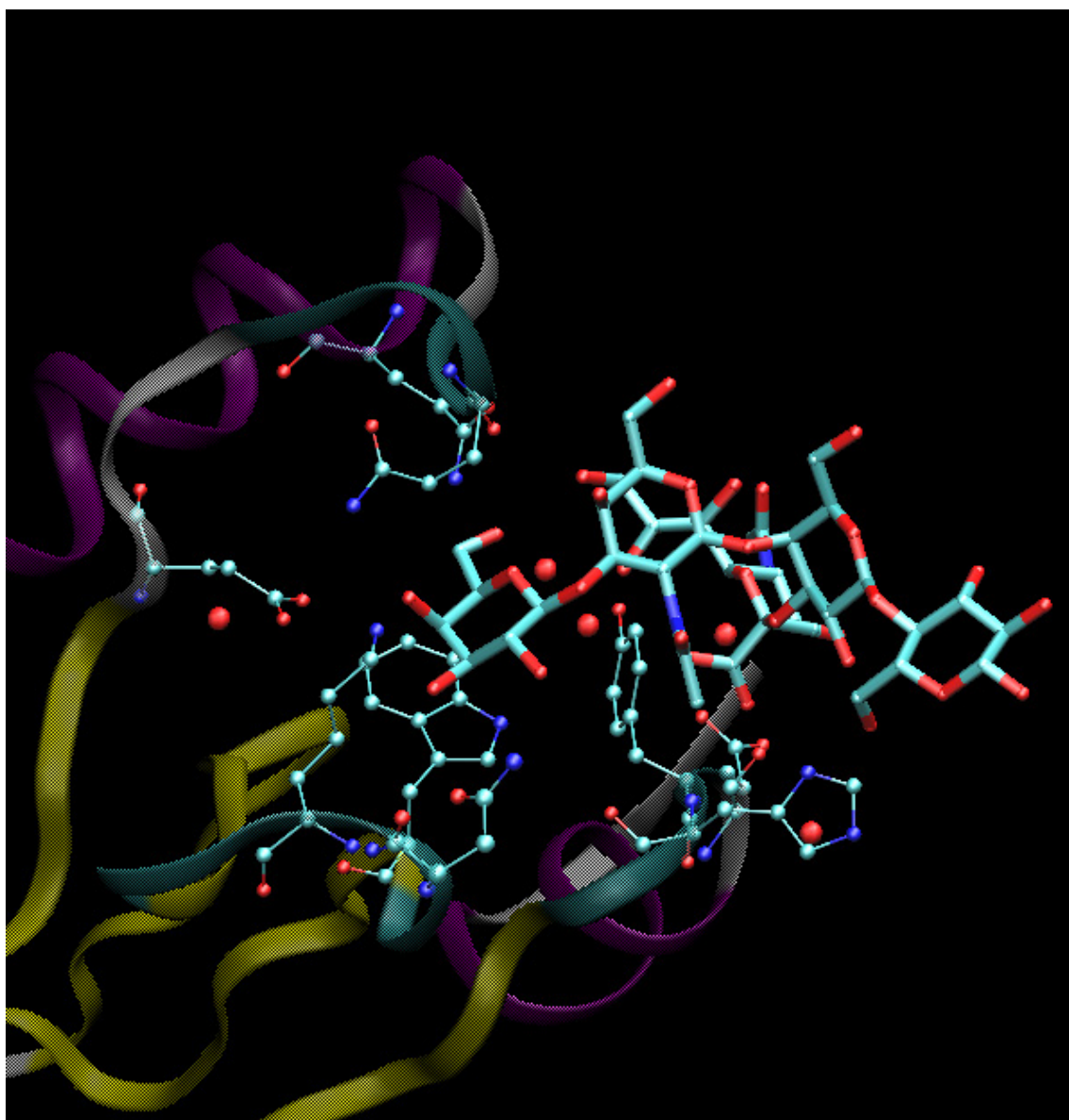


Figure 6 Structure of CT pentamer in complex with the pentasaccharide GM1 (Merritt et al. 1994)].

The Ca^{2+} forms coordination bonds with the side-chain carbonyl oxygen of a conserved asparagine residue, while the side-chain NH_2 of this asparagine donates a hydrogen bond to the sugar ligand. Likewise, one carboxylate oxygen of an acidic amino acid is bound to a water molecule that is in turn a Ca^{2+} ligand, and the other one accepts a hydrogen bond from the sugar. The Ca^{2+} coordination shell thus fixes side chains for optimal binding to the sugar.

The C-type lectins are unique among the structurally characterized lectins in that a required Ca^{2+} forms direct coordination bonds with the sugar ligand. In this case, the full noncovalent bonding potential of two vicinal OHs is used (Kolatkar et al. 1996; Ng et al. 1996; Weis et al. 1992): one lone pair of electrons from each OH forms a coordination bond with the Ca^{2+} . A Ca^{2+} also interacts directly with the

carboxylate moiety of a 4,6-0-(1-carboxyethylidene) galactoside derivative in SAP (Emsley et al. 1994). However, the Gal moiety does not interact with the Ca^{2+} , and whether binding of SAP to naturally-occurring sugars involves direct interactions with Ca^{2+} is not known. The only other direct sugar-metal interaction known in proteins is xylose isomerase, in which the sugar forms direct coordination bonds with two Mg^{2+} (Lavie et al. 1994).

Hydrogen Bonding and Water-Mediated Hydrogen Bonds

Cooperative hydrogen bonding, in which the hydroxyl group (OH) acts simultaneously as a hydrogen-bond donor and acceptor, is characteristic of the interaction of lectins and other carbohydrate-binding proteins with sugar hydroxyls. The oxygen atom of the -OH is sp^3 hybridized, giving an approximately tetrahedral arrangement of two lone pairs of electrons and a proton. The -OHs can therefore act as an acceptor of two hydrogen bonds and as a donor of a single hydrogen bond (Munoz et al. 2002; Vicente et al. 2004). Generally, one acidic side chain is used as a hydrogen bond acceptor from one or two sugar OHs. Hydrogen-bond donors come primarily from main-chain amide groups or from the side-chain amide group of asparagine and, less frequently, glutamine. Charged side-chain donors also occur with some frequency. Protein -OHs from tyrosine, serine, and threonine are much less common as either donors or acceptors of hydrogen bonds with sugar -OHs (Asensio et al. 2000). Thus, the most common hydrogen bonding scheme involving sugar -OHs is: $(\text{NH})_n \rightarrow \text{OH} \rightarrow \text{O}=\text{C}$; where NH is a hydrogen bond donor group, O=C is a carbonyl or carboxylate acceptor, OH is a non-anomeric sugar hydroxyl, and $n=1$ or 2. The hydrogen bond geometry of planar donors and acceptors is fixed, whereas an OH has torsional freedom in the placement of the proton and lone pairs. Presumably rotational freedom permits optimization of hydrogen bonds between the -OH and neighboring groups, albeit with some entropic cost due to fixing the -OH rotamer. The infrequent use of protein -OHs in hydrogen bonds with sugar OHs suggests that the entropic cost of fixing the rotamers of both sugar and protein OHs is too large.

In some cases, a pair of vicinal sugar hydroxyl groups (or one -OH group and the ring oxygen) interacts with two functional groups in a single amino acid side chain or with consecutive main-chain amide groups. Excellent hydrogen-bond geometry can be achieved between planar amino acid side chains and vicinal sugar -OHs in an equatorial/equatorial or equatorial/axial configuration: both sugar configurations give an OH-OH spacing of about 2.8 Å (Vyas 1991). For example, N δ 2 of Asn90 in CT donates to Gal 2-OH, and O δ 1 accepts a bond from Gal 3-OH (Fig. 6).

The sugar ring oxygen atom is also sp^3 hybridized and has two lone pairs of electrons in tetrahedral geometry that can act as hydrogen-bond acceptors but cannot participate in cooperative hydrogen bonds. When used in direct protein interactions, this oxygen usually shares a hydrogen bond-donating amino-acid side chain with one of the sugar hydroxyl groups. In general, this oxygen is common to all sugars and thus cannot be used to distinguish among sugars. The acetamido moiety of GlcNAc, GalNAc, and NeuNAc is often a dominant or significant recognition determinant. In contrary to the hydroxyl group, the amide group and carbonyl oxygen of the acetamido substituent have fixed,

planar geometry. The amide group acts as a hydrogen bond donor to planar carbonyl or carboxylate oxygens in all cases observed to date (see Fig. 6 as example), while the acetamido oxygen often accepts hydrogen bonds from serine.

In general, sugar functionalities that form hydrogen bonds with the lectin are those required for specific recognition and discrimination, whereas those positions that are not used as recognition elements tend to be solvent exposed and form no direct contacts with the protein. Hydrogen bonds between lectins and essential recognition determinants on sugars are shielded from bulk solvent, meaning that they exist in a lower dielectric environment and are probably enthalpically stronger than those formed with water (Quioco 1986). Moreover, the use of hydrogen bond donors and acceptors with fixed geometry may be important to specificity. A freely rotating group like the -OH of serine has some plasticity in the formation of hydrogen bonds with the sugar (Asensio et al. 1998; Asensio et al. 1995; Asensio et al. 2000) and therefore may not be capable of discriminating absolutely between, for example, epimeric hydroxyl groups.

Regarding the role water molecules, it was observed that they mediate hydrogen bonds between sugar and protein in many of the lectin-carbohydrate complex structures determined at high resolution. Studies of glycogen phosphorylase inhibitors have shown that water-mediated hydrogen bonds between sugar and protein can be as strong as direct protein-sugar hydrogen bonds (Watson et al. 1994). Comparison of a series of sugars bound to a given lectin, or a series of related lectins bound to a given sugar, sometimes reveals common water molecules, which can be presumed to be important elements in recognition. For example, three water molecules are observed to mediate identical interactions between Gal and both LT and CT (Merritt et al. 1994; Merritt et al. 1994), and in both the galactose and lactose complexes of LT (Merritt et al. 1994; Sixma et al. 1992) (Fig. 6). Also, comparison of the binding site in its free and bound states shows that some water molecules bound to the protein in the sugar-free state remain in the presence of sugar and form hydrogen bonds with the ligand. In this case the water molecules act as fixed structural elements, equivalent to hydrogen-bonding groups of the protein, and can therefore be considered part of the binding site architecture.

References

- Akama, T. O., H. Nakagawa, K. Sugihara, S. Narisawa, C. Ohyama, S. Nishimura, D. A. O'Brien, K. W. Moremen, J. L. Millan and M. N. Fukuda (2002) "Germ cell survival through carbohydrate-mediated interaction with Sertoli cells" *Science* **295**(5552): 124-7.
- Andre, S., B. Frisch, H. Kaltner, D. L. Desouza, F. Schuber and H. J. Gabius (2000) "Lectin-mediated drug targeting: selection of valency, sugar type (Gal/Lac), and spacer length for cluster glycosides as parameters to distinguish ligand binding to C-type asialoglycoprotein receptors and galectins" *Pharm Res* **17**(8): 985-90.
- Andre, S., H. Kaltner, T. Furuike, S. Nishimura and H. J. Gabius (2004) "Persubstituted cyclodextrin-based glycoclusters as inhibitors of protein-carbohydrate recognition using purified plant and mammalian lectins and wild-type and lectin-gene-transfected tumor cells as targets" *Bioconjug Chem* **15**(1): 87-98.
- Andre, S., B. Liu, H. J. Gabius and R. Roy (2003) "First demonstration of differential inhibition of lectin binding by synthetic tri- and tetravalent glycoclusters from cross-coupling of rigidified 2-propynyl lactoside" *Org Biomol Chem* **1**(22): 3909-16.
- Andre, S., P. J. Ortega, M. A. Perez, R. Roy and H. J. Gabius (1999) "Lactose-containing starburst dendrimers: influence of dendrimer generation and binding-site orientation of receptors (plant/animal lectins and immunoglobulins) on binding properties" *Glycobiology* **9**(11): 1253-61.
- Andre, S., R. J. Pieters, I. Vrasidas, H. Kaltner, I. Kuwabara, F. T. Liu, R. M. Liskamp and H. J. Gabius (2001) "Wedgelike glycodendrimers as inhibitors of binding of mammalian galectins to glycoproteins, lactose maxiclusters, and cell surface glycoconjugates" *ChemBiochem* **2**(11): 822-30.
- Arosio, D., I. Vrasidas, P. Valentini, R. M. Liskamp, R. J. Pieters and A. Bernardi (2004) "Synthesis and cholera toxin binding properties of multivalent GM1 mimics" *Org Biomol Chem* **2**(14): 2113-24.
- Asensio, J. L., F. J. Canada, M. Bruix, C. Gonzalez, N. Khiar, A. Rodriguez-Romero and J. Jimenez-Barbero (1998) "NMR investigations of protein-carbohydrate interactions: refined three-dimensional structure of the complex between hevein and methyl beta-chitobioside" *Glycobiology* **8**(6): 569-77.
- Asensio, J. L., F. J. Canada, M. Bruix, A. Rodriguez-Romero and J. Jimenez-Barbero (1995) "The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose" *Eur J Biochem* **230**(2): 621-33.
- Asensio, J. L., F. J. Canada, H. C. Siebert, J. Laynez, A. Poveda, P. M. Nieto, U. M. Soedjanaamadja, H. J. Gabius and J. Jimenez-Barbero (2000) "Structural basis for chitin recognition by defense proteins: GlcNAc residues are bound in a multivalent fashion by extended binding sites in hevein domains" *Chem Biol* **7**(7): 529-43.
- Asensio, J. L., H. C. Siebert, C. W. von Der Lieth, J. Laynez, M. Bruix, U. M. Soedjanaamadja, J. J. Beintema, F. J. Canada, H. J. Gabius and J. Jimenez-Barbero (2000) "NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N"-triacylchitotriose" *Proteins* **40**(2): 218-36.
- Barkley, A. and P. Arya (2001) "Combinatorial chemistry toward understanding the function(s) of carbohydrates and carbohydrate conjugates" *Chemistry* **7**(3): 555-63.
- Barondes, S. H. (1988) "Bifunctional properties of lectins: lectins redefined" *Trends Biochem Sci* **13**(12): 480-2.
- Bernardi, A., D. Arosio, D. Potenza, I. Sanchez-Medina, S. Mari, F. J. Canada and J. Jimenez-Barbero (2004) "Intramolecular carbohydrate-aromatic interactions and intermolecular van der Waals interactions enhance the molecular recognition ability of GM1 glycomimetics for cholera toxin" *Chemistry* **10**(18): 4395.

- Bharadwaj, S., H. Kaltner, E. Y. Korchagina, N. V. Bovin, H. J. Gabius and A. Surolia (1999) "Microcalorimetric indications for ligand binding as a function of the protein for galactoside-specific plant and avian lectins" *Biochim Biophys Acta* **1472**(1-2): 191-6.
- Brill, L. M., C. J. Evans and A. M. Hirsch (2001) "Expression of MsLEC1- and MsLEC2-antisense genes in alfalfa plant lines causes severe embryogenic, developmental and reproductive abnormalities" *Plant J* **25**(4): 453-61.
- Burmeister, W. P., R. W. Ruigrok and S. Cusack (1992) "The 2.2 Å resolution crystal structure of influenza B neuraminidase and its complex with sialic acid" *Embo J* **11**(1): 49-56.
- Carver, J. P., S. W. Michnick, A. Imberty and D. A. Cumming (1989) "Oligosaccharide-protein interactions: a three-dimensional view" *Ciba Found Symp* **145**: 6-18; discussion 18-26.
- Chouaib, S., C. Asselin-Paturel, F. Mami-Chouaib, A. Caignard and J. Y. Blay (1997) "The host-tumor immune conflict: from immunosuppression to resistance and destruction" *Immunol Today* **18**(10): 493-7.
- Coussens, L. M. and Z. Werb (2002) "Inflammation and cancer" *Nature* **420**(6917): 860-7.
- Crennell, S., E. Garman, G. Laver, E. Vimr and G. Taylor (1994) "Crystal structure of *Vibrio cholerae* neuraminidase reveals dual lectin-like domains in addition to the catalytic domain" *Structure* **2**(6): 535-44.
- Cummings, R. D. (1997) in *Lectins as tools for glycoconjugate purification and characterization*. Gabius HJ and Gabius S (eds) Glycosciences Status and Perspectives. Weinheim, Germany, Chapman & Hall, p. 191.
- Damjanov, I. (1987) "Lectin cytochemistry and histochemistry" *Lab Invest* **57**(1): 5-20.
- Danguy, A., F. Akif, B. Pajak and H. J. Gabius (1994) "Contribution of carbohydrate histochemistry to glycobiology" *Histol Histopathol* **9**(1): 155-71.
- Davis, B. G. and M. A. Robinson (2002) "Drug delivery systems based on sugar-macromolecule conjugates" *Curr Opin Drug Discov Devel* **5**(2): 279-88.
- Dougherty, D. A. (1996) "Cation- π interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp" *Science* **271**(5246): 163-8.
- Drickamer, K. and M. E. Taylor (1993) "Biology of animal lectins" *Annu Rev Cell Biol* **9**: 237-64.
- Drickamer, K. and M. E. Taylor (1998) "Evolving views of protein glycosylation" *Trends Biochem Sci* **23**(9): 321-4.
- Dunitz, J. D. (1995) "Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions" *Chem Biol* **2**(11): 709-12.
- Dwek, R. A. (1996) "Glycobiology: Toward Understanding the Function of Sugars" *Chem Rev* **96**(2): 683-720.
- Emsley, J., H. E. White, B. P. O'Hara, G. Oliva, N. Srinivasan, I. J. Tickle, T. L. Blundell, M. B. Pepys and S. P. Wood (1994) "Structure of pentameric human serum amyloid P component" *Nature* **367**(6461): 338-45.
- Ernst, E., K. Schmidt and M. K. Steuer-Vogt (2003) "Mistletoe for cancer? A systematic review of randomised clinical trials" *Int J Cancer* **107**(2): 262-7.
- Esteban, R., B. Dopico, F. J. Munoz, S. Romo and E. Labrador (2002) "A seedling specific vegetative lectin gene is related to development in *Cicer arietinum*" *Physiol Plant* **114**(4): 619-626.
- Fernandez-Alonso, M., F. J. Canada, J. Jimenez-Barbero and G. Cuevas (2005) "Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions" *J Am Chem Soc* **127**(20): 7379-86.
- Gabius, H. J. (1994) "Non-carbohydrate binding partners/domains of animal lectins" *Int J Biochem* **26**(4): 469-77.
- Gabius, H. J. (1997) "Animal lectins" *Eur J Biochem* **243**(3): 543-76.

- Gabius, H. J. (2000) in *Protein liquid chromatography*. Kastner, M. Amsterdam ; New York, Elsevier.
- Gabius, H. J. (2001) "Probing the cons and pros of lectin-induced immunomodulation: case studies for the mistletoe lectin and galectin-1" *Biochimie* **83**(7): 659-66.
- Gabius, H. J. and S. Gabius (1997) *Glycosciences : status and perspectives (Chapter: Sharon N, Lis H "Glycoproteins: structure and function")*, London ; New York, Chapman & Hall.
- Gabius, H. J., H. C. Siebert, S. Andre, J. Jimenez-Barbero and H. Rudiger (2004) "Chemical biology of the sugar code" *ChemBiochem* **5**(6): 740-64.
- Gagneux, P. and A. Varki (1999) "Evolutionary considerations in relating oligosaccharide diversity to biological function" *Glycobiology* **9**(8): 747-55.
- Goldstein, I. P., H. C. Winter and R. D. Poretz (1997) in *Glycoproteins II*. Montreuil, J., Vliegenthart, J. F. G., Schachter, H. Amsterdam ; New York, Elsevier, p. 403.
- Guo, Y., H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer (2004) "Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR" *Nat Struct Mol Biol* **11**(7): 591-8.
- Hajto, T., K. Hostanska, K. Frei, C. Rordorf and H. J. Gabius (1990) "Increased secretion of tumor necrosis factors alpha, interleukin 1, and interleukin 6 by human mononuclear cells exposed to beta-galactoside-specific lectin from clinically applied mistletoe extract" *Cancer Res* **50**(11): 3322-6.
- Hajto, T., K. Hostanska and H. J. Gabius (1989) "Modulatory potency of the beta-galactoside-specific lectin from mistletoe extract (Iscador) on the host defense system in vivo in rabbits and patients" *Cancer Res* **49**(17): 4803-8.
- Hindsgaul, O., D. P. Khare, M. Bach and R. U. Lemieux (1985) "Molecular Recognition III. The Binding of the H-type 2 Human Blood Group Determinant by the Lectin I of *Ulex Europaeus*" *Can. J. Chem.* **63**: 2653-2658.
- Hirabayashi, J., T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W. E. Muller, F. Yagi and K. Kasai (2002) "Oligosaccharide specificity of galectins: a search by frontal affinity chromatography" *Biochim Biophys Acta* **1572**(2-3): 232-54.
- Kilpatrick, D. C. (2002) "Animal lectins: a historical introduction and overview" *Biochim Biophys Acta* **1572**(2-3): 187-97.
- Kolatkari, A. R. and W. I. Weis (1996) "Structural basis of galactose recognition by C-type animal lectins" *J Biol Chem* **271**(12): 6679-85.
- Kuntz, I. D., K. Chen, K. A. Sharp and P. A. Kollman (1999) "The maximal affinity of ligands" *Proc Natl Acad Sci U S A* **96**(18): 9997-10002.
- Laine, R. A. (1994) "A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems" *Glycobiology* **4**(6): 759-67.
- Lavie, A., K. N. Allen, G. A. Petsko and D. Ringe (1994) "X-ray crystallographic structures of D-xylose isomerase-substrate complexes position the substrate and provide evidence for metal movement during catalysis" *Biochemistry* **33**(18): 5469-80.
- Lee, Y. C. (1992) "Biochemistry of carbohydrate-protein interaction" *Faseb J* **6**(13): 3193-200.
- Lemieux, R. U. (1989) "The origin of the specificity in the recognition of oligosaccharides by proteins" *Chem Soc Rev* **18**: 347-374.
- Levi, G. and V. I. Teichberg (1981) "Isolation and physicochemical characterization of electrolectin, a beta-D-galactoside binding lectin from the electric organ of *Electrophorus electricus*" *J Biol Chem* **256**(11): 5735-40.
- Liener, I. E., N. Sharon and I. J. Goldstein (1986) *Kocourek, J. in "The Lectins : properties, functions, and applications in biology and medicine"*, Orlando, Academic Press.

- Lis, H. and N. Sharon (1986) in *The Lectins : properties, functions, and applications in biology and medicine*. Liener, I. E., Sharon, N., Goldstein, I. J. Orlando, Academic Press, p. 293.
- Lis, H. and N. Sharon (1986) "Lectins as molecules and as tools" *Annu Rev Biochem* **55**: 35-67.
- Lis, H. and N. Sharon (1998) "Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition" *Chem Rev* **98**(2): 637-674.
- Loris, R., T. Hamelryck, J. Bouckaert and L. Wyns (1998) "Legume lectin structure" *Biochim Biophys Acta* **1383**(1): 9-36.
- Loukas, A., N. P. Mullin, K. K. Tetteh, L. Moens and R. M. Maizels (1999) "A novel C-type lectin secreted by a tissue-dwelling parasitic nematode" *Curr Biol* **9**(15): 825-8.
- Lumry, R. and S. Rajender (1970) "Enthalpy-entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water" *Biopolymers* **9**(10): 1125-227.
- Mammen, M., S. K. Choi and G. M. Whitesides (1998) "Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors" *Angew Chem Int Ed Engl* **37**: 2754 - 2794.
- Marcaurelle, L. A. and P. H. Seeberger (2002) "Combinatorial carbohydrate chemistry" *Curr Opin Chem Biol* **6**(3): 289-96.
- Meijer, D. K. and G. Molema (1995) "Targeting of drugs to the liver" *Semin Liver Dis* **15**(3): 202-56.
- Merritt, E. A., S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial and W. G. Hol (1994) "Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide" *Protein Sci* **3**(2): 166-75.
- Merritt, E. A., T. K. Sixma, K. H. Kalk, B. A. van Zanten and W. G. Hol (1994) "Galactose-binding site in Escherichia coli heat-labile enterotoxin (LT) and cholera toxin (CT)" *Mol Microbiol* **13**(4): 745-53.
- Munoz, E. M., M. Lopez de la Paz, J. Jimenez-Barbero, G. Ellis, M. Perez and C. Vicent (2002) "The relevance of carbohydrate hydrogen-bonding cooperativity effects: a cooperative 1,2-trans-diaxial diol and amido alcohol hydrogen-bonding array as an efficient carbohydrate-phosphate binding motif in nonpolar media" *Chemistry* **8**(8): 1908-14.
- Ng, K. K., K. Drickamer and W. I. Weis (1996) "Structural analysis of monosaccharide recognition by rat liver mannose-binding protein" *J Biol Chem* **271**(2): 663-74.
- Nishio, M., Y. Umezawa, M. Hirota and Y. Takeuchi (1995) "The CH/p interaction: Significance in molecular recognition" *Tetrahedron* **51**(32): 8665-8701.
- Osawa, T. and T. Tsuji (1987) "Fractionation and structural assessment of oligosaccharides and glycopeptides by use of immobilized lectins" *Annu Rev Biochem* **56**: 21-42.
- Peumans, W. J. and E. J. Van Damme (1998) "Plant lectins: specific tools for the identification, isolation, and characterization of O-linked glycans" *Crit Rev Biochem Mol Biol* **33**(3): 209-58.
- Quioco, F. A. (1986) "Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions" *Annu Rev Biochem* **55**: 287-315.
- Ramstrom, O., T. Bunyapaiboonsri, S. Lohmann and J. M. Lehn (2002) "Chemical biology of dynamic combinatorial libraries" *Biochim Biophys Acta* **1572**(2-3): 178-86.
- Reuter, G. and H. J. Gabius (1999) "Eukaryotic glycosylation: whim of nature or multipurpose tool?" *Cell Mol Life Sci* **55**(3): 368-422.
- Rice, K. G. (1997) in *Glycosciences : status and perspectives*. Gabius, H. J., Gabius, S. London ; New York, Chapman & Hall, p. 471.
- Rini, J. M. (1995) "Lectin structure" *Annu Rev Biophys Biomol Struct* **24**: 551-77.
- Rogers, J. C. and S. Kornfeld (1971) "Hepatic uptake of proteins coupled to fetuin glycopeptide" *Biochem Biophys Res Commun* **45**(3): 622-9.
- Rose, I. A., K. R. Hanson, K. D. Wilkinson and M. J. Wimmer (1980) "A Suggestion for Naming Faces of Ring Compounds" *Proc Natl Acad Sci U S A* **77**: 2439-2441.

- Roseman, S. (2001) "Reflections on glycobiology" *J Biol Chem* **276**(45): 41527-42.
- Roth, S. A. (1973) "A molecular model for cell interactions" *Quart. Rev. Biol.* **48**: 541-563.
- Rudd, P. M., F. Fortune, T. Patel, R. B. Parekh, R. A. Dwek and T. Lehner (1994) "A human T-cell receptor recognizes 'O'-linked sugars from the hinge region of human IgA1 and IgD" *Immunology* **83**(1): 99-106.
- Rudiger, H. and H. J. Gabius (2001) "Plant lectins: occurrence, biochemistry, functions and applications" *Glycoconj J* **18**(8): 589-613.
- Rudiger, H., H. C. Siebert, D. Solis, J. Jimenez-Barbero, A. Romero, C. W. von der Lieth, T. Diaz-Maurino and H. J. Gabius (2000) "Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets" *Curr Med Chem* **7**(4): 389-416.
- Sauter, N. K., J. E. Hanson, G. D. Glick, J. H. Brown, R. L. Crowther, S. J. Park, J. J. Skehel and D. C. Wiley (1992) "Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography" *Biochemistry* **31**(40): 9609-21.
- Schengrund, C. L. (2003) "'Multivalent' saccharides: development of new approaches for inhibiting the effects of glycosphingolipid-binding pathogens" *Biochem Pharmacol* **65**(5): 699-707.
- Science (2001) "Carbohydrates and Glycobiology" *Science, Special Issue* **291**: 2263-2502.
- Sharon, N. and H. Lis (1987) "A century of lectin research (1888–1988)" *Trends Biochem. Sci.* **12**: 483-487.
- Sharon, N. and H. Lis (1989) *Lectins*, London ; New York, Chapman and Hall.
- Sharon, N. and H. Lis (1989) "Lectins as cell recognition molecules" *Science* **246**(4927): 227-34.
- Sharon, N. and H. Lis (1993) "Carbohydrates in cell recognition" *Sci Am* **268**(1): 82-9.
- Siebert, H. C., R. Adar, R. Arango, M. Burchert, H. Kaltner, G. Kayser, E. Tajkhorshid, C. W. von der Lieth, R. Kaptein, N. Sharon, J. F. Vliegthart and H. J. Gabius (1997) "Involvement of laser photo-CIDNP (chemically induced dynamic nuclear polarization)-reactive amino acid side chains in ligand binding by galactoside-specific lectins in solution" *Eur J Biochem* **249**(1): 27-38.
- Sixma, T. K., S. E. Pronk, K. H. Kalk, B. A. van Zanten, A. M. Berghuis and W. G. Hol (1992) "Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography" *Nature* **355**(6360): 561-4.
- Sogn, J. A. (1998) "Tumor immunology: the glass is half full" *Immunity* **9**(6): 757-63.
- Solis, D., J. Jimenez-Barbero, H. Kaltner, A. Romero, H. C. Siebert, C. W. von der Lieth and H. J. Gabius (2001) "Towards defining the role of glycans as hardware in information storage and transfer: basic principles, experimental approaches and recent progress" *Cells Tissues Organs* **168**(1-2): 5-23.
- Solis, D., A. Romero, H. Kaltner, H. J. Gabius and T. Diaz-Maurino (1996) "Different architecture of the combining site of the two chicken galectins revealed by chemical mapping studies with synthetic ligand derivatives" *J Biol Chem* **271**(22): 12744-8.
- Stehle, T., Y. Yan, T. L. Benjamin and S. C. Harrison (1994) "Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment" *Nature* **369**(6476): 160-3.
- Varghese, J. N., W. G. Laver and P. M. Colman (1983) "Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution" *Nature* **303**(5912): 35-40.
- Varki, A. (1993) "Biological roles of oligosaccharides: all of the theories are correct" *Glycobiology* **3**(2): 97-130.
- Varki, A. (1996) "'Unusual' modifications and variations of vertebrate oligosaccharides: are we missing the flowers for the trees?" *Glycobiology* **6**(7): 707-10.
- Vicente, V., J. Martin, J. Jimenez-Barbero, J. L. Chiara and C. Vicent (2004) "Hydrogen-bonding cooperativity: using an intramolecular hydrogen bond to design a carbohydrate derivative with a cooperative hydrogen-bond donor centre" *Chemistry* **10**(17): 4240-51.

- Vrasidas, I., S. Andre, P. Valentini, C. Bock, M. Lensch, H. Kaltner, R. M. Liskamp, H. J. Gabius and R. J. Pieters (2003) "Rigidified multivalent lactose molecules and their interactions with mammalian galectins: a route to selective inhibitors" *Org Biomol Chem* **1**(5): 803-10.
- Vyas, N. K. (1991) "Atomic features of protein-carbohydrate interactions" *Curr Opin Struct Biol* **1**(5): 732-740.
- Watowich, S. J., J. J. Skehel and D. C. Wiley (1994) "Crystal structures of influenza virus hemagglutinin in complex with high-affinity receptor analogs" *Structure* **2**(8): 719-31.
- Watson, K. A., E. P. Mitchell, L. N. Johnson, J. C. Son, C. J. Bichard, M. G. Orchard, G. W. Fleet, N. G. Oikonomakos, D. D. Leonidas, M. Kontou and et al. (1994) "Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-C-glucosides and 1-thio-beta-D-glucose compounds" *Biochemistry* **33**(19): 5745-58.
- Weis, W., J. H. Brown, S. Cusack, J. C. Paulson, J. J. Skehel and D. C. Wiley (1988) "Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid" *Nature* **333**(6172): 426-31.
- Weis, W. I., G. V. Crichlow, H. M. Murthy, W. A. Hendrickson and K. Drickamer (1991) "Physical characterization and crystallization of the carbohydrate-recognition domain of a mannose-binding protein from rat" *J Biol Chem* **266**(31): 20678-86.
- Weis, W. I. and K. Drickamer (1996) "Structural basis of lectin-carbohydrate recognition" *Annu Rev Biochem* **65**: 441-73.
- Weis, W. I., K. Drickamer and W. A. Hendrickson (1992) "Structure of a C-type mannose-binding protein complexed with an oligosaccharide" *Nature* **360**(6400): 127-34.
- Wright, C. S. (1984) "Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II" *J Mol Biol* **178**(1): 91-104.
- Wright, C. S. (1990) "2.2 A resolution structure analysis of two refined N-acetylneuraminyllactose--wheat germ agglutinin isolectin complexes" *J Mol Biol* **215**(4): 635-51.
- Wu, A. M., J. H. Wu, M. S. Tsai, J. H. Liu, S. Andre, K. Wasano, H. Kaltner and H. J. Gabius (2002) "Fine specificity of domain-I of recombinant tandem-repeat-type galectin-4 from rat gastrointestinal tract (G4-N)" *Biochem J* **367**(Pt 3): 653-64.
- Yong, W. D., Y. Y. Xu, W. Z. Xu, X. Wang, N. Li, J. S. Wu, T. B. Liang, K. Chong, Z. H. Xu, K. H. Tan and Z. Q. Zhu (2003) "Vernalization-induced flowering in wheat is mediated by a lectin-like gene VER2" *Planta* **217**(2): 261-70.
- Yu, J. L. and J. W. Rak (2003) "Host microenvironment in breast cancer development: inflammatory and immune cells in tumour angiogenesis and arteriogenesis" *Breast Cancer Res* **5**(2): 83-8.

The three dimensional structures of oligosaccharides and glycomimetics. Factors and facts

The shape of a given oligosaccharide is determined by the conformation of the furanose/pyranose rings and the relative positioning of the rings in the chain. Based on X-ray crystallography, neutron diffraction, and homonuclear coupling constant data, the 4C_1 chair conformer (1C_4 for L-sugars) is the energetically preferred D-pyranose ring structure ((Abeygunawardana et al. 1993; Brown et al. 1965). In rare cases, for example, for L-iduronic acid, as constituent of heparan and dermatan sulfates, and to accommodate mechanical stress, conformational flexibility and elasticity of a pyranose can be generated by chair-boat transitions, which allow L-iduronic acid to acquire the skew-boat form 2S_0 (Bush et al. 1999; Canales et al. 2005; Casu et al. 1988; Marszalek et al. 1998; von der Lieth et al. 1998). However, the main contribution to defining the shape of a glycan generally does not originate from this source but rather from changes of the three dihedral angles Φ , Ψ and ω of each glycosidic bond (Fig. 7). Since the pyranose rings linked by the glycosidic bond and their exocyclic substituents are rather bulky, their size imposes topological restraints on the intramolecular movements of the oligomer.

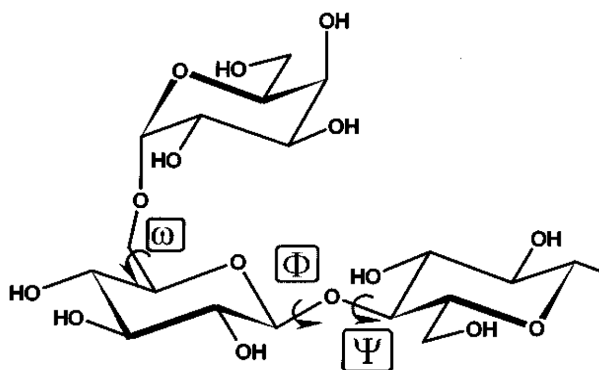


Figure 7 Schematic view of the glycosidic torsion angles which define the three dimensional shape of an oligosaccharide

Compared to oligopeptides with small side chains, the conformational space accessible to the molecule at room temperature is thus relatively restricted. That this spatial factor limits the range of interchangeable conformations has been inferred by computer-assisted molecular mechanics and dynamics calculations and convincingly documented by experimental evidence primarily from sophisticated magnetic resonance spectroscopy (Bush et al. 1999; von der Lieth et al. 1998). Exploring the actual position(s) of each oligosaccharide on the scale between high flexibility with an ensemble of conformers and almost complete rigidity definitely has salient implications for predicting its role as a coding unit. In this respect, it is also worth pointing out that a notable level of intramolecular flexibility

is not a favorable factor for crystallization. Indeed, such an extent of unrestrained conformational entropy can contribute to explaining the frequently frustrating experience in respective attempts at carbohydrate chemistry. If, on the other hand, the level of conformational entropy is confined to only very few stable conformers (keys), the presented shape distribution is a function not only of the sequence but also of external factors affecting the current status of the equilibrium. Sugar receptors as probes for distinguishing bioactive or bioinert glycan presentation modes on proteins have already provided the hypothesis experimental support (Mann et al. 1998; Noorman et al. 1998; Solis et al. 1987). In next sections a general overview of factors and facts that define the 3D-structure of a oligosaccharide is presented.

The shape of the six-membered rings

Emil Fischer introduced the classical projection formulae for sugars, with a standard orientation (carbon chain vertical, carbonyl group at the top; Fig. 8); since he used models with flexible bonds between the atoms, he could easily 'stretch' the sugar models into a position suitable for projection. He assigned to the dextrorotatory glucose (via the derived glucaric acid) the projection with the OH group at C-5 pointing to the right, well knowing that there was a 50% chance that this was wrong.

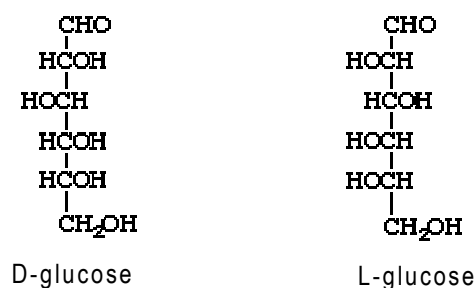


Figure 8 Fisher's projections for glucose

Towards the end of the nineteenth century, it was realized that the free sugars (not only the glycosides) existed as cyclic hemiacetals or hemiketals. Mutarotation, discovered in 1846 by Dubrunfaut, was now interpreted as being due to a change in the configuration of the glycosidic (anomeric) carbon atom. Emil Fischer assumed the cyclic form to be a five-membered ring, which Tollens designated by the symbol <1,4>, while the six-membered ring received the symbol <1,5>. In the 1920s, Haworth proposed the terms 'furanose' and 'pyranose' for the two forms. He also introduced the 'Haworth depiction' for writing structural formulae, a convention that was soon widely followed.

Cyclic forms with a three-membered ring are called oxiroses, those with a four-membered ring oxetoses, those with a five-membered ring furanoses, with a six-membered ring pyranoses, with a seven-membered ring septanoses, with an eight-membered ring octanoses, and so on.

If a cyclic form of a sugar is to be represented in the Fischer projection, a long bond can be drawn between the oxygen involved in ring formation and the (anomeric) carbon atom to which it is linked, as shown in the following formulae for cyclic forms of α -D-glucose (Fig. 9)

In the Haworth representation of the cyclic form (Fig. 10) the ring is orientated almost perpendicular to the plane of the paper, but viewed from slightly above so that the edge closer to the viewer is drawn below the more distant edge, with the oxygen behind and C-1 at the right-hand end. To define the perspective, the ring bonds closer to the viewer are often thickened.

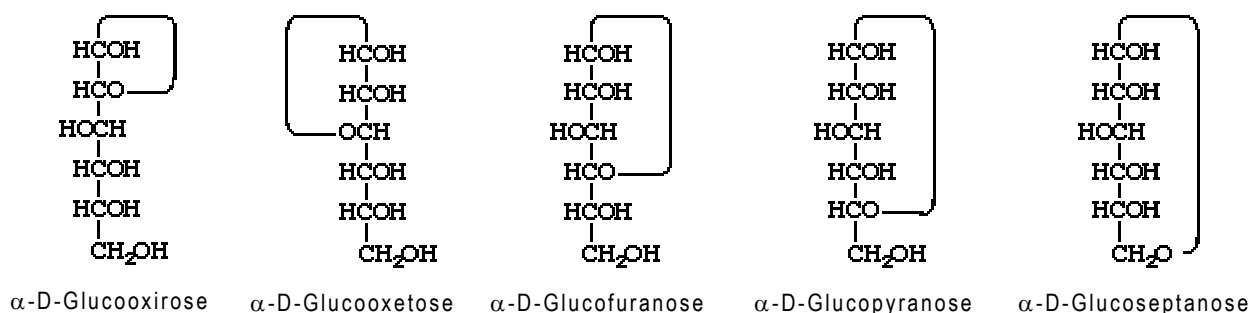


Figure 9 Fisher projections for the cyclic forms

Figure 10 shows the schematic representation of pyranose ring closure in D-glucose with the reorientation at C-5 necessary to allow ring formation.

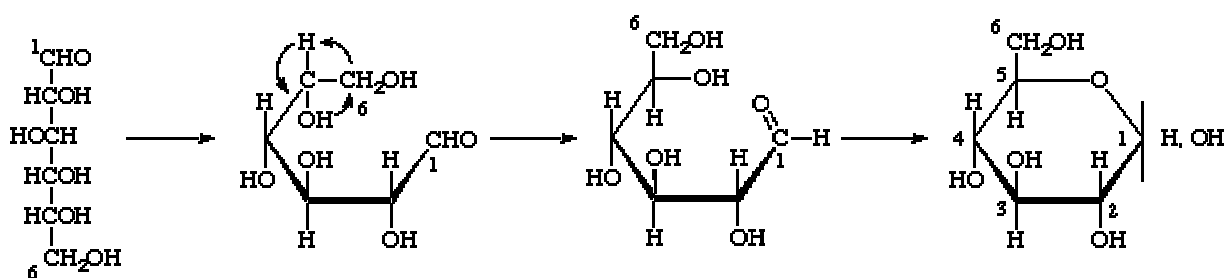


Figure 10 Haworth representation of D-glucopyranose

The orientation of the model described above results in a clockwise numbering of the ring atoms. Groups that appear to the right of the modified Fischer projection appear below the plane of the ring; those on the left appear above. In the common Haworth representation of the pyranose form of D-aldohehexoses, C-6 is above the plane.

The Haworth representation implies a planar ring. However, due to the tetrahedral geometric of carbon backbone, monosaccharides assume conformations that are not planar. The conformation and the spatial arrangements of the ring atoms of a monosaccharide in the cyclic form is well described by the IUPAC-IUB Joint Commission on Biochemical Nomenclatureⁱ and is indicated with a conformational descriptor at the end of the name of the monosaccharide. Figure 11 lists some examples of conformational descriptor within sugar types.

| Type of | Conf. | Atoms of | Above | Belo | Notation | Example |
|---------|-------|----------|-------|------|----------|---------|
|---------|-------|----------|-------|------|----------|---------|

ⁱ Conformational nomenclature for five- and six-membered ring forms of monosaccharides and their derivatives (Recommendations 1980), Eur.J.Biochem., 111, 295-298 (1980); Arch. Biochem. Biophys., 207, 469-472 (1981); Pure Appl. Chem., 53, 1901-1905 (1981); ref. 2, pp. 158-161.

| sugar | reference plane | plane | w plane | | | |
|-------------------------|-----------------|-----------------|---------|---------|------------------|--|
| Aldofuranose (1) | envelope | O-4,C-1,C-3,C-4 | - | C-2 | E ₂ | |
| Aldofuranose (2) | envelope | C-1,C-2,C-4,O-4 | C-3 | - | 3E | |
| Aldofuranose (3) | twist | C-1,O-4,C-4 | C-3 | C-2 | 3T ₂ | |
| Aldofuranose (4) | twist | C-3,C-4,O-4 | C-2 | C-1 | 2T ₁ | |
| Aldopyranose (5) | chair | C-2,C-3,C-5,O-5 | C-4 | C-1 | 4C ₁ | |
| Pyranoid lactone (6) | chair | C-2,C-3,C-5,O-5 | C-1 | C-4 | 1C ₄ | |
| Aldopyranose (7) | boat | O-5,C-1,C-3,C-4 | C-2,C-5 | - | 2,5B | |
| Aldopyranose (8) | boat | C-2,C-3,C-5,O-5 | - | C-1,C-4 | B _{1,4} | |
| Aldopyranose (9) | skew | C-2,C-4,C-5,O-5 | C-1 | C-3 | 1S ₃ | |

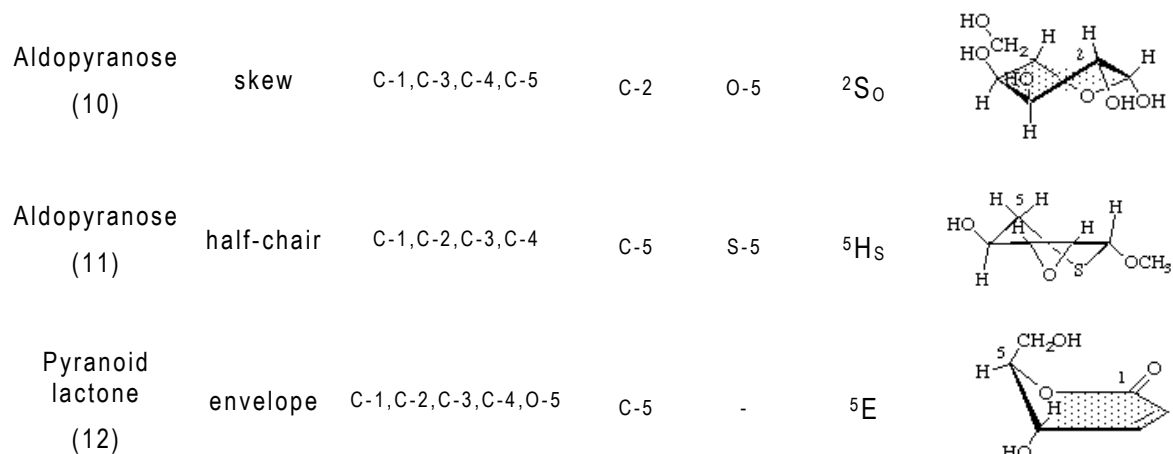


Figure 11 Notation for ring shape. The appropriate letters are as follows. Five-membered rings: E for envelope (examples 1 and 2) and T for twist (examples 3 and 4); six-membered rings: C for chair (examples 5 and 6), B for boat (examples 7 and 8), S for skew (examples 9 and 10), H for half-chair (example 11), and E for envelope (example 12).

The rigorous description of the shape of the six-membered ring can be defined in term of its torsion angles. Again, following IUPAC indicationsⁱⁱ, torsion angles are defined as ν angles (Fig. 12):

| | Hexapyranose (A) | Pentafuranose (B) |
|---------|------------------|-------------------|
| ν_0 | C4-O4-C1-C2 | C5-O5-C1-C2 |
| ν_1 | O4-C1-C2-C3 | O5-C1-C2-C3 |
| ν_2 | C1-C2-C3-C4 | C1-C2-C3-C4 |
| ν_3 | C2-C3-C4-O4 | C2-C3-C4-C5 |
| ν_4 | C3-C4-O4-C1 | C3-C4-C5-O5 |
| ν_5 | - | C4-C5-O5-C1 |

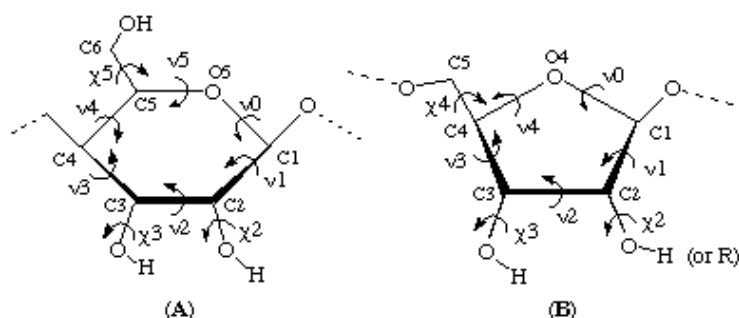


Figure 12. Notation for torsion angles ν for (A) a hexapyranose and (B) a pentafuranose unit. The exocyclic torsion angles χ are also defined.

ⁱⁱ IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN) Conformational nomenclature for five and six-membered ring forms of monosaccharides and their derivatives, Recommendations 1980. Arch. Biochem. Biophys. 207, 469-472 (1981); Eur. J. Biochem. 111, 295-298 (1980); Pure Appl. Chem. 53, 1901-1905 (1981)

It follows that the pyranose ring have 38 basic conformations(Fernandez-Alonso et al. 2005; Stoddart 1971) that have been recently reviewed (Bércecs et al. 2001) in terms of conformational descriptors (Table 1).

Table 1. Endocyclic torsion angles of the 38 ideal basic conformations of pyranose

| conformational descriptor | ν_1 | ν_2 | ν_3 | ν_4 | ν_5 | ν_0 |
|---------------------------|---------|---------|---------|---------|---------|---------|
| 1C_4 | 60 | -60 | 60 | -60 | 60 | -60 |
| 4C_1 | -60 | 60 | -60 | 60 | -60 | 60 |
| ${}^{1,4}B$ | 0 | 60 | -60 | 0 | 60 | -60 |
| $B_{2,5}$ | 60 | 0 | -60 | 60 | 0 | -60 |
| ${}^{0,3}B$ | 60 | -60 | 0 | 60 | -60 | 0 |
| $B_{1,4}$ | 0 | -60 | 60 | 0 | -60 | 60 |
| ${}^{2,5}B$ | -60 | 0 | 60 | -60 | 0 | 60 |
| $B_{0,3}$ | -60 | 60 | 0 | -60 | 60 | 0 |
| 1S_5 | 30 | 30 | -60 | 30 | 30 | -60 |
| 0S_2 | 60 | -30 | -30 | 60 | -30 | -30 |
| 3S_1 | 30 | -60 | 30 | 30 | -60 | 30 |
| 5S_1 | -30 | -30 | 60 | -30 | -30 | 60 |
| 2S_0 | -60 | 30 | 30 | -60 | 30 | 30 |
| 1S_3 | -30 | 60 | -30 | -30 | 60 | -30 |
| 1H_2 | 45 | -15 | 0 | -15 | 45 | -60.2 |
| 3H_2 | 60 | -45 | 15 | 0 | 15 | -45 |
| 3H_4 | 45 | -60 | 45 | -15 | 0 | -15 |
| 5H_4 | 15 | -45 | 60 | -45 | 15 | 0 |
| 5H_0 | 0 | -15 | 45 | -60 | 45 | -15 |
| 1H_0 | 15 | 0 | 15 | -45 | 60 | -45 |
| 4H_5 | -15 | 45 | -60 | 45 | -15 | 0 |
| 0H_5 | 0 | 15 | -45 | 60 | -45 | 15 |
| 0H_1 | -15 | 0 | -15 | 45 | -60 | 45 |
| 2H_1 | -45 | 15 | 0 | 15 | -45 | 60 |
| 2H_3 | -60 | 45 | -15 | 0 | -15 | 45 |
| 4H_3 | -45 | 60 | -45 | 15 | 0 | 15 |
| 1E | 30 | 0 | 0 | -30 | 60 | -60 |
| E_2 | 60 | -30 | 0 | 0 | 30 | -60 |
| 3E | 60 | -60 | 30 | 0 | 0 | -30 |
| E_4 | 30 | -60 | 60 | -30 | 0 | 0 |
| 5E | 0 | -30 | 60 | -60 | 30 | 0 |
| E_0 | 0 | 0 | 30 | -60 | 60 | -30 |
| 4E | -30 | 60 | -60 | 30 | 0 | 0 |
| E_5 | 0 | 30 | -60 | 60 | -30 | 0 |
| 0E | 0 | 0 | -30 | 60 | -60 | 30 |
| E_1 | -30 | 0 | 0 | 30 | -60 | 60 |
| 2E | -60 | 30 | 0 | 0 | -30 | 60 |
| E_3 | -60 | 60 | -30 | 0 | 0 | 30 |

Regarding ring substituents, it should designate as being axial or equatorial in a given conformation. For the precise specification of the orientation, it is necessary to specify the torsion angle about the exocyclic bond. The exocyclic torsion angle is denoted by χ (chi), followed by the atoms to which it refers. e.g. χ (C1-C2-O2-H). or, if no ambiguity arises, simply by χ_2 . The reference atom in the ring is the carbon atom with the number one lower than that of the substituted carbon; e.g. for χ_5 substituted carbon is C5 and reference atom is C4 (see Fig. 13).

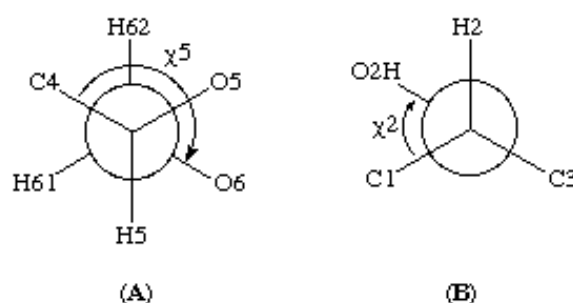


Figure 13. Newman projections showing the exocyclic torsion angles χ_5 (A) and χ_2 (B).

Of particular interest is the relative position of the hydroxymethyl group in position 6 due to the fact that different populations for the three possible rotamers (see Fig. 14) are strictly correlated with the global structure of the sugar and hence its properties (Asensio et al. 1995; Bouwstra et al. 1990; Kirschner et al. 2001; Leeflang et al. 1990) According to IUPAC convention describe here above, rotation around C5-C6 bond could be described by χ_5 torsion angle, even if the use of dihedral angle ω , as will be described in the next section, is generally employed.

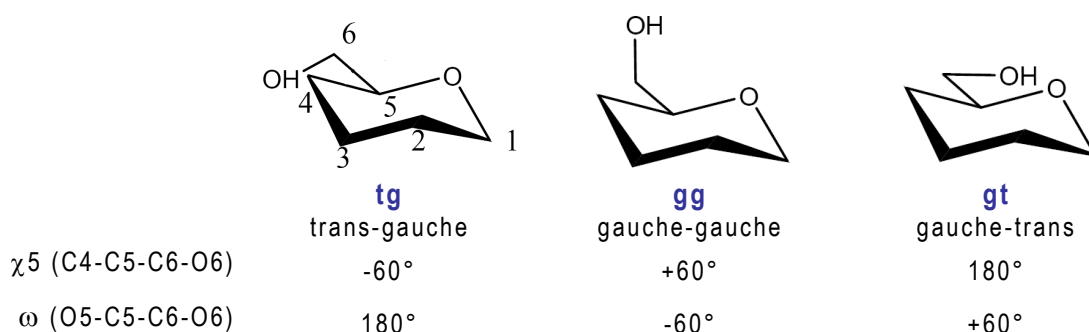


Figure 14. Preferred orientation of the hydroxymethyl group OH6 in the pyranose ring

In the nomenclature introduced in Fig. 14 (tg, gg and gt), the first letter define the relative position of the endocyclic oxygen respect the exocyclic one (O5 and O6 respectively); whereas the second letter define the relative position of C4 with respect to O6. In general, galactose moiety adopts preferently gauche-trans and trans-gauche orientations, due to the steric hindrance that occurs between OH4 and OH6 in the gg rotamer. For the same reason preferred orientation in glucose are the gauche-gauche and the gauche-trans.

With respect to the secondary OH groups, they tend to adopt those positions that allow them to create the maximum number of hydrogen bonds. In that sense, the possible orientation of OHs is reduced to two possibilities: "c", clockwise orientation and "r", reverse-clockwise orientation (Fig. 15).

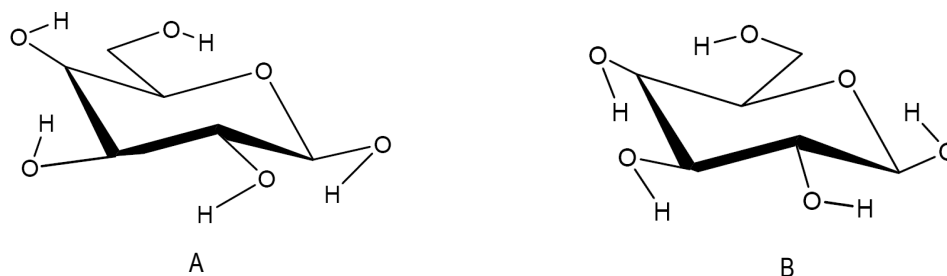


Figure 15. Possible orientations of hydroxyl groups in the pyranose ring. (A) clockwise and (B) reverse clockwise.

Finally just a brief mention to the concept of α and β anomerism. α is used to denote the anomer where the absolute stereochemistry of the anomeric position and the most remote stereocenter in the sugar chain are the same; β is used for the case where they have opposite configurations (Fig. 16).

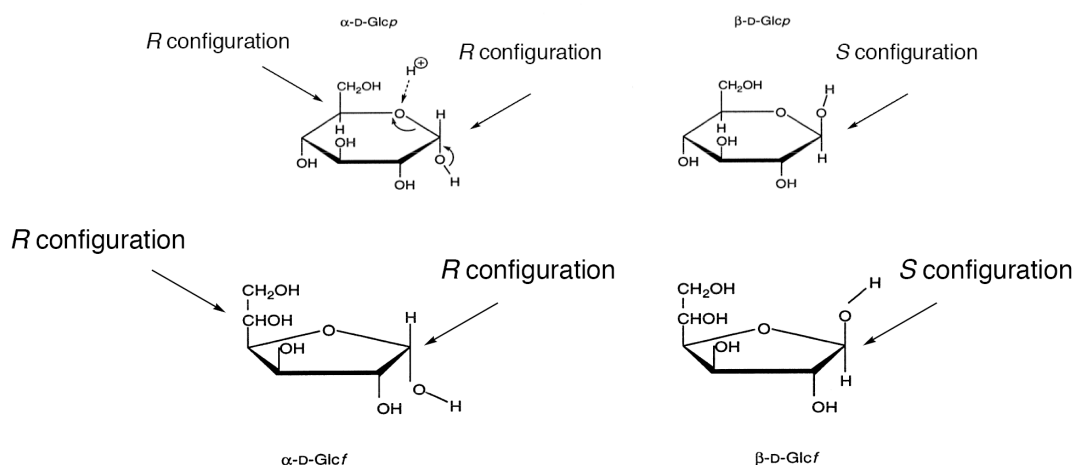


Figure 16. α and β anomerism in pyranose and furanose galactose rings.

Common assumption is that α = axial, which is correct for D-pyranose ring, but it is particularly misleading in the case of furanose sugars: in 5 membered rings, conformational preferences are often subtle and the term "axial" can be ambiguous (Fig. 16).

Glycosidic and aglyconic linkages

When the number of sugar rings is increased, a new variable, the glycoside bond, and consequently a new flexibility degree has to be introduced. The orientations of the glycosidic bonds are specified by the corresponding dihedral angles Φ e Ψ (respectively H1-C1-O1-Cx and C1-O1-Cx-Hx with x being the aglyconic linkage position, see Fig. 17).

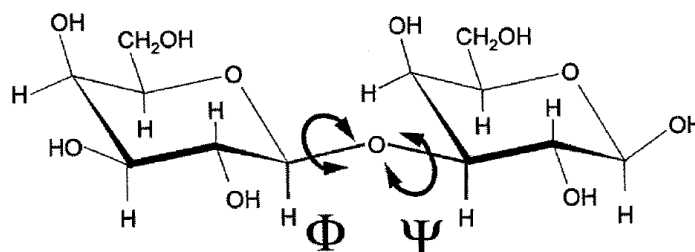


Figure 17. Depiction of the main source of conformational flexibility of the disaccharide Gal β 1 \rightarrow 3Gal by independent rotations about the two dihedral angles ϕ and ψ of the glycosidic bond.

When the glycosidic bond does not involve a carbon atom located in the ring, but on a side chain, the ω angle is introduced. In the case of 1 \rightarrow 6 linkage, ω is defined as O5-C5-C6-O6 (Fig. 18).

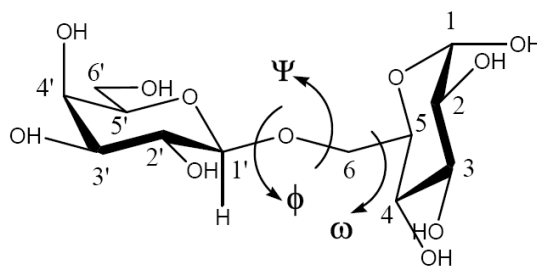


Figure 18. The three torsion angles that define a 1 \rightarrow 6 linkage

The anomeric effects

Anomeric and exoanomeric effects are two important factors that determine the three dimensional structures of carbohydrates (REF LEMIEUX). Both effects refer to the preferred orientation of the substituents at anomeric carbon. For the anomeric effectⁱⁱⁱ (Altona et al. 1963; Epiotis et al. 1976; Juaristi et al. 1995; Romers et al. 1969) an axial configuration of the electronegative substituent at C1 is preferred (Fig. 19A). In fact, this configuration allows an hyperconjugative delocalization between the non bonding electron pair of the endocyclic oxygen and the σ^* orbital of C1-O1 bond as illustrated in Figure 19A.

The exo-anomeric effect, introduced by Lemieux in 1979 (Lemieux et al. 1979), defined the preferred orientation around the interglycosidic linkage and hence defines the most populated dihedral angles. As illustrated in figure 19B, a favorable interaction occurs between the electron lone pair of the exocyclic oxygen bond to C1 and orbital σ^* of endocyclic C1-O5.

ⁱⁱⁱ IUPAC "Compendium of Chemical Terminology", 2nd Edition, 1997

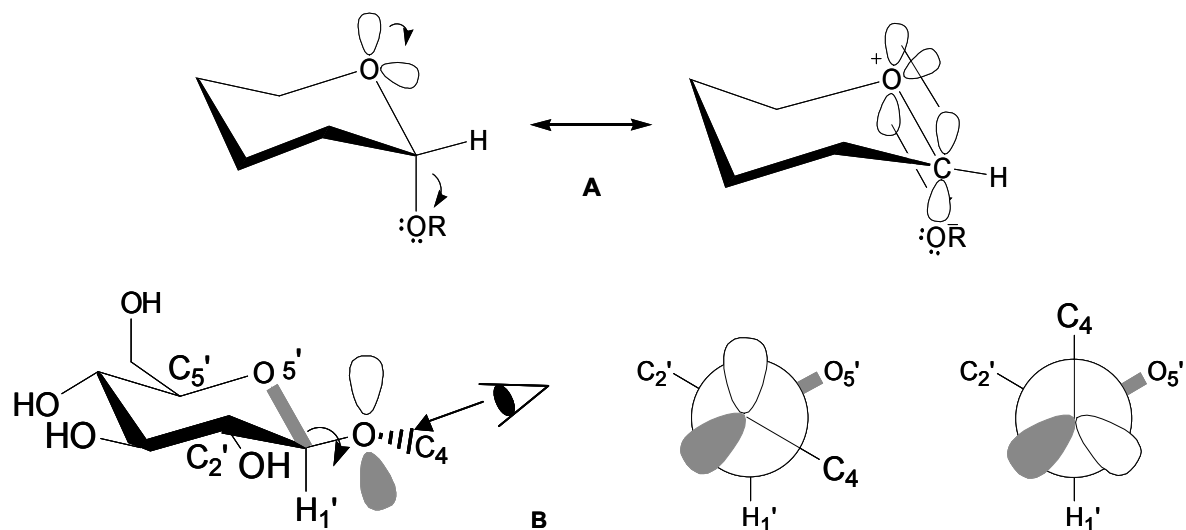


Figure 19. Anomeric and exoanomeric effects. (A) Schematic representation of interactions responsible of axial-preferred configuration; (B) exoanomeric effect for a $\beta(1\rightarrow4)$ disaccharide.

Symbol Representation for Glycans

The Consortium for Functional Glycomics, NLM/NCBI and The Glycobiology Research and Training Center are developing a new nomenclature for glycans that better interact with glycobiology. Since numerous papers have been already published following this new symbolism, it would be interesting to point out the general principle and the essential symbols employed (Table 2) (Varki 1999)^{iv}.

The aim of this new nomenclature system could be defined with the following points:



































- Avoid same shape/color but different orientation to represent different sugars, since annotation of mass spectra is hard to show as horizontal cartoons.
- Choice of symbols should be logical and simple to remember.
- Each monosaccharide type (e.g. hexose) should have the same shape, and isomers are to be differentiated by color/black/white/shading.
- Use same shading/color for different monosaccharides of same stereochemical designation e.g., Gal, GalNAc, GalA should all be the same shading/color.
- To minimize variations, sialic acids and uronic acids are same shape. Only major uronic and sialic acid types are represented.
- While color is useful, the system should also function with black and white, and colored representations should survive black-and white printing or Xeroxing.

^{iv} and webpage <http://grtc.ucsd.edu/>

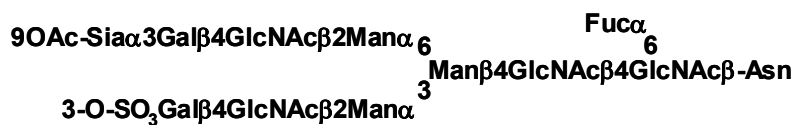
- Positions of linkage origin are assumed to be common ones unless indicated. Anomeric notation and destination linkage indicated without spacing/dashes.
- Modifications of monosaccharides (e.g., sulfation, O-acetylation) indicated by attached small letters, with numbers indicating linkage positions, if known.
- Only common monosaccharides in vertebrate systems are assigned a specific symbol. All other monosaccharides are represented by an open Hexagon, and defined in the figure legend. If there is more than one type of undesignated monosaccharide in a figure, an letter designation internal to the Hexagon can be included to differentiate between them - again, specified in the figure legend.

Symbol Representations

Table 2. Essential second edition symbols. Color (in parenthesis the corresponding RGB values are indicated) and Black&White representation.

| LEGEND | | SYMBOL | | | | |
|---|---|---|---|---|---|---|
| Hexoses: Circles N-Acetylhexosamines: Squares Hexosamines: Squares divided diagonally | | RGB colors | Black&White | | | |
| Galactose stereochemistry: YELLOW (255,255,0) with Black outline |  |  |  |  |  |  |
| Glucose stereochemistry: BLUE (0,0,250) with Black outline |  |  |  |  |  |  |
| Mannose stereochemistry: GREEN (0,200,50) with Black outline |  |  |  |  |  |  |
| Fucose: RED (250,0,0) with Black outline | |  | |  | | |
| Xylose: (5-pointed star) ORANGE (250,100,0) with Black outline | |  | |  | | |
| <u>Acidic Sugars (Diamonds)</u> | | | | | | |
| Neu5Ac: PURPLE (125,0,125) with Black outline | |  | |  | | |
| Neu5Gc: LIGHT BLUE (200,250,250) with Black outline | |  | |  | | |
| KDN: GREEN (0,200,50) with Pattern & Black outline | |  | |  | | |
| GlcA: BLUE (0,0,250) /Upper segment with Black outline | |  | |  | | |
| IdoA: TAN (150,100,50) /Lower segment with Black outline | |  | |  | | |
| GalA: RED (250,0,0) /Left segment with Black outline | |  | |  | | |

SIMPLIFIED REPRESENTATION (IUB-IUPAC JCBN Recommendations, J.Biol.Chem. 262:13-18, 1985)



SYMBOLIC REPRESENTATIONS

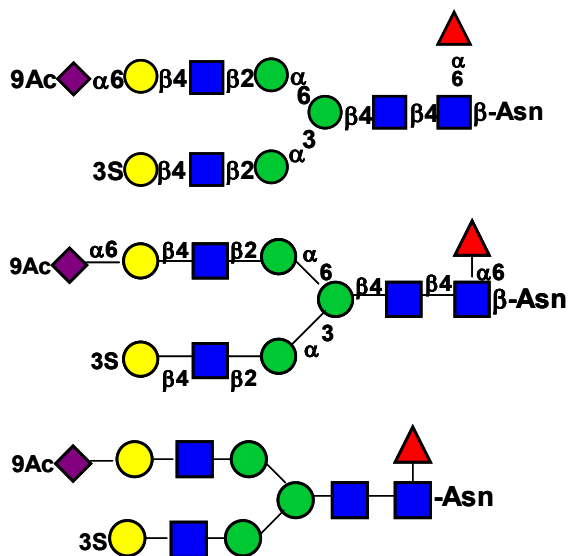


Figure 20. Different representations of the same oligosaccharide.

References

- Abeygunawardana, C. and C. A. Bush (1993) "Determination of the chemical structure of complex polysaccharides by heteronuclear NMR spectroscopy" *Adv. Biophys. Chem.* **3**: 199-249.
- Altona, C., C. Knobler and C. Romers (1963) "The conformation of non-aromatic ring compounds. VII. Crystal structure of trans-2,5-dichloro-1,4-dioxane at 125°C" *Acta Cryst.* **16**: 1217-1225.
- Asensio, J. L. and J. Jimenez-Barbero (1995) "The use of the AMBER force field in conformational analysis of carbohydrate molecules: determination of the solution conformation of methyl alpha-lactoside by NMR spectroscopy, assisted by molecular mechanics and dynamics calculations" *Biopolymers* **35**(1): 55-73.
- Bérces, A., D. M. Whitfield and T. Nukada (2001) "Quantitative description of six-membered ring conformations following the IUPAC conformational nomenclature" *Tetrahedron* **57**: 477-491.
- Bouwstra, J. B., E. C. Spoelstra, P. De Waard, B. R. Leeftang, J. P. Kamerling and J. F. Vliegthart (1990) "Conformational studies on the N-linked carbohydrate chain of bromelain" *Eur J Biochem* **190**(1): 113-22.
- Brown, G. M. and H. A. Levy (1965) "Alpha-D-Glucose: Precise Determination of Crystal and Molecular Structure by Neutron-Diffraction Analysis" *Science* **147**: 1038-9.
- Bush, C. A., M. Martin-Pastor and A. Imberty (1999) "Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides" *Annu Rev Biophys Biomol Struct* **28**: 269-93.
- Canales, A., J. Angulo, R. Ojeda, M. Bruix, R. Fayos, R. Lozano, G. Gimenez-Gallego, M. Martin-Lomas, P. M. Nieto and J. Jimenez-Barbero (2005) "Conformational flexibility of a synthetic glycosylaminoglycan bound to a fibroblast growth factor. FGF-1 recognizes both the (1)C(4) and (2)S(O) conformations of a bioactive heparin-like hexasaccharide" *J Am Chem Soc* **127**(16): 5778-9.
- Casu, B., M. Petitou, M. Provasoli and P. Sinay (1988) "Conformational flexibility: a new concept for explaining binding and biological properties of iduronic acid-containing glycosaminoglycans" *Trends Biochem Sci* **13**(6): 221-5.
- Epiotis, N. D., R. L. Yates, F. Bernardi and S. Wolfe (1976) *J Am Chem Soc* **98**: 5435.
- Fernandez-Alonso, M. C., J. Canada, J. Jimenez-Barbero and G. Cuevas (2005) "Theoretical study of inversion and topomerization processes of substituted cyclohexanes: the relevance of the energy 3D hypersurface" *Chemphyschem* **6**(4): 671-80.
- Juaristi, E. and G. Cuevas (1995) *The anomeric effect*, Boca Raton, CRC Press.
- Kirschner, K. N. and R. J. Woods (2001) "Solvent interactions determine carbohydrate conformation" *Proc Natl Acad Sci U S A* **98**(19): 10541-5.
- Leeftang, B. R., J. B. Bouwstra, J. Kerekgyarto, J. P. Kamerling and J. F. Vliegthart (1990) "ROESY studies and HSEA calculations on xylose-containing oligosaccharides related to N-glycoproteins" *Carbohydr Res* **208**: 117-26.
- Lemieux, R. U. and K. Bock (1979) "The conformations of the Lewis blood group determinants, sucrose and kanamycin A" *Jpn J Antibiot* **32 Suppl**: S163-77.
- Mann, P. L. and R. E. Waterman (1998) "Glycocoding as an information management system in embryonic development" *Acta Anat (Basel)* **161**(1-4): 153-61.
- Marszalek, P. E., A. F. Oberhauser, Y. P. Pang and J. M. Fernandez (1998) "Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring" *Nature* **396**(6712): 661-4.
- Noorman, F., M. M. Barrett-Bergshoeff and D. C. Rijken (1998) "Role of carbohydrate and protein in the binding of tissue-type plasminogen activator to the human mannose receptor" *Eur J Biochem* **251**(1-2): 107-13.
- Romers, C., C. Altona, H. R. Buys and E. Havinga (1969) *Topics Stereochem.* **4**(39).

- Solis, D., D. Estremera, P. Usobiaga and T. Diaz-Maurino (1987) "Differential binding of mannose-specific lectins to the carbohydrate chains of fibrinogen domains D and E" *Eur J Biochem* **165**(1): 131-8.
- Stoddart, J. F. (1971) *Stereochemistry of Carbohydrates*, New York, John Wiley and Sons.
- Varki, A. (1999) *Essentials of glycobiology*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
- von der Lieth, C., H. Siebert, T. Kozar, M. Burchert, M. Frank, M. Gilleron, H. Kaltner, G. Kayser, E. Tajkhorshid, N. V. Bovin, J. F. Vliegthart and H. Gabius (1998) "Lectin ligands: new insights into their conformations and their dynamic behavior and the discovery of conformer selection by lectins" *Acta Anat (Basel)* **161**(1-4): 91-109.

Investigations of sugar conformations: the Free State

Like polypeptides and polynucleotides, complex oligosaccharides may adopt three-dimensional (3D) conformations that are relatively fixed, or that may exist as several conformations in equilibrium, thereby leading to a flexible model. The monosaccharides, in all mammalian glycoproteins and glycolipids, as well as almost all the bacterial polysaccharides, are composed of pyranosides. Homonuclear NMR coupling constant data show that for the great majority of cases, these six-membered rings adopt fixed chair conformations, a 4C_1 chair for D sugars and a 1C_4 chair for L sugars (Abeygunawardana et al. 1993). For fixed pyranoside ring conformations, the two dihedral angles, ϕ and ψ , about the glycosidic linkage are the main internal coordinates specifying the oligosaccharide conformation. Although this parallels the polypeptide case, the pyranoside rings are larger and bulkier, introducing greater stereochemical restriction, which leads to disaccharide maps with smaller minimum energy space than occurs in dipeptide maps. The furanosides found in nucleic acids are much more flexible. Although our understanding of the conformation of complex carbohydrates remains primitive by comparison with the structures of proteins and nucleic acids, there is no obvious reason why the spectacular success of X-ray crystallography, NMR spectroscopy, and computer molecular modeling cannot be duplicated for this challenging problem. Hopefully appropriate applications of these methods are yielding progress toward the understanding of carbohydrate conformation and interactions.

X-Ray Crystallography

The methodology used for extracting structural data from crystalline carbohydrates depends on the size of the molecule and its ability to form crystals. For compounds as small as disaccharides or oligosaccharides, crystallization is the main difficulty, although the cause is not obvious. It could result in part from the high polarity of the hydroxyl groups, from the flexibility of the carbohydrate linkage, or simply from the lack of pure samples in sufficient quantity and quality. As a result, only about 50 structures of disaccharides are available in the Cambridge Structural Database. For larger oligosaccharides, fewer than 20 structures have been solved (excluding cyclodextrins), most of them sucrose containing molecules. In fact, only two trisaccharides that are part of complex glycans have been crystallized, i.e. the $\alpha\text{Man}(1\rightarrow3)\beta\text{Man}(1\rightarrow4)\text{GlcNAc}$ (Perez et al. 1996; Warin et al. 1979) and the Lewis-X histo—blood group antigen (Perez et al. 1996). The information that can be extracted from crystal structures of di- and oligosaccharides includes, not only the conformational features, but also the intermolecular interaction. For example, networks of hydrogen bonding that involve one or two water molecules have been extensively studied by Jeffrey & Saenger (Jeffrey et al. 1991). In other cases, the analysis of the crystal packing can be directly compared with biological data. In the Lewis-X crystal structure (Perez et al. 1996), some interactions observed between adjacent rows of trisaccharide can provide the basis for the previously postulated involvement of Lex-Lex interaction in cell-cell adhesion (Eggens et al. 1989). Although oligosaccharides are difficult to crystallize in their native state, they can be “trapped” in protein crystals, either covalently linked to the peptide chain (glycoproteins) or as

ligands to the protein. More than 100 of the protein structures available from the Protein Data Bank (Bernstein et al. 1977) are N-glycosylated. The conformational features of the GlcNAc-Asparagine linkage have been analyzed from the structures solved at high resolution (Imberty et al. 1995). Because the glycan moiety is often poorly resolved, only one or two carbohydrate residues can generally be seen in glycoprotein crystal structures. More information can be obtained from complexes between protein and carbohydrate ligands, as discussed below. Because it is impossible to obtain single crystals from polysaccharides suitable for X-ray analysis, different strategies have been used for diffractometry. One possibility is to stretch the polysaccharide into an oriented fiber. Since the beginning of X-ray fiber diffraction, in the early 1950s, more than 200 biological and synthetic polymer helices have been successfully investigated, among them over 50 polysaccharides (Chandrasekaran 1997). This includes the major plant polysaccharides, amylose and cellulose, but also many macromolecules from animal and bacterial origin. An alternate strategy is to use electron diffraction for studying very small crystals, needles, or platelets that can be obtained from polysaccharides (Perez et al. 1989). In both X-ray fiber diffraction and electron diffraction, the amount of data collected is small, and building a model by molecular mechanics is necessary for the resolution of the 3D structure.

NMR Methods for Structure Determination

During the past decades, structural assignment of complex carbohydrate structures by NMR spectroscopy has demonstrated its vast importance and this is expected to continue in the future. As already mentioned, building blocks in oligosaccharides are more diverse in Nature than in proteins or nucleic acids. Carbohydrates often differ only from each other in the stereochemistry, and the pattern of interresidue linkages can be very heterogeneous (Rudiger et al. 2000). The information capacity of the carbohydrates is much larger than in proteins, particularly due to branched structures. Consequently, the structure determination of complex oligosaccharides is a challenging problem. A wealth of techniques are currently being applied in the identification of known oligosaccharides or the determination of new structures; e.g. capillary electrophoresis (El Rassi et al. 1997), fluorescence detected HPLC (Anumula et al. 1998) and mass spectrometry (Ahn et al. 1999; Sato et al. 2000). However, sugar stereochemistry cannot be solved by current methodology used routinely in mass spectrometry and for NMR spectroscopy represented a new challenge (Jimenez-Barbero et al. 2003).

Assignment Methods

In this Thesis, the basis and the fundamentals of NMR technique will not be discussed. I will focus on the methodologies that address to the structure determination for a sugar molecule.

There are several ways to perform a primary structural analysis of a mono-, oligo-, or polysaccharide by NMR spectroscopy and a good starting point is the anomeric proton chemical shift. Integration of the anomeric resonances offers an initial estimate on the number of different monosaccharide residues present. The anomeric proton resonances are found in the shift range 4.4-5.5 ppm. The remaining ring proton resonances are found in the range 3-4.2 ppm in unprotected

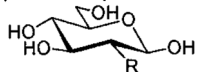
oligosaccharides. Additionally, the number of anomeric C₁ resonances present in a 1D ¹³C NMR spectrum will confirm the number of different residues. Such results can also be obtained from 2D ¹³C-¹H HSQC (Bodenhausen et al. 1980) HMQC (Bax et al. 1983; Mueller 1979) or HMBC (Bax et al. 1986) spectra, which in many cases are more sensitive than a 1D ¹³C spectrum. Homonuclear TOCSY and DQF-COSY spectra are useful in the identification of individual monosaccharide residues. In TOCSY spectra of oligosaccharides acquired with a fairly long mixing time (>100 ms), it is often possible to measure the size of the coupling constants and the correlations to reveal the identity of the residue. In cases with significant overlap in the bulk region (3-4.2 ppm), a 1D TOCSY may be useful in resolving ambiguities. In Table 2 the most common sugars are listed with their NMR characteristics and coupling pattern, and NMR spectra of selected monosaccharide residues are shown. Both ¹H and ¹³C chemical shifts for most monosaccharides can be found in the literature (Tables 2 and 3), and based on such values, an assignment of the individual residues can be made. Relevant database are also available online like CarbBank and SUGABASE* database (Doubet et al. 1989; van Kuik et al. 1992; van Kuik et al. 1992) or CASPER (Hermansson et al. 1992; Jansson et al. 1991; Jansson et al. 1991) where both prediction and attendance in the assignment are available for the main oligosaccharides and polysaccharides.

The ¹³C chemical shift values can easily be obtained from a HSQC or HMQC spectrum. The TOCSY and HSQC (or HMQC) data may also be obtained simultaneously using the 2D version of the HSQC-TOCSY (de Beer et al. 1994; Kover et al. 1997) or the HMQC-TOCSY (Lerner et al. 1986) experiments. These experiments are useful and give additional dispersion in the carbon dimension, which may facilitate the assignment of individual spin systems. For NeuAc derivatives (see Table 2), without an anomeric proton, characteristic signals as the H3eq or H3ax protons are a good starting point for the assignments. ¹³C chemical shifts are often obtained from multidimensional heteronuclear experiments using an inverse detection probe, because the 1D carbon-observed experiment requires more compound. With the introduction of pulse-field gradients, many of the NMR experiments have gained additional sensitivity, because the gradient-enhanced experiments give fewer spectral artifacts, allow for better solvent suppression, and require a shorter phase cycle (Keeler et al. 1994). Many 1D, 2D, 3D, and 4D homonuclear and heteronuclear experiments have been adapted to gradient versions (Keeler et al. 1994)

* available at <http://www.boc.chem.uu.nl/sugabase/carbbank.html>

Table 2. Characteristic chemical shift, J-coupling pattern and reference spectrum of the most common monosaccharides

β -Glc and β -GlcNAc



R = OH, NHAc

Glc: J_{α} (Hz): 3.6, 9.5, 9.5, 9.5¹

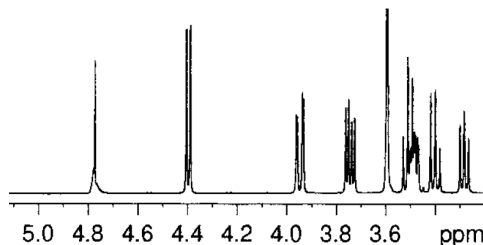
J_{β} (Hz): 7.8, 9.5, 9.5, 9.5¹

GlcNAc: the coupling pattern is similar as Glc

δ NAc~2ppm

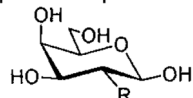
upfield shift:

$\delta C_{2,\alpha}$ ~55.4ppm; $\delta C_{2,\beta}$ ~58ppm^{1,2}



¹H spectrum of Methyl α -D-glucopyranoside

β -Gal and β -GalNAc



R = OH, NHAc

Gal: J_{α} (Hz): 3.8, 10, 3.8, 11

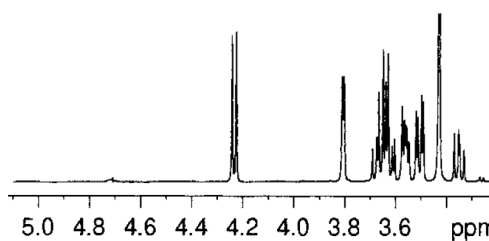
J_{β} (Hz): 8, 10, 3.8, 11

GalNAc: the coupling pattern is similar as Gal

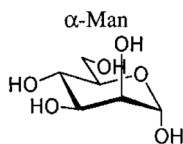
δ NAc~2ppm

upfield shift: $\delta C_{2,\alpha}$ ~51.4ppm;

$\delta C_{2,\beta}$ ~54.9ppm^{1,2}

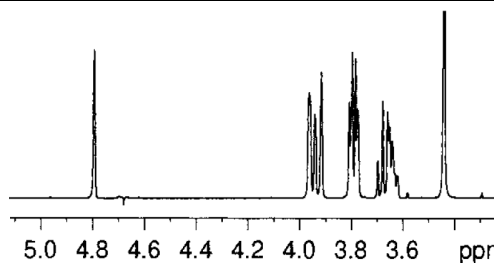


¹H spectrum of Methyl α -D-galactopyranoside

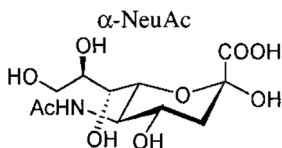


J_{α} (Hz): 1.8, 3.8, 10, 9.8¹

J_{β} (Hz): 1.5, 3.8, 10, 9.8¹



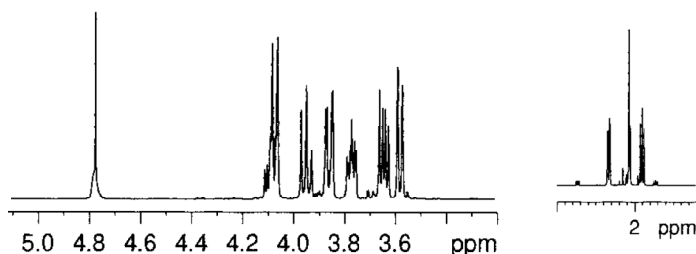
¹H spectrum of Methyl α -D-mannopyranoside



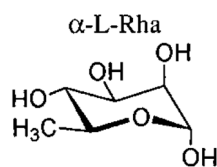
δH_{3ax} ~1.9ppm; δH_{3eq} ~2.3ppm³

upfield shift at δC_3

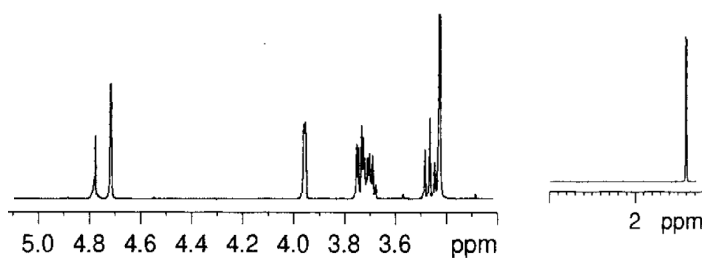
$J_{\alpha,C1-H3ax}$ ~6Hz; $J_{\beta,C1-H3ax}$ <1Hz^{4,5}



¹H spectrum of N-acetylneuraminic acid



$\delta_{C6} \sim 18$ ppm; $\delta_{H6} \sim 1.2$ ppm¹
 Similar coupling pattern to Man,
 $J_{56} = 6.2$ Hz¹



¹H spectrum of Methyl α -L-rhamnoside

1. Bock, K.; Thøgersen, H. *Annu. Rep. NMR Spectrosc.* 1982, 13, 1-57
2. Bock, K.; Pedersen, C. *Adv. Carbohydr. Chem. Biochem.* 1983, 41, 27-66
3. Prytulla, S.; Lambert, J.; Lauterwein, J.; Klessinger, M.; Thiem, J. *Magn. Reson. Chem.* 1990, 28, 888-901
4. Haverkamp, J.; Spoomaker, T.; Dorland, L.; Vliegthart, J. F. G.; Schauer, R. *J. Am. Chem. Soc.* 1979, 101, 4851-4853.
5. Prytulla, S.; Lauterwein, J.; Klessinger, M.; Thiem, J. *Carbohydr. Res.* 1991, 215, 345-349

A general summary of chemical shift tables collections appeared in the last years is presented in Table 3.

Table 3. Chemical Shift Table Collections

| author | title | year |
|---|---|--------------------|
| De Bruyn, A.; Anteunis, M.; Verhegge, G. | ¹ H-NMR study of the diglucopyranoses in D ₂ O | 1975 ⁹¹ |
| Bock, K.; Thøgersen, H. | Nuclear magnetic resonance spectroscopy in the study of mono- and oligosaccharides | 1982 ⁷⁵ |
| Bock, K.; Pedersen, C. | Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides | 1983 ⁷⁶ |
| Vliegthart, J. F. G.; Dorland, L.; van Halbeek, H. | High resolution, ¹ H-nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins | 1983 ⁵⁶ |
| Bock, K.; Pedersen, C.; Pedersen, H. | Carbon-13 nuclear magnetic resonance data for oligosaccharides. | 1984 ⁹² |
| Shashkov, A. S.; Nifant'ev, N. E.; Amochaeva, V. Y.; Kochetkov, N. K. | ¹ H and ¹³ C NMR data for 2- <i>O</i> -, 3- <i>O</i> - and 2,3-Di- <i>O</i> -glycosylated methyl α - and β -D-glucopyranosides and β -D-galactopyranosides | 1993 ⁹³ |
| Bock, K.; Duus, J. Ø. | A conformational study of hydroxymethyl groups in carbohydrates investigated by ¹ H NMR spectroscopy | 1994 ⁹⁴ |
| Hounsell, E. F. | ¹ H NMR in the structural and conformational analysis of oligosaccharides and glycoconjugates | 1995 ⁹⁶ |
| Hobley, P.; Howarth, O.; Ibbett, R. N. | ¹ H and ¹³ C NMR shifts for aldopyranose and aldofuranose monosaccharides: Conformational analysis and solvent dependence | 1996 ⁹⁵ |
| Khatuntseva, E. A.; Shashkov, A. S.; Nifant'ev, N. E. | ¹ H and ¹³ C NMR data for 3- <i>O</i> -, 4- <i>O</i> - and 3,4-di- <i>O</i> -glycosylated methyl α -L-rhamnopyranosides | 1997 ⁹⁶ |

The proton and carbon chemical shifts are sensitive to the attachment of a non-carbohydrate group like a methyl, acetyl, sulfate, or a phosphate group. Attachment of such a group will affect the proton and carbon resonances where the group is located. Normally downfield shifts ~ 0.2 - 0.5 ppm are observed (Van Halbeek 1996) for protons and higher $\Delta\delta$ values for ¹³C. This places these resonances in a less crowded area of the spectra and helps the identification of modified residues. Such appended groups may also contain NMR-active nuclei, which may give rise to additional splittings due to couplings (e.g., ³¹P-¹H long-range couplings). The use of other homo- or heteronuclear correlations may help in the determination of their position.

Scalar Coupling Constants

One feature that distinguishes the structural analysis of oligosaccharides by NMR from biomolecules such as proteins is that, in general, the number and distribution of the NOEs obtained from conventional experiments is not sufficient to fully characterize the solution conformation with reasonable certainty (Homans 1990), especially when flexibility is present among the glycosidic linkages (Xu et al. 1996). If the interpretation of interresidue NOE data in terms of a unique structure is impossible, a combination of conformations, or a virtual conformation is required (Cumming et al. 1987). For these situations, NOE data generally must be combined with other sources of structural constraints, such as those obtained from scalar coupling constants. In flexible structures, scalar coupling constants have the advantage that calculation of average values over an ensemble of conformations is simpler than for NOE. In addition to the $1/r^6$ dependence on distance, NOE has a complex time dependence, in which the overall molecular tumbling may interact with kinetics of the conformational exchange involved in the internal motion. The scalar coupling values for a flexible structure are simple linear averages over the ensemble of individual conformers. Recent developments in multidimensional NMR have made possible a number of new methods for measurement of the required heteronuclear coupling constants. The poor sensitivity characteristic of ^{13}C detection methods in natural abundance can be improved by such ^1H detection methods as heteronuclear multiple bond correlation (HMBC) (Pope et al. 1991; Pope et al. 1991). Recently, the access to ^{13}C enriched carbohydrates (Jones et al. 1989) has allowed the use of new 2D and 3D NMR experiments for the determination of long-range H-C coupling constants, such as HMQC-NOESY (Xu et al. 1996) and HMQC-TOCSY (Gitti et al. 1994). With highly enriched oligosaccharides and polysaccharides, it is also possible to measure long-range C-C coupling constants that provide additional information on the glycosidic dihedral angles (Martin-Pastor et al. 2003; Milton et al. 1998; Xu et al. 1998).

In general, homonuclear $^3J_{\text{HH}}$ coupling constants characterize the disposition of axial and equatorial substituents in a pyranose and furanose residue (Tab. 2). For instance, vicinal coupling constant between the anomeric H1 and the H2 indicates the relative orientation of the two protons: If they are both in an axial configuration in pyranose structures, a large coupling constant (7-8 Hz) is observed, whereas if they are equatorial-axial, this is smaller ($J_{1,2} \sim 4$ Hz), and for axial-equatorial or equatorial-equatorial oriented protons, even smaller coupling constants are observed (< 2 Hz) (Jansson et al. 1987) used when assigning the relative orientation of protons in a hexopyranose ring as first demonstrated by Lemieux et al. in 1958 (Lemieux et al. 1958).

In addition, the homonuclear $^3J_{\text{H5-H6proR}}$ and $^3J_{\text{H5-H6proS}}$ couplings are related to the ω torsion angle (defined by atoms O5-C5-C6-O6) of the hydroxymethyl group in pyranose rings. These couplings are usually interpreted in terms of a combination of the three possible staggered conformations, gauche-gauche ($\omega = -60^\circ$), gauche-trans ($\omega = 60^\circ$), and trans-gauche ($\omega = 180^\circ$). The recent use of experiments such as HCCH-COSY or HCCH-TOCSY in ^{13}C -enriched carbohydrates allows accurate measurements of these homonuclear coupling constants, which are significant to the overall conformation of oligosaccharides that have a $1 \rightarrow 6$ linkage (van Halbeek 1994). But most importantly,

the one bond ^{13}C - ^1H coupling constants in pyranoses can be used to determine the anomeric configuration unequivocally (Bock et al. 1974). For D sugars in the $^4\text{C}_1$ conformation, a $^1J_{\text{C}_1, \text{H}_1} \sim 170$ Hz indicates an α -anomeric sugar configuration whereas $^1J_{\text{C}_1, \text{H}_1} \sim 160$ Hz indicates a β -anomeric sugar configuration (Bock et al. 1974). This is reversed for L sugars. The use of one-bond coupling constants in furanose structures does not correlate in the same way with the anomeric structure. For sugars such as NeuAc with no anomeric proton, the anomeric configuration can be obtained from measurements of the carbonyl-H3eq/H3ax coupling constants (Haverkamp et al. 1979; Prytulla et al. 1991) the C2-H3 coupling constants (Prytulla et al. 1990), or from chemical shifts of the H4 and H6 protons, respectively. Also, heteronuclear coupling constants are extremely important to deduce the relative disposition of the monomers within the chain and hence the geometry of the glycosidic linkage in term of its dihedral angles (Cano et al. 1987).

The interpretation of heteronuclear coupling constants $^3J_{\text{COCH}}$ across a glycosidic linkage requires a Karplus relationship to correlate the data with the ϕ and ψ glycosidic torsion angles. The Equation 1 (Tvaroska et al. 1989) similar to one proposed by Mulloy et al (Mulloy et al. 1988),

$$^3J_{\text{COCH}}(\theta) = 5.7 \cos^2(\theta) - 0.6 \cos(\theta) + 0.5$$

represents the correlation of the coupling constants $^3J_{\text{COCH}}$ across the glycosidic linkage which allows to deduce structural constraints. Nevertheless, important restraints for the ϕ and ψ torsion angles can also be deduced from the the vicinal coupling constants, $^3J_{\text{C}_2\text{C}_1\text{O}_1\text{C}_x}$, $^3J_{\text{C}_1\text{O}_1\text{C}_x\text{C}_{x-1}}$, and $^3J_{\text{C}_1\text{O}_1\text{C}_x\text{C}_{x+1}}$, and a considerable effort has been dedicated to this topic (Milton et al. 1998).

Nuclear Overhauser Effect: theory and related experiments

Consider two spins, **I** and **S** which are not coupled. The two spins have between them four energy levels, which can be labelled according to the spin states of the two spins (Fig. 21).

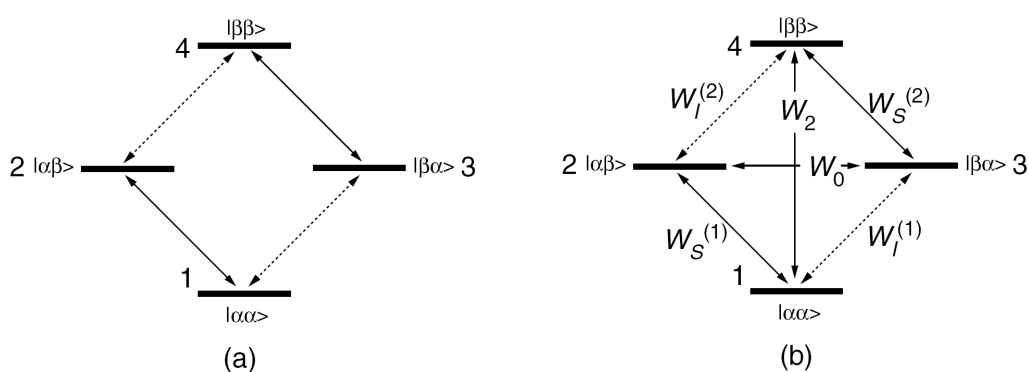


Figure 21. (a) shows the energy levels of a two spin system; the levels are labelled with the spin of **I** first and the spin of **S** second. The dashed arrows indicate allowed transitions of the **I** spin, and the solid arrows indicate allowed transitions of the **S** spin. Diagram (b) shows the relaxation induced transitions which are possible amongst the same set of levels.

It turns out that in such a system it is possible to have relaxation induced transitions between all possible pairs of energy levels, even those transitions which are forbidden in normal spectroscopy. The rate constants for the two allowed **I** spin transitions will be denoted $W_1^{(1)}$ and $W_1^{(2)}$, and likewise for the spin **S** transitions. The rate constant for the transition between the $\alpha\alpha$ and $\beta\beta$ states is denoted W_2 , that it is a double quantum transition. Finally, the rate constant for the transition between the $\alpha\beta$ and $\beta\alpha$ states is denoted W_0 , a zero quantum transition (Fig. 21).

The basis of the nuclear Overhauser effect can readily be seen from the Solomon equation:

$$\frac{d(I_z - I_z^0)}{dt} = -R_I(I_z - I_z^0) - \sigma_{IS}(S_z - S_z^0) \quad \text{[equation 1]}$$

in which it is clear that if the **S** spin magnetization deviates from equilibrium there will be a change in the **I** spin magnetization at a rate proportional to the cross-relaxation rate, σ_{IS} , and to the extent of the deviation of the **S** spin from equilibrium. This change in the **I** spin magnetization will manifest itself as a change in the intensity in the corresponding spectrum, and this change in intensity of the **I** spin when the **S** spin is perturbed is denominated the nuclear Overhauser effect. Plainly, there will be no such effect unless σ_{IS} is non-zero, which requires the presence of the W_2 and W_0 relaxation pathways. Such pathways are only present when there is dipolar relaxation between the two spins and that the resulting cross-relaxation rate constants have a strong dependence on the distance between the two spins. The observation of a nuclear Overhauser effect is therefore diagnostic of dipolar relaxation and hence the proximity of pairs of spins. The effect is of enormous value, therefore, in structure determination by NMR.

A simple experiment, which reveals the NOE, is to invert just the **S**, spin by applying a selective 180° pulse to its resonance (the so called transient NOE experiment). The **S** spin is then not at equilibrium so magnetization is transferred to **I** spin by cross-relaxation. After a suitable period, called the mixing time, τ_m , a non-selective 90° pulse is applied and the spectrum recorded.

After the selective pulse the situation is

$$I_z(0) = I_z^0 \quad S_z(0) = -S_z^0 \quad \text{[equation 2]}$$

where I_z has been written as $I_z(t)$ to emphasize that it depends on time and likewise for **S**. To work out what will happen during the mixing time the differential equations

$$\begin{aligned} \frac{dI_z(t)}{dt} &= -R_I(I_z(t) - I_z^0) - \sigma_{IS}(S_z(t) - S_z^0) \\ \frac{dS_z(t)}{dt} &= -\sigma_{IS}(I_z(t) - I_z^0) - R_S(S_z(t) - S_z^0) \end{aligned} \quad \text{[equations 3]}$$

need to be solved (integrated) with this initial condition. One simple way to do this is to use the initial rate approximation. This involves assuming that the mixing time is sufficiently short that it can be assumed that the initial conditions set out in Eq. [2] apply; so, for the first equation

$$\begin{aligned}\frac{dI_z(t)}{dt} \Big|_{\text{init}} &= -R_I(I_z^0 - I_z^0) - \sigma_{IS}(-S_z^0 - S_z^0) \\ &= 2\sigma_{IS}S_z^0\end{aligned}\quad \text{[equation 4]}$$

This is now easy to integrate as the right-hand side has no dependence on $I_z(t)$

$$\begin{aligned}\int_0^{\tau_m} dI_z(t) &= \int_0^{\tau_m} 2\sigma_{IS}S_z^0 dt \\ I_z(\tau_m) - I_z(0) &= 2\sigma_{IS}\tau_m S_z^0 \\ I_z(\tau_m) &= 2\sigma_{IS}\tau_m S_z^0 + I_z^0\end{aligned}\quad \text{[equation 5]}$$

In Eq.[5] for zero mixing time, the I magnetization is equal to its equilibrium value, but as the mixing time increases, the I magnetization has an additional contribution, which is proportional to the mixing time and the cross-relaxation rate, σ_{IS} . This latter term results in a change in the intensity of the I spin signal, and this change is called an NOE enhancement. The normal procedure for visualizing these enhancements is to record a reference spectrum in which the intensities are unperturbed. In terms of z-magnetizations this means that $I_{z,\text{ref}} = I_z^0$. The difference spectrum, defined as (perturbed spectrum – unperturbed spectrum) corresponds to the difference

$$\begin{aligned}I_z(\tau_m) - I_{z,\text{ref}} &= 2\sigma_{IS}\tau_m S_z^0 + I_z^0 - I_z^0 \\ &= 2\sigma_{IS}\tau_m S_z^0\end{aligned}\quad \text{[equation 6]}$$

The NOE enhancement factor, η , is defined as

$$\eta = \frac{\text{intensity in enhanced spectrum} - \text{intensity in reference spectrum}}{\text{intensity in reference spectrum}}\quad \text{[equation 7]}$$

so in this case η is

$$\eta(\tau_m) = \frac{I_z(\tau_m) - I_{z,\text{ref}}}{I_{z,\text{ref}}} = \frac{2\sigma_{IS}\tau_m S_z^0}{I_z^0}\quad \text{[equation 8]}$$

and if I and S are of the same nuclear species (e.g. both proton), their equilibrium magnetizations are equal so that

$$\eta(\tau_m) = 2\sigma_{IS}\tau_m\quad \text{[equation 9]}$$

Hence, a plot of η against mixing time will give a straight line of slope σ_{IS} ; this is a method used for measuring the cross-relaxation rate constant. A single experiment for one value of the mixing time will reveal the presence of NOE enhancements. The dynamics of the NOE in 2D-NOESY experiments are very similar to those for the transient NOE experiment. The key difference is that instead of the

magnetization of the **S** spin being inverted at the start of the mixing time, the magnetization has an amplitude label, which depends on the evolution during t_1 .

The time dependence and size of the NOE enhancement depends on the relative sizes of the cross-relaxation rate constant σ_{IS} and the self-relaxation rate constants R_I and R_S . It turns out that these self-rates are always positive, but the cross-relaxation rate constant can be positive or negative. The reason for this is that $\sigma_{IS} = (W_2 - W_0)$ and it is quite possible for W_0 to be greater or less than W_2 .

A positive cross-relaxation rate constant means that if spin **S** deviates from equilibrium cross-relaxation will increase the magnetization on spin **I**. This leads to an increase in the signal from **I**, and hence a positive NOE enhancement. This situation is typical for small molecules in non-viscous solvents.

A negative cross-relaxation rate constant means that if spin **S** deviates from equilibrium cross-relaxation will decrease the magnetization on spin **I**. This leads to a negative NOE enhancement, a situation typical for large molecules in viscous solvents. Also, W_0 and W_2 can become equal and then the NOE enhancement goes to zero.

Thus, σ_{IS} can be either positive or negative. If σ_{IS} is positive, the diagonal and cross peaks will be of opposite sign, whereas if σ_{IS} is negative all the peaks will have the same sign.

In conclusion, the dipolar mechanism is the only common relaxation mechanism, which can cause transitions in which more than one spin flips. Specifically, the dipolar mechanism gives rise to transitions between the $\alpha\alpha$ and the $\beta\beta$ states (W_2) and between the $\alpha\beta$ and $\beta\alpha$ states (W_0) and hence to the NOE effect. The relaxation induced by the dipolar coupling is proportional to the square of the coupling, and it goes as

$$\gamma_I^2 \gamma_S^2 \frac{1}{r_{IS}^6}$$

where γ_I and γ_S are the gyromagnetic ratios of the two nuclei involved, and r_{IS} is the distance between them. It is evident, hence, that NOE enhancement could be strictly correlated with the geometry of the molecule.

Residual Dipolar Coupling: Theory and General Concepts

This alignment tensor is a key concept that is crucial to understand residual dipolar couplings. However, in our experience, there are conceptual difficulties to fully understand the physical meaning of the alignment tensor, which sometimes lead to serious misconceptions. This may result in part from the common practice in literature, to derive the alignment tensor using mathematically elegant, but not very intuitive approaches based on spherical harmonics, their addition theorems, Legendre polynomials, Wigner rotation matrices, and a confusing number of angles between various axes.

F. Kramer *et al.* in "Residual Dipolar Coupling Constant: An Elementary Derivation of Key Equations", Concepts Magn. Resonance (2004) 21A, 10-21

The use of residual dipolar couplings (RDCs) in the analysis of biomolecular structure and dynamics has expanded rapidly since its potential as a source of structural information on proteins was demonstrated in the mid 1990s (Tjandra et al. 1997; Tolman et al. 1995). The development of RDC applications has been reviewed periodically since their introduction to the structural biology field (see Table 3 for references list, Jiménez-Barbero and Peters, 2002)) and applications have spread to nucleic acid structure, carbohydrate structure, protein-ligand interactions, protein domain relationships, high-throughput strategies for structural genomics, and studies of motional amplitudes in flexible assemblies.

Table 3. Residual Dipolar Coupling reviews

| Authors | Title | Year |
|--|---|------|
| Zhou, H.; Vermeulen, A.; Jucker, F. M.; Pardi, A. | Incorporating Residual Dipolar Couplings into the NMR Solution Structure Determination of Nucleic Acids | 1999 |
| Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J.R. | NMR structures of biomolecules using field oriented media and residual dipolar couplings | 2000 |
| Bax, A.; Kontaxis, G.; Tjandra, | Dipolar couplings in macromolecular structure determination | 2001 |
| Brunner, E. | Residual Dipolar Couplings in Protein NMR | 2001 |
| Tolman, J. R. | Dipolar couplings as a probe of molecular dynamics and structure in solution | 2001 |
| Al-Hashimi, H. M.; Patel, D. J. | Residual dipolar couplings: synergy between NMR and structural genomics | 2002 |
| de Alba, E.; Tjandra, N. | NMR dipolar couplings for the structures determination of biopolymers in solution | 2002 |
| MacDonald, D.; Lu, P. | Residual dipolar couplings in nucleic acid structure determination | 2002 |
| Bax, A. D. | Weak alignment offers new NMR opportunities to study protein structure and dynamics | 2003 |
| Lipsitz RS, Tjandra N. | Residual dipolar couplings in NMR structure analysis | 2004 |
| Prestegard JH, Bougault CM, Kishore AI. | Residual dipolar couplings in structure determination of biomolecules | 2004 |

If we consider two spins I and S (Fig. 22), we can identify their dipolar coupling with the internuclear vector $\mathbf{R}^{\#}$ and its orientation respect the external magnetic field.

The formal analytical expression of dipolar coupling is represented in Eq. 10:

$$D = \frac{\kappa}{R^3} \left(\cos^2\theta - \frac{1}{3} \right) \quad [\text{Equation 10}]$$

[#] vector entities are here indicated with their corresponding bold letters.

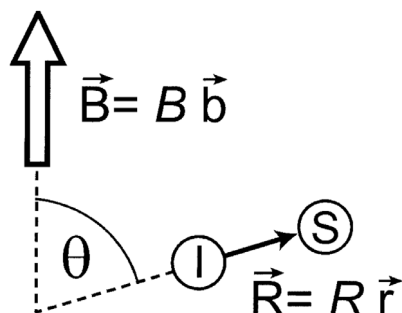


Figure 22. Definition of the angle θ between the internuclear vector \mathbf{R} (connecting spins I and S) and the magnetic field vector \mathbf{B} . The unit vectors \mathbf{r} and \mathbf{b} point in the direction of \mathbf{R} and \mathbf{B} , respectively.

where θ is the angle between the internuclear vector and the magnetic field (see Fig. 22) and the term κ :

$$\kappa = -\frac{3}{8\pi^2} \gamma_I \gamma_S \mu_0 \hbar \quad \text{[Equation 11]}$$

only depends on γ_I and γ_S , gyromagnetic ratios of spin I and S respectively; the Planck constant $\hbar = h/2\pi$ and the permeability of vacuum μ_0 . For example, for ^1H - ^1H , ^{13}C - ^1H , and ^{15}N - ^1H spin pairs, $\kappa = -360.3$, -90.6 , and $36.5 \text{ kHz } \text{\AA}^3$, respectively. The maximum possible value of $\cos^2 \theta$ is 1 (for $\theta = 0$ or π), and hence, according to Eq. [10], the maximum possible dipolar coupling constant is

$$D_{\max} = \kappa/R^3(1 - 1/3) = (2/3)\kappa/R^3 \quad \text{[Equation 12]}$$

which corresponds, e.g., to 21.7 kHz for a ^{15}N - ^1H spin pair with distance $R = 1.04 \text{ \AA}$.

It turns out that although dipolar coupling are the dominant interactions in solid state NMR of spin -1/2 nuclei, they are averaged to zero for isotropically reorienting molecules in the liquid state. This makes it possible to achieve high-resolution spectra with relative ease in liquid state NMR. On the other hand, a wealth of structural information is lost if dipolar couplings vanish. Fortunately, in liquid state NMR, molecules can be partially aligned, e.g., by external fields (magnetic or electric) or by anisotropic solvents. For example, in liquid crystalline solvents, the dissolved molecules are partially aligned through steric and anisotropic interactions with the solvent molecules, and dipolar couplings can be observed (Saupe et al. 1963). The recent success and wide use of such residual dipolar couplings is due to the development and characterization of several new alignment media such as bicelles (Tjandra et al. 1997), filamentous phage Pf1 (Hansen et al. 1998), and polyacrylamide gels (Sass et al. 2000; Tycko et al. 2000), which make it possible to create a relatively small, tunable degree of alignment. This allows to adjust the alignment in such a way that the size of the average dipolar coupling is on the order of the J couplings. In this case, the resulting spectra are still simple, and dipolar coupling constants can be measured, in a relatively easy manner, by comparing line splitting in isotropic and in aligned samples. In a weak alignment medium, the total Hamiltonian (sum of dipolar and scalar Hamiltonians)

directly corresponds to the experimentally observed line splitting in units of Hz. By repeating the same measurements in isotropic conditions, only scalar coupling can be measured and hence by just subtracting this result to the previous, only dipolar contribution can be extracted.

In the last couple of years, RDCs measurements have become fairly common methodologies, especially for solving relative stereochemistry of remote centers (Aroulanda et al. 2003; Yan et al. 2003) that are otherwise impossible to elucidate by employing standard approaches like scalar couplings and NOEs contacts. In these cases, just the experimental RDC values, without any further mathematical analysis have solved the relative orientation of C-H vectors (cis versus trans or axial versus equatorial).

The qualitative approach for RDCs is evidently not enough when incorporating them in molecular modeling protocols as additional restrains to the NOEs. In particular, back calculations of RDCs require a previous knowledge of the orientation and of the three components of the so-called alignment tensor. In a similar geometric approach of the original derivation from Saupe (Saupe et al. 1963). The alignment tensor could be derived from the related probability tensor. So, from a geometrical point of view, the $\cos \theta$ term in Eq. 10 could be interpreted as the scalar product between the two unit vectors \mathbf{b} and \mathbf{r} (figure 22) and the $\cos \theta$ term can also be expressed in the form

$$\cos \theta = \vec{b}^T \vec{r} \quad \text{[Equation 13]}$$

where \mathbf{b}^T is a row vector (the transpose of the column vector \mathbf{b}) that allows to write the scalar product of the two vectors as a usual matrix product between the 1X 3 matrix \mathbf{b}^T and the 3X 1 matrix \mathbf{r} .

When we consider the two spins I and S, we know that they are part of a molecule in solution. In the lab frame, the magnetic field vector \mathbf{B} is constant (pointing along the z^L axis), but the internuclear vector \mathbf{R} is time-dependent [see Fig. 23A]. For simplicity, we assume that the molecule is rigid (no internal dynamics with a constant distance, R), such that the time dependence of \mathbf{R} is solely due to the rotational tumbling motion of the molecule. Hence, the term $\cos \theta$ (and as a result also the dipolar coupling constant D and the dipolar coupling Hamiltonian) is time-dependent. For proteins, the rotational correlation time is on the order of nanoseconds and on the time scale of the NMR experiment, only the time-averaged dipolar Hamiltonian $\overline{H_D}$ gives rise to splitting in the spectrum. The time-averaged dipolar coupling constant

$$\bar{D} = \frac{\kappa}{R^3} \left(\overline{\cos^2 \theta} - \frac{1}{3} \right) \quad \text{[Equation 14]}$$

represents the so-called residual dipolar coupling constant, which depends on the average alignment of the molecule.

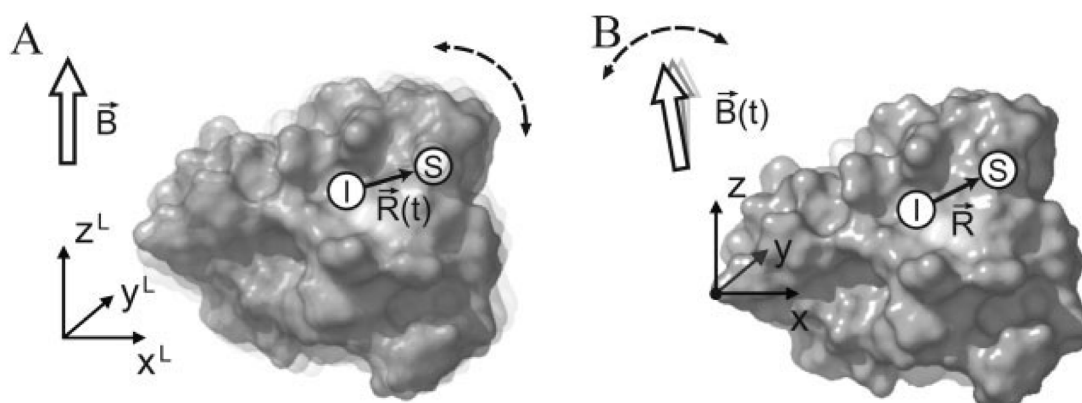


Figure 23. Effect of molecular tumbling of a rigid molecule as seen (A) from the lab frame of reference (with axes x^L , y^L , z^L) and (B) from an arbitrary molecular frame of reference (with axes x , y , z). In the lab frame (A), the magnetic field \mathbf{B} is constant and points by definition along the z^L axis, whereas the internuclear vector \mathbf{R} keeps changing its direction. In a molecular frame (B), the situation is reversed: here, any given internuclear vector is constant, whereas the orientation of the magnetic field is time-dependent.

When deriving a general approach for the calculation of \overline{D} , the easiest way is to use a second reference system different to the one depicted in Fig 23A, and hence move from the lab frame to a frame of reference that is fixed to the molecule (Fig. 23B). In this arbitrarily chosen molecular frame, with axes (x, y, z) , a given internuclear vector \mathbf{R} is constant (still assuming a rigid molecule without internal dynamics) and equal to

$$\tilde{\mathbf{R}} = R\tilde{\mathbf{r}} = R \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}$$

while the direction of the magnetic field vector \mathbf{B} is time-dependent:

$$\tilde{\mathbf{B}} = B\tilde{\mathbf{b}} = B \begin{pmatrix} b_x(t) \\ b_y(t) \\ b_z(t) \end{pmatrix}.$$

And now, the definition of $\cos \theta$ can be rewritten in this new reference system as:

$$\cos \theta = \tilde{\mathbf{b}}^T \cdot \tilde{\mathbf{r}} = \begin{pmatrix} b_x(t) & b_y(t) & b_z(t) \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix}$$

$$= b_x(t)r_x + b_y(t)r_y + b_z(t)r_z,$$

[Equation 15]

and

$$\begin{aligned}
\cos^2\theta &= (b_x(t)r_x + b_y(t)r_y + b_z(t)r_z)^2 \\
&= b_x^2(t)r_x^2 + b_x(t)b_y(t)r_xr_y + b_x(t)b_z(t)r_xr_z \\
&\quad + b_y(t)b_x(t)r_yr_x + b_y^2(t)r_y^2 + b_y(t)b_z(t)r_yr_z \\
&\quad + b_z(t)b_x(t)r_zr_x + b_z(t)b_y(t)r_zr_y + b_z^2(t)r_z^2. \quad [\text{Equation 16}]
\end{aligned}$$

Rearranging Eq. 16, it is possible to express the time average $\cos^2\theta$ in terms of the probability matrix, \mathbf{P} :

$$\begin{aligned}
\overline{\cos^2\theta} &= (r_x \quad r_y \quad r_z) \begin{pmatrix} \overline{b_x^2(t)} & \overline{b_x(t)b_y(t)} & \overline{b_x(t)b_z(t)} \\ \overline{b_x(t)b_y(t)} & \overline{b_y^2(t)} & \overline{b_y(t)b_z(t)} \\ \overline{b_x(t)b_z(t)} & \overline{b_y(t)b_z(t)} & \overline{b_z^2(t)} \end{pmatrix} \begin{pmatrix} r_x \\ r_y \\ r_z \end{pmatrix} \\
&= \vec{r}^T \mathbf{P} \vec{r}. \quad [\text{Equation 17}]
\end{aligned}$$

And from the knowledge of the probability matrix, \mathbf{P} , the residual dipolar coupling constant can be calculated from Eq. 18:

$$\bar{D} = \frac{\kappa}{R^3} \left(\vec{r}^T \mathbf{P} \vec{r} - \frac{1}{3} \right) \quad [\text{Equation 18}]$$

The matrix \mathbf{P} is real, symmetric, and has a trace of 1 because

$$\begin{aligned}
\text{tr}\{\mathbf{P}\} &= P_{xx} + P_{yy} + P_{zz} = \overline{b_x^2(t)} + \overline{b_y^2(t)} + \overline{b_z^2(t)} \\
&= \overline{(b_x^2(t) + b_y^2(t) + b_z^2(t))} = 1 \quad [\text{Equation 19}]
\end{aligned}$$

Since by definition, \mathbf{b} is a unit vector, hence, $b_x^2(t) + b_y^2(t) + b_z^2(t) = 1$ for all times t . Therefore, \mathbf{P} is fully specified by only five independent parameters.

In the NMR literature, it is not customary to consider the probability tensor \mathbf{P} , but to use its traceless part, which is called the alignment tensor \mathbf{A} :

$$\mathbf{A} = \mathbf{P} - \frac{1}{3} \mathbf{1} \quad [\text{Equation 20}]$$

Or alternatively, it is possible to use the principal values S_x , S_y and S_z of the Saupe matrix (or order matrix) \mathbf{S} :

$$\mathbf{S} = 3/2 \mathbf{A} \quad [\text{Equation 21}]$$

these definitions allow to introduce the *axial component* A_a and the *rhombic component* A_r of the alignment tensor as:

$$A_a = \frac{3}{2} A_z = S_z \quad \text{and} \quad A_r = A_x - A_y = \frac{2}{3} (S_x - S_y)$$

And hence the three most important parameters when studying RDCs in weak alignment media are

- Rhombicity $R = \frac{A_r}{A_a}$
- Asymmetry parameter $\eta = \frac{A_x - A_y}{A_z} = \frac{S_x - S_y}{S_z} = \frac{3}{2} R$
- Generalized degree of order (GDO) $GDO = \sqrt{\frac{3}{2}} |A| = \sqrt{\frac{2}{3}} |S|$

Finally, one of the last frontiers in the application of RDCs, and the most interesting for the purpose of this work, is the study of the dynamics of a carbohydrate (Martin-Pastor et al. 2005; Tian et al. 2001; Yi et al. 2004), which in combination with molecular dynamic simulations can facilitate the ability to separately defined structures and motions and could drastically improve the physical description of the conformers sampled.

NMR and molecular modeling

Molecular modeling, also known as molecular mechanics, is a method to calculate the structure and energy of molecules based on nuclear motions. Electrons are not considered explicitly, but rather it is assumed that they will find their optimum distribution, once the positions of the nuclei are known. This assumption is based on the Born-Oppenheimer approximation of the Schrödinger equation. The



Born-Oppenheimer approximation states that nuclei are much heavier and move much more slowly than electrons. Thus, nuclear motions, vibrations and rotations can be studied separately from electrons; the electrons are assumed to move fast enough to adjust to any movement of the nuclei. In a very simple sense, molecular modeling treats a molecule as a collection of weights connected with springs, where the weights represent the nuclei and the springs represent the

bonds.

A *force field* is used to calculate the energy and geometry of a molecule. It is a collection of atom types (to define the atoms in a molecule), parameters (for bond lengths, bond angles, etc.) and equations (to calculate the energy of a molecule). In a force field, a given element may have several atom types. For example, ethylbenzene contains both sp^3 -hybridized carbons and aromatic carbons. sp^3 -Hybridized carbons have a tetrahedral bonding geometry, while aromatic carbons have a trigonal

bonding geometry. The C-C bond in the ethyl group differs from a C-C bond in the phenyl ring, and the C-C bond between the phenyl ring and the ethyl group differs from all other C-C bonds in ethylbenzene. The force field (Perez et al. 1998) contains parameters for these different types of bonds and the most common in carbohydrate computing are CHARMM (Ha et al. 1988; Reiling et al. 1996), AMBER (Homans 1990; Woods et al. 1995), Macromodel-AMBER* (Senderowitz et al. 1997), GROMOS (Ott et al. 1996), OPLS (Damm et al. 1997), CVFF (Hwang et al. 1998), TRIPOS (Perez et al. 1995), and MM3 (Allinger et al. 1990).

The total energy of a molecule is divided into several parts, called force potentials, or potential energy equations. Force potentials are calculated independently, and summed to give the total energy of the molecule. Examples of force potentials are the equations for the energies associated with bond stretching, bond bending, torsional strain and van der Waals interactions. These equations define the potential energy surface of a molecule.

$$E_{\text{TOTAL}} = E_{\text{STRETCH}} + E_{\text{BEND}} + E_{\text{TORSION}} + E_{\text{VDW}} + E_{\text{DP-DP}}$$

Energy due to Bond Stretching

Whenever a bond is compressed or stretched the energy goes up. The energy potential for bond stretching and compressing is described by an equation similar to Hooke's law for a spring:



$$E_s = \frac{k_s}{2} (l - l_0)^2 \quad [\text{equation 22}]$$

where k_s is the force constant in $\text{mdyn}/\text{\AA}$

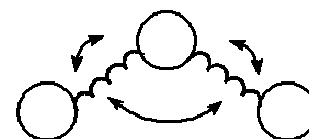
l_0 is the natural bond length in \AA

l is the actual bond length in \AA

For a typical alkene, C-C bond $k_s = 4.4 \text{mdyn}/\text{\AA}$ and $l_0 = 1.523 \text{\AA}$

Energy due to Bond Angle Bending

As angles are bent from their norm the energy increases. The potential function below works very well for bends of up to about 10 degrees.



$$E_\Theta = \frac{k_\Theta}{2} (\Theta - \Theta_0)^2 \quad [\text{equation 23}]$$

where k_Θ is the force constant in $\text{mdyn}/(\text{\AA} \text{rad}^2)$

Θ_0 is the natural bond angle in degrees

Θ is the actual bond angle in degrees

For an alkane C-C-C bond angle k_Θ is $0.45 \text{mdyn}/(\text{\AA} \text{rad}^2)$ and Θ is 109.5°

Energy due to Torsional Strain

Intramolecular rotations (rotations about torsion or dihedral angles) require energy. For example, it takes energy for cyclohexane to go from the chair conformation to the boat conformation. The torsion potential is a Fourier series that accounts for all 1-4 through-bond relationships:



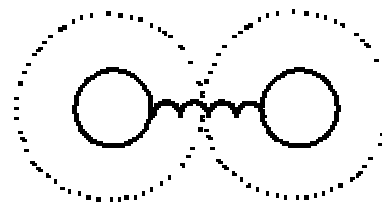
$$E_{\text{tor}} = \frac{V_1}{2}(1 + \cos \omega) + \frac{V_2}{2}(1 + \cos 2\omega) + \frac{V_3}{2}(1 + \cos 3\omega) \quad [\text{equation 24}]$$

where V_1, V_2, V_3 are force constants in the Fourier series in kcal/mol
 ω is the torsion angle from 0° to 180°

For a typical torsion angle in an alkane, $V_1=0.20, V_2=0.27, V_3=0.0093$

Energy due to van der Waals Interactions

The van der Waals radius of an atom is its effective size. As two non-bonded atoms are brought together, the van der Waals attraction between them increases (a decrease in energy). When the distance between them equals the sum of the van der Waals radii, the attraction is at a maximum. If the atoms are brought still closer together, there is strong van der Waals repulsion (a sharp increase in energy):



$$E_{\text{vdw}} = \varepsilon \left[\left(\frac{r_0}{r_V} \right) - 2 \left(\frac{r_V}{r_0} \right)^6 \right] \quad [\text{equation 25}]$$

where ε is the energy parameter which sets the depth of the potential energy well

r_V is the sum of the Van der Waals radii of the interacting atoms

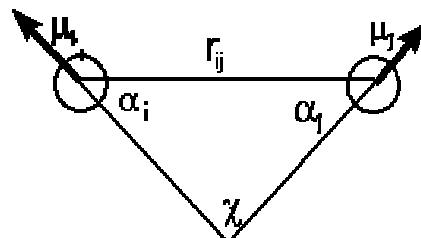
r_0 is the distance between the interacting centers

For carbon-carbon interactions ε is 0.044

For carbon-hydrogen interactions ε is 0.046

Energy due to Dipole-Dipole Interactions

In some force fields, the electrostatic interactions are accounted for by atomic point charges. In other force fields, such as MM2 and MMX, bond dipole moments are used to represent electrostatic contributions. One can readily see that the equation below stems from Coulomb's law. The energy is calculated by considering all dipole-dipole interactions in a molecule. If the molecule has a net charge (e.g., NH_4^+), charge-charge and charge-dipole calculations must also be carried out.



$$E_{\text{dipole}} = \frac{\mu_i \mu_j}{D(r_{ij})^3} (\cos \chi - 3 \cos \alpha_i \cos \alpha_j) \quad [\text{equation 26}]$$

where D is the dielectric constant of the solvent

χ is the angle between two the dipoles μ_i and μ_j

α_i, α_j are the angles between the dipoles and the vector connecting the two dipoles

r_{ij} is the distance between the dipoles

The derivative of the force potential energy, and hence the derivative of each term of the sum, determine the forces that act on each atom defined for the molecule.

$$F_i = -\frac{\partial V}{\partial x_i} = m_i a_i \quad [\text{equation 27}]$$

Consequently, minimization of this function for all the possible existing conformers allow to predict which are the geometries at the equilibrium and their corresponding relative energies. Knowledge of the forces, and hence of the acceleration and velocity of each atom, allows to solve the classical Newton equation, but at atomic level. This is the task of Molecular Dynamic simulations, where atom positions are predicted during the time on the basis of the force potential energy. As consequent, the correct choice of which force field has to be applied became fundamental before running any molecular simulation (Jensen 1999), including this particular case of carbohydrate molecules (Woods 2003).

Molecular Modeling and Nuclear Overhauser Enhancement

Quantitative interpretation of the NOEs became very common protocols since NOEs are the main constraints available from NMR spectroscopy, especially when modeling macromolecules. Several approaches to quantify distances from NOE-signal intensities exist, and here are presented the main four:

- Isolated spin pair approximation (ISPA);
- Direct comparison of cross and diagonal peaks (DIRECT);
- Reference structure based iterative methods (IRMA, MARDIGRAS);
- Simulation of nOe (NOEPROM).

Given a molecule with known distance relations, the cross peak intensities for a special mixing time, $V(\tau_m)$ in a 2D NOESY spectrum can be recalculated from known diagonal peak intensities $V(0)$ at mixing time $\tau_m=0$ by:

$$\mathbf{V}(\tau_m) = \exp(-\mathbf{R}\tau_m) \mathbf{V}(0) \quad [\text{equation 28}]$$

Here, \mathbf{R} is the relaxation matrix whose element are defined by:

$$R_{ij} = \sigma_{ij} = \frac{\gamma_i^2 \cdot \gamma_j^2 \cdot h^2}{40 \cdot \pi^2 \cdot r_{ij}^6} \cdot [6J_{2,ij}(\omega) - J_{0,ij}(\omega)]$$

[equation 29]

$$R_{ii} = \rho_i = \frac{\gamma_i^2 \cdot \gamma_j^2 \cdot h^2}{40 \cdot \pi^2 \cdot r_{ij}^6} \cdot \sum_{i,j} [J_{0,ij}(\omega) + 3J_{1,ij}(\omega) + 6J_{2,ij}(\omega) + A_{1,i}]$$

[equation 30]

$A_{1,i}$ is the contribution by non-dipolar relaxation mechanisms, which can be neglected in the absence of paramagnetic nuclei. Cross and diagonal elements of the relaxation matrix are a function of the spectral density J that describes the frequency dependence of a motion:

$$J_{n,ij}(\omega) = \frac{\tau_{ij}}{\left(1 + n^2 \omega^2 \tau_{ij}^2\right)}$$

[equation 31]

Isolated Spin Pair Approximation (ISPA)

Although the relationship between NOE-peak intensity and distance is nonlinear, the time course can be linearized under several experimental conditions (Gronenborn et al. 1985):

- short mixing time;
- short correlation time.

Truncation of a Taylor series expansion for [Eq.16] after the second term describes this linear relationship:

$$\exp(-R\tau_m) \approx 1 - R\tau_m + \frac{R^2 \tau_m^2}{2} - \dots + \left[\frac{(-1)^n}{n!} \right] R^n \tau_m^n \approx V(t_m)$$

[equation 32]

The cross-peak intensities depend only on the corresponding off-diagonal element in \mathbf{R} . Calibration of the unknown interproton distances is then possible using the peak intensity and distance of a proton pair with a fixed distance r_{ref} as reference by:

$$r_{ij} = r_{ref} \cdot \left(\frac{V_{ref}}{V_{ij}} \right)^{\frac{1}{6}}$$

[equation 33]

The ISPA method shows a small error for long-range distances and, therefore, can be used to reveal secondary structure elements and tertiary structures. In general, distances smaller than the reference are underestimated; others are overestimated by a factor of 5-6 between the percentages in distance

and intensity error. A solution for errors arising from inadequate reference distances is the utilization of two different references spanning a wider distance range and a modified formula [Eq.22] for calculation of distances.

$$r_{ij} = \frac{r_{ref1}V_{ref1}^6 - r_{ref2}V_{ref2}^6}{V_{ref1}^6 - V_{ref2}^6} + \left[\frac{r_{ref1} - r_{ref2}}{V_{ref1}^{-6} - V_{ref2}^{-6}} \right] \cdot (V_{ij})^{-\frac{1}{6}} \quad \text{[equation 34]}$$

Direct Comparison of cross and diagonal peaks

The DIRECT method (Olejniczak 1986) implies that the cross- and diagonal-peak intensities for all protons can be identified and accurately quantified. Rearrangement of [Eq.16] allows obtaining the distance estimates directly from the intensity matrix:

$$R = \frac{-(\ln V(\tau_m) \cdot V(0)^{-1})}{\tau_m} \quad \text{[equation 35]}$$

The diagonal-peak intensities at mixing time = 0 can be derived by extrapolating a series of NOE spectra run at different mixing times (build-up rate measurements).

Under typical experimental conditions not all diagonal peaks are resolved and cross peaks may be disturbed by overlaps and close proximity to the diagonal. Hence, the ideal and complete relaxation matrix representing the spatial arrangement of all protons will not be evaluated.

Reference structure based iterative methods

Iterative algorithms are designed to cope with incomplete intensity informations and are based on equation 23. All approaches aim to minimize the deviation between measured and calculated intensities while they refer to a proposed structural model. The programs differ in the modified subject to reach agreement between calculated and measured intensities, and just for citing some of these examples are MARDIGRAS (Borgias 1990), IRMA (Boelens 1988).

Simulation of nOe: NOEPROM Software

Noeprom software was developed in our laboratories (Martin-Pastor 1997) and can now be free downloadable at Dr. M. Martin Pastor's web page <http://desoft03.usc.es/mmartin/software.html> where also a complete documentation can be found. It can simulate and hence back calculate NOEs as 1D-NOE, 2D-NOESY or 2D-ROESY starting from a single structure or a bunch of structures originated from a Conformational Search (Boltzmann dependence) or from Molecular Dynamics simulations. Input files can be normal PDB or Macromodel (extension .out or .dat) files. General experimental inputs that have to be declared before running the calculation are: spectrometer frequency, mixing time, temperature, correlation time of the molecule (or an estimation of it) and the model to simulate molecular motions

(rigid isotropic or Lipari-Szabo model). The software runs in Windows platform, albeit a Linux version is available from the author upon request as beta-version; in the case of Windows a nice interface (Fig. 24) guides the user within all the possible entries and setup.

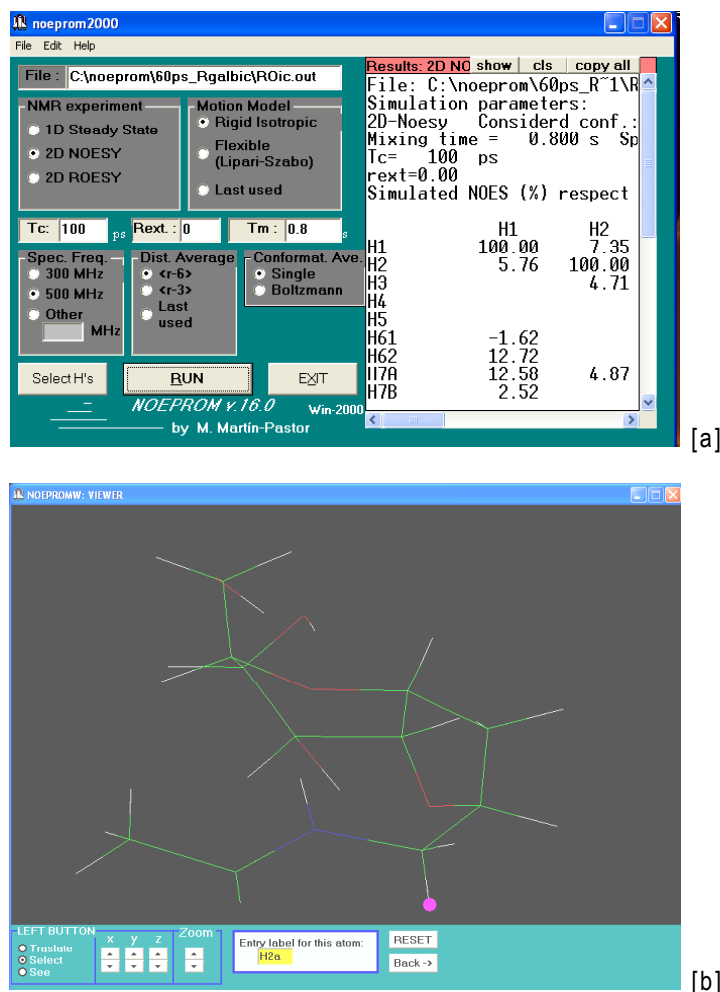


Figure 24. Main window of Noeprom software displaying the resulting .sol file [a]; the viewer window, which allows an easy atom selection from the input structure [b].

As illustrated in Fig. 24-b, in the first step, the atom selection is required. The Viewer window allows to customize atom labels and to save it as *.hhh file.

After calculation, Noeprom generates two different files as output:

- *.sol: this is the log file that contains the simulation results (noe percentages), and appear in the main window of the program (Fig.24-a)
- *.NOE: this file contains the average distances calculated for each atom pair selected. This distances file could be use as input for a new calculation with Noeprom

The application of this software will be found in the second part of the present work.

References

- Abeygunawardana, C. and C. A. Bush (1993) "Determination of the chemical structure of complex polysaccharides by heteronuclear NMR spectroscopy" *Adv. Biophys. Chem.* **3**: 199-249.
- Ahn, Y. H. and J. S. Yoo (1999) "Effect of the labeling group in structural analyses of malononitrile-labeled oligosaccharides by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and electrospray ionization mass spectrometry" *Rapid Commun Mass Spectrom* **13**(20): 1985-90.
- Allinger, N. L., M. Rahman and J.-H. Lii (1990) "A molecular mechanics force field (MM3) for alcohol and ethers." *J Am Chem Soc* **112**: 8293-8307.
- Anumula, K. R. and S. T. Dhume (1998) "High resolution and high sensitivity methods for oligosaccharide mapping and characterization by normal phase high performance liquid chromatography following derivatization with highly fluorescent anthranilic acid" *Glycobiology* **8**(7): 685-94.
- Aroulanda, C., V. Boucard, F. Guibe, J. Courtieu and D. Merlet (2003) "Weakly oriented liquid-crystal NMR solvents as a general tool to determine relative configurations" *Chemistry* **9**(18): 4536-9.
- Bax, A., R. H. Griffey and H. B. L. (1983) "Correlation of proton and nitrogen-15 chemical shifts by multiple quantum NMR" *J Magn Reson* **55**: 301-315.
- Bax, A. and M. F. Summers (1986) "Proton and carbon-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR" *J Am Chem Soc* **108**: 2093-2094.
- Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi (1977) "The Protein Data Bank: a computer-based archival file for macromolecular structures" *J Mol Biol* **112**(3): 535-42.
- Bock, K., C. Pedersen and H. Heding (1974) "A ¹³C-NMR spectroscopic study of alpha- and beta-streptomycin" *J Antibiot (Tokyo)* **27**(2): 139-40.
- Bodenhausen, G. and D. J. Ruben (1980) "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy" *Chem Phys Lett* **69**(1): 185-189.
- Boelens, R. (1988) *J Mol Struct* **173**: 299-311.
- Borgias, B. A. (1990) *J Magn Reson* **87**: 475-487.
- Cano, F. H., C. Foces-Foces, J. Jimenez-Barbero, A. Alemany, M. Bernabe and M. Martin-Lomas (1987) "Experimental Evidence Of Deviations From A Karplus-Like Relationship Of Vicinal Carbon-Proton Coupling Constants In Some Conformationally Rigid Carbohydrate Derivatives" *J Org Chem* **52**: 3367-3372.
- Chandrasekaran, R. (1997) "Molecular architecture of polysaccharide helices in oriented fibers" *Adv Carbohydr Chem Biochem* **52**: 311-439.
- Cumming, D. A. and J. P. Carver (1987) "Reevaluation of rotamer populations for 1,6 linkages: reconciliation with potential energy calculations" *Biochemistry* **26**(21): 6676-83.
- Damm, W., A. Frontera, J. Tirado-Rives and W. L. Jorgensen (1997) "OPLS all-atom force field for carbohydrates." *J Comput Chem* **18**: 1955-1970.
- de Beer, T., C. W. van Zuylen, K. Hard, R. Boelens, R. Kaptein, J. P. Kamerling and J. F. Vliegthart (1994) "Rapid and simple approach for the NMR resonance assignment of the carbohydrate chains of an intact glycoprotein. Application of gradient-enhanced natural abundance ¹H-¹³C HSQC and HSQC-TOCSY to the alpha-subunit of human chorionic gonadotropin" *FEBS Lett* **348**(1): 1-6.
- Doubet, S., K. Bock, D. Smith, A. Darvill and P. Albersheim (1989) "The Complex Carbohydrate Structure Database" *Trends Biochem Sci* **14**(12): 475-7.

- Eggens, I., B. Fenderson, T. Toyokuni, B. Dean, M. Stroud and S. Hakomori (1989) "Specific interaction between Lex and Lex determinants. A possible basis for cell recognition in preimplantation embryos and in embryonal carcinoma cells" *J Biol Chem* **264**(16): 9476-84.
- El Rassi, Z. and Y. S. Mechref (1997) *Capillary Electrophoresis, Theory and Practice*, London, CRC Press.
- Gitti, R., G. Long and C. A. Bush (1994) "Measurement of long-range ^{13}C - ^1H coupling constants of 95% uniformly ^{13}C -labeled polysaccharide from *Streptococcus mitis* J22" *Biopolymers* **34**(10): 1327-38.
- Gronenborn, A. M. and G. M. Clore (1985) "Investigation of the solution structures of short nucleic acid fragments by means of nuclear Overhauser enhancement measurements" *Prog Nucl Magn Reson Spectrosc* **17**: 1-32.
- Ha, S. N., A. Giammona, M. Field and J. W. Brady (1988) "A revised potential-energy surface for molecular mechanics studies of carbohydrates" *Carbohydr Res* **180**(2): 207-21.
- Hansen, M. R., L. Mueller and A. Pardi (1998) "Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions" *Nat Struct Biol* **5**(12): 1065-74.
- Haverkamp, J., T. Spoormaker, L. Dorland, J. F. G. Vliegthart and R. Schauer (1979) "Determination of the β -anomeric configuration of cytidine 5'-monophospho-N-acetylneuraminic acid by carbon- ^{13}C NMR spectroscopy" *J Am Chem Soc* **101**: 4851-4853.
- Hermansson, K., P. E. Jansson, L. Kenne, G. Widmalm and F. Lindh (1992) "A ^1H and ^{13}C NMR study of oligosaccharides from human milk. Application of the computer program CASPER" *Carbohydr Res* **235**: 69-81.
- Homans, S. W. (1990) "A molecular mechanical force field for the conformational analysis of oligosaccharides: comparison of theoretical and crystal structures of Man α 1-3Man β 1-4GlcNAc" *Biochemistry* **29**(39): 9110-8.
- Homans, S. W. (1990) "Oligosaccharide conformations: Application of NMR and energy calculations" *Prog Nucl Magn Reson Spectrosc* **22**: 55-81.
- Hwang, M. J., X. Ni, M. Waldman, C. S. Ewig and A. T. Hagler (1998) "Derivation of class II force fields. VI. Carbohydrate compounds and anomeric effects" *Biopolymers* **45**(6): 435-68.
- Imberty, A. and S. Perez (1995) "Stereochemistry of the N-glycosylation sites in glycoproteins" *Protein Eng* **8**(7): 699-709.
- Jansson, P. E., L. Kenne and G. Widmalm (1987) "Casper—a computerised approach to structure determination of polysaccharides using information from n.m.r. spectroscopy and simple chemical analyses" *Carbohydr Res* **168**(1): 67-77.
- Jansson, P. E., L. Kenne and G. Widmalm (1991) "CASPER: a computer program used for structural analysis of carbohydrates" *J Chem Inf Comput Sci* **31**(4): 508-16.
- Jansson, P. E., L. Kenne and G. Widmalm (1991) "Computer-assisted structural analysis of oligosaccharides using CASPER" *Anal Biochem* **199**(1): 11-7.
- Jeffrey, G. A. and W. Saenger (1991) *Hydrogen bonding in biological structures*, Berlin ; New York, Springer-Verlag.
- Jensen, F. (1999) *Introduction to Computational Chemistry*, England, John Wiley & Sons.
- Jimenez-Barbero, J. and T. Peters (2003) *NMR spectroscopy of glycoconjugates*, Weinheim, Wiley-VCH.
- Jones, D. N. K. and J. K. M. Sanders (1989) "Biosynthetic studies using carbon- ^{13}C -COSY: the *Klebsiella* K3 serotype polysaccharide" *J Am Chem Soc* **111**: 5132-5137.
- Keeler, J., R. T. Clowes, A. L. Davis and E. D. Laue (1994) "Pulsed-field gradients: theory and practice" *Methods Enzymol* **239**: 145-207.
- Kover, K. E., V. J. Hruby and D. Uhrin (1997) "Sensitivity- and gradient-enhanced heteronuclear coupled/decoupled HSQC-TOCSY experiments for measuring long-range heteronuclear coupling constants" *J Magn Reson* **129**(2): 125-9.

- Lemieux, R. U., R. K. Kullnig, H. J. Bernstein and W. G. Schneider (1958) "Configurational Effects on the Proton Magnetic Resonance Spectra of Six-membered Ring Compounds" *J Am Chem Soc* **80**: 6098-6105.
- Lerner, L. and A. Bax (1986) "Sensitivity-enhanced two-dimensional heteronuclear relayed coherence transfer NMR spectroscopy" *J Magn Reson* **69**: 375-380.
- Martin-Pastor, M. (1997) "Estructura y Dinamica de oligo y polisacaridos mediante RMN" *PhD Thesis UAM*.
- Martin-Pastor, M., A. Canales, F. Corzana, J. L. Asensio and J. Jimenez-Barbero (2005) "Limited flexibility of lactose detected from residual dipolar couplings using molecular dynamics simulations and steric alignment methods" *J Am Chem Soc* **127**(10): 3589-95.
- Martin-Pastor, M., A. Canales-Mayordomo and J. Jimenez-Barbero (2003) "NMR experiments for the measurement of proton-proton and carbon-carbon residual dipolar couplings in uniformly labelled oligosaccharides" *J Biomol NMR* **26**(4): 345-53.
- Milton, M. J., R. Harris, M. A. Probert, R. A. Field and S. W. Homans (1998) "New conformational constraints in isotopically (¹³C) enriched oligosaccharides" *Glycobiology* **8**(2): 147-53.
- Mueller, L. (1979) "Sensitivity enhanced detection of weak nuclei using heteronuclear multiple quantum coherence" *J Am Chem Soc* **101**: 4481-4484.
- Mulloy, B., T. A. Frenkiel and D. B. Davies (1988) "Long-range carbon—proton coupling constants: Application to conformational studies of oligosaccharides" *Carbohydr Res* **184**: 39-46.
- Olejniczak, E. T. (1986) *J Magn Reson* **67**: 28-41.
- Ott, K. H. and B. Meyer (1996) "Parametrization of GROMOS force field for oligosaccharide and assessment of efficiency of molecular dynamics simulations." *J Comput Chem* **17**: 1068-1084.
- Perez, S. and H. Chanzy (1989) "Electron crystallography of linear polysaccharides" *J Electron Microsc Tech* **11**(4): 280-5.
- Perez, S., A. Imberty, S. B. Engelsen, J. Gruza, K. Mazeau, J. Jimenez-Barbero, A. Poveda, J. F. Espinoza, B. Van Eyck, G. Johnson, A. French, M. L. Kouwijzer, P. Grootenhuis, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten and K. Rasmussen (1998) "A Comparison and Chemometrics Analysis of Several Molecular Mechanics Force Fields and Parameters Sets for Carbohydrates" *Carbohydr Res* **314**: 141-155.
- Perez, S., C. Meyer and A. Imberty (1995) in *Modelling of Biomolecular Structures and Mechanisms*. Practical tools for accurate modeling of complex carbohydrates and their interactions with proteins. A. Pullman, Jortner, J. and Pullman, B. Dordrecht, Kluwer Academic, p. 425-454.
- Perez, S., N. Mouhous-Riou, N. E. Nifant'ev, Y. E. Tsvetkov, B. Bachet and A. Imberty (1996) "Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide" *Glycobiology* **6**(5): 537-42.
- Poppe, L. and H. Van Halbeek (1991) "¹H detected measurements of long-range heteronuclear coupling constants. Application to a trisaccharide" *J Magn Reson* **92**: 636-641.
- Poppe, L. and H. Van Halbeek (1991) "Selective, inverse-detected measurements of longrange C-H coupling constants. Application to a disaccharide." *J Magn Reson* **93**: 214-217.
- Prytulla, S., J. Lambert, M. Klessinger and J. Thiem (1990) *Magn Reson Chem* **28**: 888-901.
- Prytulla, S., J. Lauterwein, M. Klessinger and J. Thiem (1991) "Configurational assignment of N-acetylneuraminic acid and analogues via the vicinal C,H coupling constants" *Carbohydr Res* **215**(2): 345-349.
- Reiling, S., M. Schlenkrich and J. Brickmann (1996) "Force field parameters for carbohydrates" *J Comput Chem* **17**: 450-468.
- Rudiger, H., H. C. Siebert, D. Solis, J. Jimenez-Barbero, A. Romero, C. W. von der Lieth, T. Diaz-Marino and H. J. Gabius (2000) "Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets" *Curr Med Chem* **7**(4): 389-416.

- Sass, H. J., G. Musco, S. J. Stahl, P. T. Wingfield and S. Grzesiek (2000) "Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes" *J Biomol NMR* **18**(4): 303-9.
- Sato, Y., M. Suzuki, T. Nirasawa, A. Suzuki and T. Endo (2000) "Microsequencing of glycans using 2-aminobenzamide and MALDI-TOF mass spectrometry: occurrence of unique linkage-dependent fragmentation" *Anal Chem* **72**(6): 1207-16.
- Saupe, A. and G. Englert (1963) "High-resolution nuclear magnetic resonance spectra of oriented molecules." *Phys Rev Lett* **11**: 462-464.
- Senderowitz, H. and W. C. Still (1997) "A quantum mechanically derived all-atom force field for pyranose oligosaccharides. AMBER* parameters and free energy simulations." *J Org Chem* **62**: 1427-1438.
- Tian, F., H. M. Al-Hashimi, J. L. Craighead and J. H. Prestegard (2001) "Conformational analysis of a flexible oligosaccharide using residual dipolar couplings" *J Am Chem Soc* **123**(3): 485-92.
- Tjandra, N. and A. Bax (1997) "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium" *Science* **278**(5340): 1111-4.
- Tolman, J. R., J. M. Flanagan, M. A. Kennedy and J. H. Prestegard (1995) "Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution" *Proc Natl Acad Sci U S A* **92**(20): 9279-83.
- Tvaroska, I., M. Hricovini and E. Petrakova (1989) "An attempt to derive a new Karplus-type equation of vicinal proton-carbon coupling constants for C---O---C---H segments of bonded atoms." *Carbohydr Res* **189**: 359-362.
- Tycko, R., F. J. Blanco and Y. Ishii (2000) "Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR." *J Am Chem Soc* **122**: 9340 -9341.
- van Halbeek, H. (1994) "NMR developments in structural studies of carbohydrates and their complexes." *Curr Opin Struct Biol* **4**: 697-709.
- Van Halbeek, H. (1996) in *Carbohydrates and Glycoconjugates*. Encyclopedia of NMR. New York, Wiley.
- van Kuik, J. A., K. Hard and J. F. Vliegthart (1992) "A 1H NMR database computer program for the analysis of the primary structure of complex carbohydrates" *Carbohydr Res* **235**: 53-68.
- van Kuik, J. A. and J. F. Vliegthart (1992) "Databases of complex carbohydrates" *Trends Biotechnol* **10**(6): 182-5.
- Warin, V., F. Baert and R. Fouret (1979) "The crystal and molecular structure of O- α -D-mannopyranosyl-(1-3)-O- β -D-mannopyranosyl-(1-4)-2-acetamido-2-deoxy- α -D-glucopyranose" *Carbohydr Res* **76**: 11-22.
- Woods, R. J. (2003) in *NMR spectroscopy of glycoconjugates*. Jimenez-Barbero, J. Peters, T. Weinheim, Wiley-VCH.
- Woods, R. J., R. A. Dwek and C. J. Edge (1995) "Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM-93 parameter development." *J Am Chem Soc* **99**: 3832-3846.
- Xu, Q. and C. A. Bush (1996) "Molecular modeling of the flexible cell wall polysaccharide of *Streptococcus mitis* J22 on the basis of heteronuclear NMR coupling constants" *Biochemistry* **35**(46): 14521-9.
- Xu, Q. and C. A. Bush (1998) "Measurement of long-range carbon-carbon coupling constants in a uniformly enriched complex polysaccharide" *Carbohydr Res* **306**(3): 335-9.
- Xu, Q., S. Mohan and C. A. Bush (1996) "A flexible model for the cell wall polysaccharide of *Streptococcus mitis* J22 determined by three-dimensional ^{13}C edited nuclear overhauser effect spectroscopy and ^{13}C - ^1H long-range coupling constants combined with molecular modeling" *Biopolymers* **38**(3): 339-53.

- Yan, J., A. D. Kline, H. Mo, M. J. Shapiro and E. R. Zartler (2003) "A novel method for the determination of stereochemistry in six-membered chairlike rings using residual dipolar couplings" *J Org Chem* **68**(5): 1786-95.
- Yi, X., A. Venot, J. Glushka and J. H. Prestegard (2004) "Glycosidic torsional motions in a bicelle-associated disaccharide from residual dipolar couplings" *J Am Chem Soc* **126**(42): 13636-8.

The bound state

The process underlying most NMR-based ligand binding screening experiments can be described by Eq.36:



Equation 36 represents a dynamic equilibrium involving three species: the free receptor E, the free ligand L, and the receptor-ligand complex, EL. The unimolecular rate constant k_{off} is inversely proportional to the mean lifetime τ_B of the receptor-ligand complex. The bimolecular rate constant, k_{on} , measures the probability of a productive encounter between the free receptor and the ligand and is often assumed to be controlled by diffusion. Thus, k_{on} can vary between 10^7 and $10^9 \text{ M}^{-1} \text{ s}^{-1}$. The binding affinity can be described by the temperature-dependent equilibrium dissociation constant, K_D :

$$K_D = \frac{[E] \cdot [L]}{[EL]} = \frac{k_{off}}{k_{on}} \quad \text{[Equation 37]}$$

the bound receptor fraction is defined by

$$P_B^E = \frac{[EL]}{[E] + [EL]} \quad \text{[Equation 38]}$$

Eq. 25 and 26 can be combined yielding equation 27:

$$P_B^E = \frac{[L]}{[L] + K_D} \quad \text{[Equation 39]}$$

where P_B^E is the fractional occupation of the receptor binding site by ligand L.

When $[L] \ll K_D$, P_B^E is proportional to $[L]$. When $[L] = K_D$, the receptor is half-saturated; that is, half of the receptor molecules exist in a one-to one complex with the ligand. When $[L] \gg K_D$, the receptor is completely saturated and $P_B^E = 1.0$; in this limit, every receptor binding site is occupied by one ligand, which on average exchanges with a second distinct ligand every $\approx 1/k_{off}$ seconds. According to these points, ligands with weak affinity have large K_D and thus require the addition of more ligand to saturate the receptor binding site.

In the two-state equilibrium given by Eq. 36, ligand and receptor molecules will exist in either a free (L, E) or complexed (EL) state. In the free state, both receptor and ligand retain their intrinsic NMR parameters (e.g. chemical shifts, relaxation rates, translational diffusion coefficients). In each other's presence, the mutual binding affinity of ligand and receptor drives an exchange process that toggles both sets of molecules between the free and complexed states. At equilibrium, they adopt free and bound state populations ($[E]$, $[L]$, $[EL]$) consistent with Eq 36. Under these conditions, the ligand transiently adopts NMR parameters characteristic of the typically much larger receptor. Alternatively, from the receptor's perspective, the ligand transiently perturbs the binding site microenvironment(s), which may alter the distribution of conformations sampled by the ensemble of receptor molecules. In

either case, the exchange modulates the NMR parameters of both molecules. It is the ability to experimentally distinguish these exchange-modulated parameters from those of the free state what enables NMR detection of the receptor-ligand interaction and hence NMR-based screening.

A complete discussion of how chemical exchange modifies the NMR parameters (dynamic NMR) should include a description of the modified Bloch equation formalism of Hahn, Maxwell, and McConnell (Hahn et al. 1952; McConnell 1958). This formalism (HMM) provides an excellent theoretical framework to describe the majority of the exchange phenomena that occur in NMR screening experiments, under all exchange regimes. Such a treatment is beyond the scope of the present Thesis but can be found in a recent review (Peng et al. 2004). For most applications described here, it is sufficient to consider only the case of fast exchange.

Fast Exchange Approximation

The solutions to the HMM equations (Hahn et al. 1952; McConnell 1958) describe the behavior of system magnetization on arbitrary exchange time scales. In NMR-based screening, however, these equations are almost never solved, and fast exchange is simply assumed. This assumption is made for two reasons. First, the experimental conditions for ligand-based NMR screening are often well suited to fast exchange. These experiments are typically carried out with $L_T/E_T > 10$ (where L_T and E_T are total ligand and total receptor concentration respectively) and the binding compounds have $K_D \geq 100\mu\text{M}$. If k_{on} is well approximated by a diffusion-limited value (10^7 - $10^9 \text{ M}^{-1} \text{ s}^{-1}$), then the slowest exchange rate k_{ex} values lie in the range $1000 < k_{ex} < 100\,000 \text{ s}^{-1}$. Ligand-based NMR screening methods are primarily ^1H based and consequently, k_{ex} exceeds most differences in intrinsic ^1H relaxation rates and rotating frame precession frequencies, thus ensuring that the fast exchange assumption is valid. A second motivation for assuming fast exchange is the resulting algebraic simplicity. Generally, the NMR parameters Q become the simple averages.

$$Q_{avg} = P_B Q_B + P_F Q_F \quad [\text{Equation 40}]$$

Here, Q_{avg} is the observed exchange-averaged parameter for the ligand (or receptor) in the presence of the receptor (or ligand). Observed differences between Q_{avg} and Q_F provide a signature of receptor binding and indicate a hit in a NMR screening based on that parameter. In the case of Eq.28, Q_{avg} is a simple population-weighted average and it applies to those parameters Q for which chemical shift modulations are not relevant. These parameters include longitudinal autorelaxation and cross-relaxation rates, rotating-frame spin-locking autorelaxation and cross-relaxation rates, and translational diffusion coefficients. The bound state contribution in Eq.28 is $P_B Q_B$. The ability to detect binding with adequate sensitivity depends critically on $P_B Q_B$, being significant relative to $P_F Q_F$. However, typical screening conditions where $L_T \gg E_T$ make $P_B \ll P_F$. For this reason, it is much preferred to measure NMR parameters Q that become amplified in the bound state (i.e. $Q_B \gg Q_F$).

Ligand-Based versus Receptor-Based Screening

Screening may proceed by ligand- or receptor-based methods. Receptor-based methods observe and compare the NMR parameters of the receptor molecule resonances in the presence and absence of compound mixtures. Thus far, the receptor-based methods have focused mostly on proteins. Such methods incorporate site-specific characterization afforded by assigned protein NMR spectra along with *a priori* knowledge of the protein's three-dimensional structure (either from X-ray or NMR) to drive lead generation. By identifying perturbations of assigned protein resonances, not only are ligands identified, but also their binding sites are localized. Localization of binding sites also enables one to immediately distinguish specific from nonspecific binding. Finally, unlike ligand-based methods, receptor-based methods do not rely on fast exchange to retrieve bound state information. Observation of receptor resonances permits the characterization of both higher and lower affinity hits.

NMR methods demand physicochemical properties of the protein target that present progressively more difficult challenges. First, milligram quantities of soluble, non-aggregated protein must be over-expressed and purified. Then, suitable expression hosts must be found that permit isotope enrichment (e.g. ^{13}C , ^{15}N , ^2H) critical for the resonance assignment of typically large (>30000 Da) therapeutic targets. After sufficient quantities of labeled protein are available, it must be ensured that the sample is stable for the time required for sequential resonance assignment. Although new data acquisition approaches promise to accelerate resonance assignment, it can still be a relatively lengthy process (weeks) for the large monomeric proteins (>30000 Da) routinely encountered in pharmaceutical research. Unfortunately, the time required for NMR assignment of such targets inevitably favors other approaches, such as X-ray crystallography, that can provide high-resolution structural information to medicinal chemistry on a faster time scale (if crystals are available). Alternatively, the typical implementation of ligand-based methods compares the NMR parameters of a mixture of compounds in the presence and absence of the receptor molecules. This approach renders the molecular weight of the receptor molecule irrelevant. In fact, the most powerful ligand-based approaches become more sensitive when dealing with larger receptors. Additionally, ligand observation bypasses the need to produce milligram quantities of isotope labeled receptor. Depending on the approach, less than a milligram of unlabeled protein is required for these experiments (receptor concentration is often $\leq 1\mu\text{M}$, and no assignments are necessary). This allows the scientist to evaluate new targets more rapidly and to contribute on a time scale useful for chemistry and high-throughput screens. An obvious disadvantage of ligand-based approaches is the inability to localize the binding site of the small molecule hits on the receptor. Also, ligand-based approaches rely on the exchange-mediated transfer of bound state information to the free state. This requirement biases ligand-based methods towards the identification of weakly binding ligands (rapid exchange) and the use of large ligand molar excesses ($L_T/E_T \gg 1$). The consequent risk is that, under these conditions, ligand may start to occupy, weaker affinity, nonspecific binding sites.

Both receptor- and ligand-based approaches have distinct advantages and disadvantages. Clearly, if receptor-based methods are possible, then the potentially higher information content that can

be obtained makes these the methods of choice. However, due to the scarcity of low-molecular weight drug targets, ligand-based screening is, in general, of broader applicability and places less demands on other research disciplines and infrastructure. The main research summarized in this Thesis refers to ligand-based NMR methods and in particular to tr-NOE and STD techniques. Next sections will explain, in detail, the fundamental of these methodologies.

Ligand-Based NMR Methods

Ligand-based methods rely on the possibility to differentiate between free and bound states from the NMR point of view. Generally, ligand-based screening methods use molecules with masses < 1kDa. These compounds exhibit small relaxation rates $R_1=1/T_1$ and $R_2=1/T_2$, 2D-NOESY cross-peaks that have opposite sign of diagonal peaks, and large translational diffusion coefficients, D . Bound compounds share the NMR properties of the much larger receptor (mass > 30000 Da). Therefore, bound compounds have large R_2 (and large selective R_1), 2D-NOESY cross-peaks of the same sign of diagonal peaks, and highly efficient spin diffusion, along with smaller molecular diffusion coefficients D .

These distinct differences imply that changes in NMR spectral parameters of the ligand can be monitored as a way to assess target binding. Most ligand-based NMR experiments detect binding by one of two mechanisms: (1) exploiting the differential mobility of the ligand in the free versus bound state, and (2) exploiting a ^1H magnetization transfer process from the receptor.

Transferred NOE

It is well established that NOE effects (NOEs) are extremely useful in determining the 3D structure of molecules in solution (see above) (Neuhaus et al. 2000) When ligand molecules bind to receptor proteins, the NOEs undergo drastic changes leading to the observation of transferred NOEs (trNOEs). These changes are the basis for a variety of experimental schemes that are designed to detect and characterize binding activity. The observation of trNOEs relies on the existence of rather different tumbling times τ_c for the free and bound molecules. Low- or medium-molecular-weight molecules ($M_w < 1000\text{Da}$) have a short correlation time τ_c and, as a consequence, such molecules exhibit positive NOEs. Large molecules, however, exhibit strongly negative NOEs. When a small molecule (ligand) is bound to a large-molecular weight protein (the protein receptor molecule), it behaves as a part of the large molecule and adopts the corresponding NOE behavior, that is, it shows strong negative NOEs, so-called trNOEs. These trNOEs reflect the bound conformation of the ligand. Binding of a ligand to a receptor protein can thus easily be distinguished by looking at the sign and size of the observed NOEs (Figure 25). Furthermore, the discrimination between trNOEs originating from the bound state and NOEs of the ligand in solution can also be achieved by the build-up rate, that is, the time required to achieve maximum intensity, which for trNOEs is in the range of 50 to 100 ms, whereas for small, non-binding, molecules it is four- to ten-times longer. Therefore, the maximum enhancement for trNOEs is observed at significantly shorter mixing times τ_m than for isolated small molecules in solution. Various experimental implementations have been explored in the last two decades, ranging from 1D selective

steady-state experiments to 1D and 2D transient NOE experiments (Jimenez-Barbero et al. 2003; Neuhaus et al. 2000).

However, one of the major drawbacks of this experiment is the possible existence of spin diffusion effects, which are typical for large molecules. In this case, apart from direct enhancements between protons close in space, other spins (including those of the receptor) may mediate the exchange of magnetization, thus producing negative cross peaks between protons far apart in the macromolecule. Thus, protein-mediated, indirect trNOE effects may lead to interpretation errors in the analysis of the ligand bound conformation. However, using trNOEs in the rotating frame (TR-ROESY) experiments (Arepalli et al. 1995; Asensio et al. 1995), it is possible to distinguish which cross peaks are dominated by an indirect effect, usually mediated by a protein proton, and hence distinguishing direct from indirect enhancements.

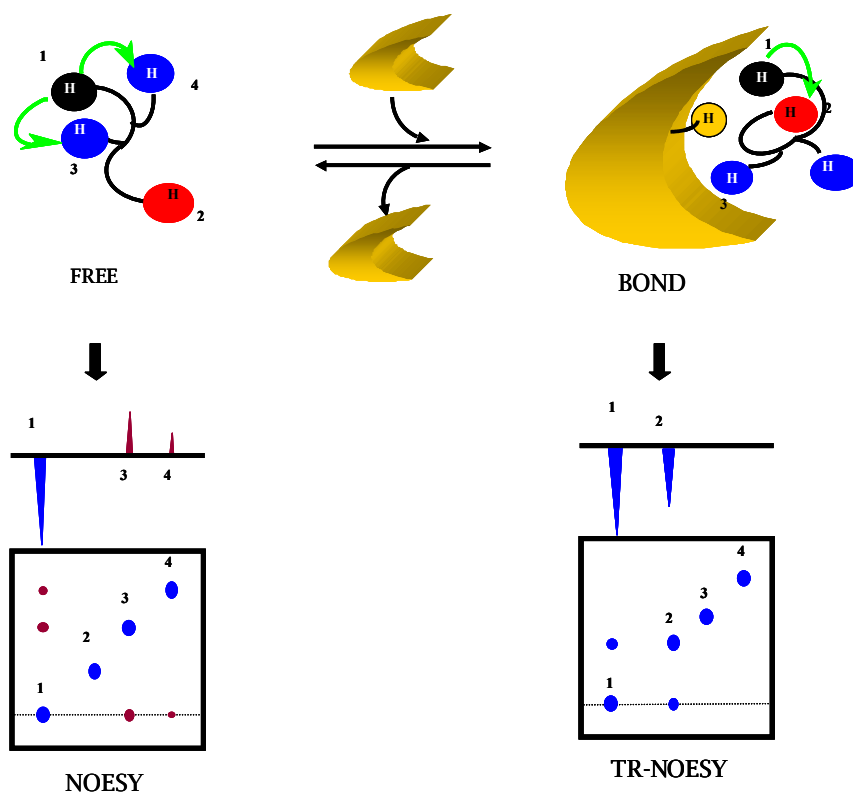


Figure 25. Left. Schematic representation of a NOESY spectrum for a free sugar. Cross peaks and diagonal peaks have different signs. Right. Schematic representation of a TR-NOESY spectrum recorded for an exchanging ligand–protein system. Cross peaks and diagonal peaks have the same signs, as expected for a large molecule, thus indicating binding to the protein. The relative sizes of the peaks and the appearance of new ones may be used to detect conformational variations. (Poveda et al. 1998)

From a more rigorous point of view, the dipole-dipole cross relaxation rate constants in the laboratory frame (σ^{NOE}) and in the rotating frame (σ^{ROE}) under fast exchange, became the population-weighted averages of the free and the bound states:

$$\sigma_{avg}^{NOE} = P_F \sigma_F^{NOE} + P_B \sigma_B^{NOE}$$

[Equations 41a,b]

$$\sigma_{avg}^{ROE} = P_F \sigma_F^{ROE} + P_B \sigma_B^{ROE}$$

In terms of spectral density functions, they have the form:

$$\sigma_{ij}^{NOE} = \frac{\hbar^2 \gamma_H^4}{4} \frac{1}{r_{ij}^6} \{6J_{ij}(2\omega_H) - J_{ij}(0)\}$$

[Equations 42a,b]

$$\sigma_{ij}^{ROE} = \frac{\hbar^2 \gamma_H^4}{4} \frac{1}{r_{ij}^6} \{2J_{ij}(0) + 3J_{ij}(\omega_H)\}$$

where i and j are the two proton pairs.

The changing in the sign of the cross-peaks passing from the free to the bound-state can be seen in Eq. 41a. In the free state, the molecule is characterized by small τ_c , satisfying $\omega_H \tau_c \ll 1$, $J_{ij}(2\omega_H) \approx J_{ij}(0)$, and $\sigma_F^{NOE} \approx +5 J_{ij}(0)$. In the bound state, large τ_c characterize the system, satisfying $\omega_H \tau_c \gg 1$, $J_{ij}(2\omega_H) \approx 0$, and $\sigma_B^{NOE} \approx -J_{ij}(0)$. By contrast, σ^{ROE} does not flip in sign. The estimated range of binding affinities that can be probed by transferred σ^{NOE} is $100\text{nM} \leq K_D \leq 1\text{mM}$ (Mayer et al. 2000).

In conclusion, the trNOE method allows fast screening of possible binders respect to a specific target receptor and, at the same time, permits the knowledge of the recognized conformation of the ligand bound to the receptor, what has considerable implication for a rational structure-based drug design.

Saturation Transfer Difference

Saturation-transfer NMR spectroscopy has been used for a long time in the characterization of binding between ligands and receptors (Akasaka 1979). More recently, Bernd Mayer and Thomas Peters have developed a method based on the transfer of saturation from the protein to the bound ligands, which in turn, by exchange, is moved into solution, where it is detected (Mayer et al. 1999). By utilizing difference spectroscopy, homonuclear methods, especially proton NMR experiments can easily be used to obtain well-resolved spectra of the ligand. By subtracting one spectrum in which the protein resonances are saturated from a second one, without protein saturation, a difference spectrum is produced in which only the signals of the ligand(s) that have felt the transfer of saturation from the protein do appear (Figure 26).

The irradiation frequency is set at a position where only resonances from the protein nuclei and no resonances from the ligand nuclei are located. Therefore, in the on-resonance experiment, selective saturation of the signals of the protein nuclei is achieved. For the on-resonance irradiation frequency

values around -1 ppm are practical because no ligand nuclei resonances are found in this spectral region, whereas the significant line width of protein signals still allows selective saturation. If the ligands show no resonance signals in the aromatic proton spectral region, the saturation frequency may also be placed there (7 ppm) or even further downfield ($\delta=11-12$ ppm). In order to achieve the desired selectivity and to avoid side-band irradiation, shaped pulses are employed for the saturation of the protein signals (Meyer et al. 2003).

The degree of ligand saturation naturally depends on the residence time of the ligand in the protein-binding pocket. The dissociation of the ligand will then transfer this saturation into solution, where the free ligand again gives rise to resonance signals with narrow line widths.

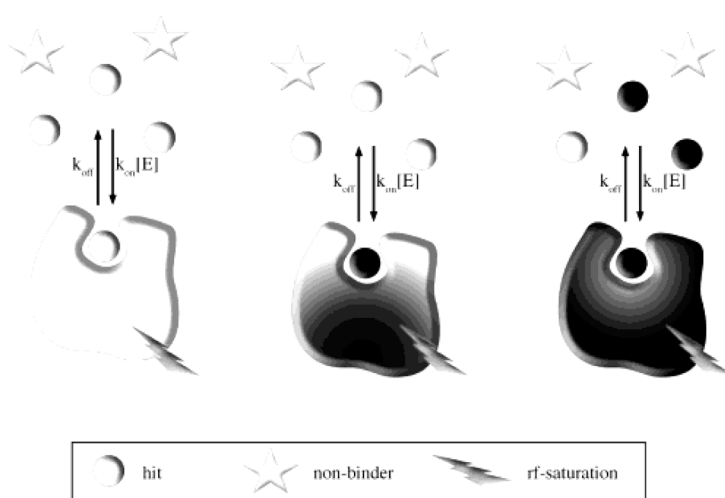


Figure 26. Detection of binding using the saturation transfer difference (STD) experiment. (Meyer et al. 2003; Peng et al. 2004) Frequency selective irradiation (lightning bolt) causes selective ^1H saturation (shading) of the target receptor. Irradiation is applied for a sustained interval during which saturation spreads throughout the entire receptor via ^1H - ^1H cross-relaxation (spin diffusion). Saturation is transferred to binding compounds (circles) during their residence time in the receptor binding site. The number of ligands having experienced saturation transfer increases as more ligand exchanges on and off the receptor during the sustained saturation period. Nonbinding compounds (stars) are unaffected.

For those ligand protons that interact with protein protons through an intermolecular NOE, a decrease in intensity is observed. However, in the presence of other molecules such as impurities and other non-binding components it is not usually possible to identify such attenuated signals. Therefore, in a second experiment, the irradiation frequency is set at a value that is far from any signal, ligand or protein, for example, 40 ppm (off-resonance spectrum). The spectrum is recorded and yields a normal NMR spectrum of the mixture. Subtraction of the on-resonance from the off-resonance spectra leads to a difference spectrum, in which only the signals of the protons that were attenuated by saturation transfer are visible. All molecules without binding activity are cancelled out (Figure 27).

Saturation of the protein and the bound ligand is very fast (about 100 ms). Therefore, a fast off rate of the ligand transfers the information about saturation quickly into solution. If a large excess of the ligand is present, one binding site can be used to saturate many ligand molecules in a few seconds.

Ligands in solution lose their information by normal T_1/T_2 relaxation, which is in the order of one second for small molecules.

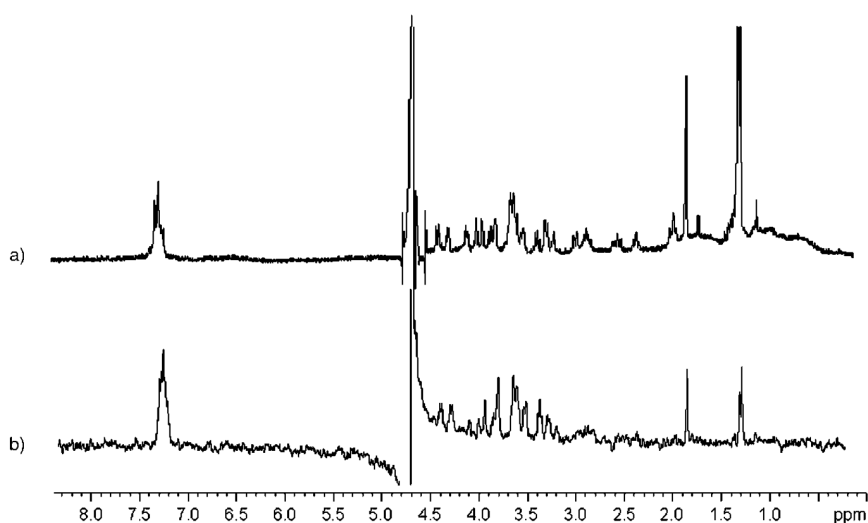


Figure 27. a) Reference 1D ^1H NMR spectrum of a GM1 mimic with CTB5 (ratio 25:1) in D_2O at 300 K. b) STD spectrum with 1 s presaturation of the protein envelope protons. Ligand proton signals are evident in the aromatic region. Protons are those belonging to the Gal and GalNAc moieties are clearly visible, while the other protons of the mimic do not appear in the STD spectrum. Details about this experiments will be found in the second part of this Thesis.

Thus, the proportion of saturated ligands in solution increases during the saturation time, and so the information about the bound state resulting from the saturated protein is amplified, which means that only a relatively small amount of protein is required. On the other hand, if binding is very tight, and consequentially off rates are in the range of 0.1–0.01 Hz, the saturation transfer to ligand molecules is not very efficient. This is usually the case for K_D values less than 1 nM. If the K_D values are 100 nM, or larger, fast exchange of free and bound ligands leads to a very efficient build up of saturation of the ligand molecules in solution.

The observed intensities of the signals arising from the ligand in the STD NMR spectrum are not directly proportional to the binding strength. STD NMR effects depend largely on the off rate. As outlined above, larger off rates should result in larger STD signals. However, when binding becomes very weak the probability of the ligand being in the receptor site becomes very low, which results in weak STD signals. STD NMR spectroscopy can be used from tight binding up to a K_D of about 10 mM.

The intensity of the STD signals depends, among other things, on the irradiation time/saturation time (Figures 28) and on the excess of ligand molecules used (Meyer et al. 2003). The more ligand that is used and the longer the irradiation time, the stronger the STD signal is. In general, the irradiation time is 2 s and a 100-fold excess of ligand is used to give good results. Upon dissociation, the saturation of the ligand is transported into solution, where it accumulates as a result of the slow decline of the saturation by relaxation processes in solution. Before ligands in solution have lost their saturation, the process of association (followed by dissociation) can occur many times and thus put many more saturated ligands into solution. The maximum net effect of saturation on ligand protons

occurs if a large excess of ligand is used, because it is very unlikely that a ligand with saturation re-enters the binding site. From the high ligand:protein ratios it is clear that the amount of protein required for the measurements is very small. At 500 MHz, an amount of approximately 0.3 nmol of protein is sufficient to record STD spectra. At a molecular weight of 50 kDa, this translates into about 15 μ g of protein.

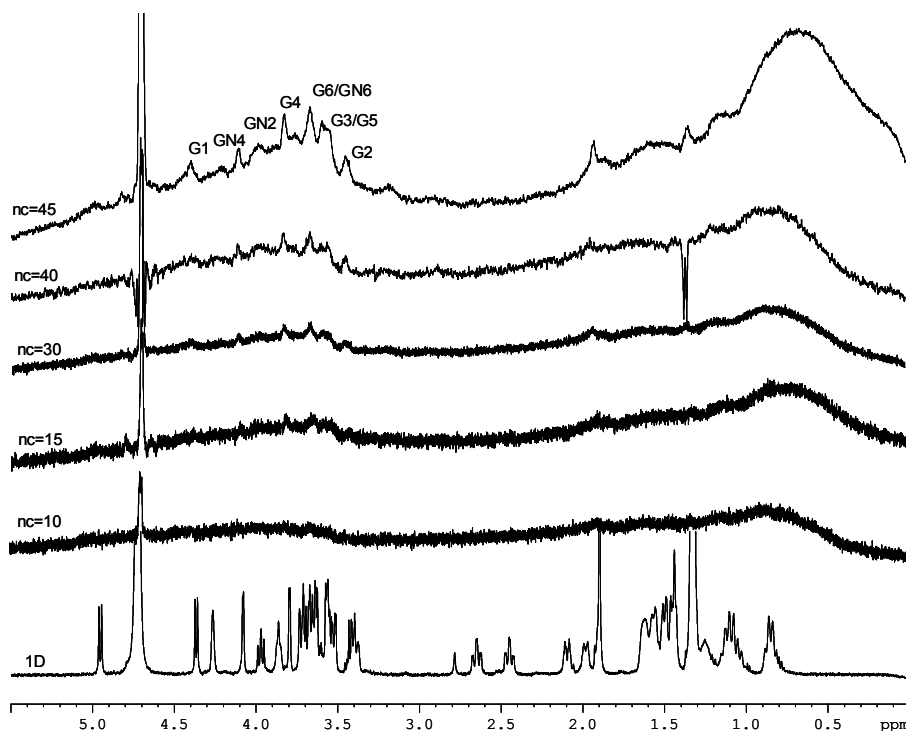


Figure 28. Personal results on dependence of STD signals with respect saturation time (here represented as number of loop cycles, nc) for the case of a GM1 mimic bound to VAA lectin (molar ratio 50:1). Each nc last for 50 ms, and hence the total saturation time varies from 500ms (second spectrum from the bottom) to more than 2s (first spectrum from the top). Reference 1H spectrum is also illustrated (first spectrum on the bottom).

It has been already shown (Figure 27) that STD-NMR is an excellent technique for determining the binding epitope of the ligand, information that is of primary importance for a rational drug design process. As consequence of the intermolecular saturation transfer process described above, the ligand saturation is higher for those protons that are in closer contact with the receptor. This kind of information can be accurately analyzed (Jayalakshmi et al. 2002) and the STD effect can be quantitatively analyzed in a similar way to that mentioned above for quantitative estimation of trNOEs. The results discussed in this Thesis refer exclusively to a qualitative analysis of STD experiments.

Applications of RDCs to the study of multiple fragments or complexes

One of the inherent limitations that arises for the analysis of dipolar data stems from the fact that dipolar coupling measurements do not have a one-to-one correspondence to a pair of polar angles

describing the two spin pairs interaction vectors. A given residual dipolar coupling measurement can only restrict the orientation of the corresponding internuclear vector to two cones (or distorted cones), as shown in Fig. 29. The addition of data corresponding to other vectors, if they are rigidly related to the first one, can reduce the number of possible solutions, but it could not reduce the solution to a single point.

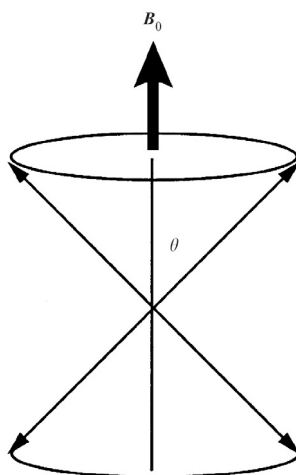


Figure 29. In the absence of motional averaging, a single residual dipolar coupling measurement restricts the orientation of an internuclear vector to two cones of orientations, subtended by the angle θ , relative to the magnetic field.

The degeneracy arises because the relationship between the individual vectors and the order tensor principle axes depend on $\cos^2\theta$, and are, thus, insensitive to inversion of any of the principal axes of the alignment tensor.

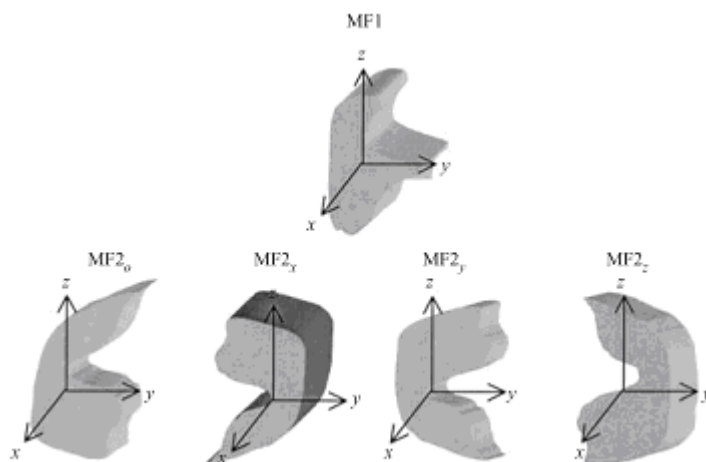
Hence, two molecular fragments (two protein subunits or a ligand interacting with its receptor) can be assembled relative to one another, but only with a four-fold degeneracy; “n” molecular fragments can be assembled only with a 4^{n-1} -fold degeneracy (Fig. 30a). Such degeneracy clearly would limit our ability to define protein structures, if only dipolar data were used. In many circumstances, NOE data and/or bonding constraints might be used to overcome this limitation. However, the most reliable method for reducing this degeneracy is by employing, at least, a second aligning medium.

The approach used in combining data is illustrated in Fig. 30b (Al-Hashimi et al. 2000; Ramirez et al. 1998). Briefly, measurement of five or more independent residual dipolar couplings in each molecular fragment, MF1 and MF2, allows determination of order tensor elements, and a principal

alignment frame for each. Structure determination then proceeds by aligning the resulting principal order frames. For two molecular fragments, MF1 and MF2, this gives rise to four unique relative orientations that are consistent with residual dipolar coupling measurements (Fig. 30a). If a new set of residual dipolar coupling data is recorded by using a second aligning medium, this choice produces a new non-coincident order tensor orientation.

Analogously, it is possible to generate a new set of four possible orientations of MF2 relative to MF1. Fragment MF1 can now be oriented identically for the two sets (Fig. 30b), and the relative orientations for MF2 in the two media can be compared. If the media do not bias the conformation of the molecule or supramolecule, the relative orientation of the fragments should be identical in only one case. This corresponds to the correct orientation of MF1 relative to MF2 (dark grey fragments). This approach has been demonstrated for two fragments in the protein rubredoxin (Al-Hashimi et al. 2000). The determined orientation of two molecular fragments using this procedure proved to be in excellent agreement with the available crystal structure.

a) Alignment Medium 1



b) Alignment Medium 2

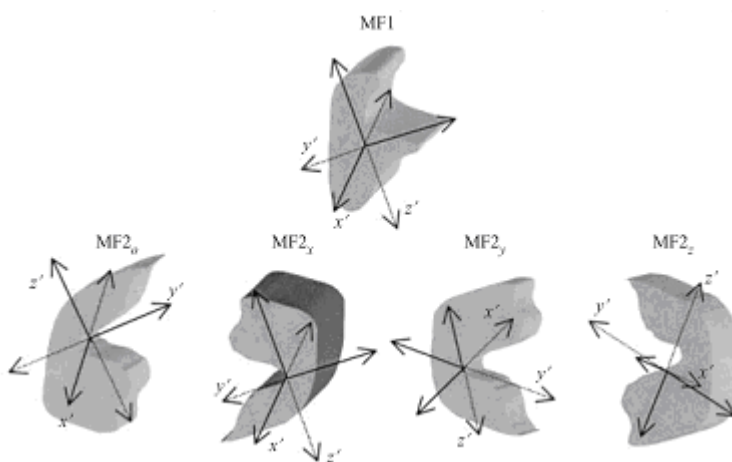


Figure 30. Resolving orientational degeneracy using two non-coincident order tensors. (a) The four possible ways to assemble two molecular fragments using a single order tensor frame from alignment medium 1. Note that while the molecular orientations MF2_o, MF2_x, MF2_y and MF2_z will always be distinguishable for non-symmetrical molecules, the designation 'o', 'x', 'y' and 'z' is arbitrary. (b) The orientation of the order tensor frame (x' , y' , z') using alignment medium 2 from the point of view of the molecular orientations). Only one of the four possible orientations (MF2_x shaded in dark gray) has a coincident order tensor frame with MF1.

References

- Akasaka, K. (1979) *J Magn Reson* **36**: 135-140.
- Al-Hashimi, H. M., H. Valafar, M. Terrell, E. R. Zartler, M. K. Eidsness and J. H. Prestegard (2000) "Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings" *J Magn Reson* **143**(2): 402-6.
- Arepalli, S. R., C. P. J. Glaudemans, G. D. Daves, P. Kovac and A. Bax (1995) "Identification of protein-mediated indirect NOE effects in a disaccharide-Fab' complex by transferred ROESY" *J Magn Reson Ser B* **106**: 195-198.
- Asensio, J. L., F. J. Canada, M. Bruix, A. Rodriguez-Romero and J. Jimenez-Barbero (1995) "The interaction of hevein with N-acetylglucosamine-containing oligosaccharides. Solution structure of hevein complexed to chitobiose" *Eur J Biochem* **230**(2): 621-33.
- Hahn, E. L. and D. E. Maxwell (1952) "Spin Echo Measurements of Nuclear Spin Coupling in Molecules" *Phys Rev* **88**(5): 1070-1084.
- Jayalakshmi, V. and N. Rama Krishna (2002) "Complete Relaxation and Conformational Exchange Matrix (CORCEMA) Analysis of Intermolecular Saturation Transfer Effects in Reversibly Forming Ligand-Receptor Complexes" *J Magn Reson* **155**: 106-118.
- Jimenez-Barbero, J. and T. Peters (2003) *NMR spectroscopy of glycoconjugates*, Weinheim, Wiley-VCH.
- Mayer, M. and B. Meyer (2000) "Mapping the active site of angiotensin-converting enzyme by transferred NOE spectroscopy" *J Med Chem* **43**(11): 2093-9.
- Mayer, M. and C. Meyer (1999) "Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy" *Angew Chem Int Ed Engl* **38**(12): 1784 - 1788.
- McConnell, H. M. (1958) "Reaction Rates by Nuclear Magnetic Resonance" *J Chem Phys* **28**(3): 430-431.
- Meyer, B. and T. Peters (2003) "NMR Spectroscopy Techniques for Screening and Identifying Ligand Binding to Protein Receptors" *Angew Chem Int Ed Engl* **42**(8): 864 - 890.
- Neuhaus, D. and M. P. Williamson (2000) *The nuclear Overhauser effect in structural and conformational analysis*, New York, J. Wiley.
- Peng, J. W., J. M. Moore and N. Abdul-Manan (2004) "NMR experiments for lead generation in drug discovery" *Prog Nucl Magn Reson Spectrosc* **44**: 225-256.
- Poveda, A. and J. Jimenez-Barbero (1998) "NMR studies of carbohydrate-protein interactions in solution" *Chem Soc Rev* **27**: 133-143.
- Ramirez, B. E. and A. Bax (1998) "Modulation of the alignment tensor of macromolecules dissolved in a dilute liquid crystalline medium" *J Am Chem Soc* **120**: 9106-9107.



Results and Discussion

..Is it just a matter how much curious we are?

Objectives

In this Thesis, we have addressed the study of the molecular recognition of glycomimetics and natural sugars by using NMR spectroscopy. Apart from my personal formation issue, we have selected this topic because we feel that it is timely, scientifically interesting, and very few examples have been described from a rigorous view point. Different biologically relevant systems have been selected that vary from toxins to enzymes, passing through a variety of plant and animal lectins. In all these cases, apart of the intrinsic nature of the investigation, we have tried to implement new methods and to extend the use of different NMR protocols.

The experimental work is presented here in a format that includes a brief introduction to the topic followed by the published paper/s (or documents in submission) in scientific journals. Five different topics are here discussed, defined with regard to possible targets with particular biomedical interest. Table 4 summarizes the fields and the related publications.

Table 4

| TOPIC | REFERENCE |
|---------------------------------------|--|
| Cholera Toxin B ₅ pentamer | Mimics of ganglioside GM1 as cholera toxin ligands: replacement of the GalNAc residue Intramolecular carbohydrate-aromatic interactions and intermolecular van der Waals interactions enhance the molecular recognition ability of GM1 glycomimetics for cholera toxin Synthesis and conformational analysis of galactose-derived bicyclic scaffolds |
| DC-SIGN | 1D saturation transfer difference NMR experiments on living cells: the DC-SIGN/oligomannose interaction |
| Glycosidase inhibitors | The Conformational Behavior of Novel Glycosidase Inhibitors with Substituted Azepan Structures: An NMR and Modeling Study |
| Carbamannose glycomimetic | Synthesis, Conformational Studies and Mannosidase Stability of a Mimic of 1,2-Mannobioside |
| VAA, CG14 and CTB5 | Conformer selection of a phenyl-lactic GM1 glycomimetic by three different lectins |

Following this section, a general view of my participation to international congresses as poster contributions or oral communications/lectures are presented.

The Cholera Toxin B₅ pentamer and its ligands

The enterotoxin produced by *Vibrio Cholera* (CT) belongs to a family of toxins with common functional properties like, for instance, ADP-ribosylation of ATPase. This family includes also *Escherichia Coli* enterotoxin (LT), diphtheria toxin, pertossis toxin and the exotoxin A. The structures of these proteins are the major causes for severe enteritis and bacterial diarrhea, both in human (CT and LT) and animals (LT). CT and LT are strictly related in terms of their physiological activity and from the structural point of view, both are hexamers AB₅-type, with a sequence homology of more than 80% (Finkelstein 1988).

The structures of these proteins (Merritt et al. 1994; Sixma et al. 1993; Sixma et al. 1991) and of some of their complexes with ligands (Jobling et al. 1991; Merritt et al. 1994; Merritt et al. 1994; Sixma et al. 1992) have been investigated by X ray diffraction and electron microscopy. LT and CT toxins are synthesized in the cytosol of the bacteria as A and B monomers, which consequently are assembled in the periplasmatic space to form the AB₅ complex. Some *Vibrio Cholerae* strands directly produce extracellular CT toxin, whereas some enterotoxigenic strands of *Escherichia coli* could secrete the toxin within membrane fragments. In a second phase, the free toxin interactS with the host intestinal epithelial cells by specific recognition between B₅ pentamer and the GM1 ganglioside, which is present in the external part of the double layer cell membrane (Figure 31). The A-subunit is the actual toxic part of the molecule. In order to produce the toxic effect, proteolysis between fragment 192 and 194 is required as well as reduction of the disulfide bridge between 186 and 199.

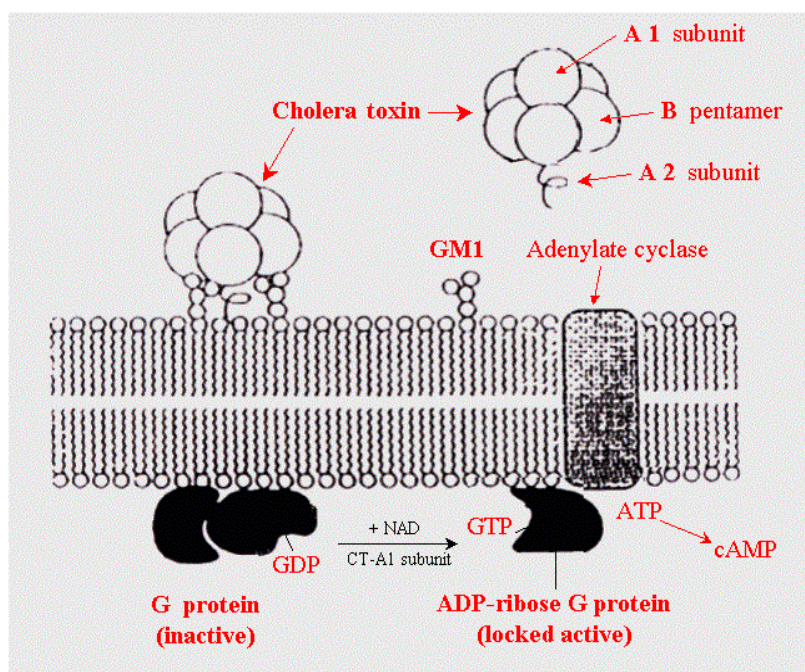


Figure 31. Schematic representation of CT toxin activity

These processes generate two fragments: the so-called A1 (constituted by residues 1-192(194)), responsible of the pathological activity of the toxin and probably for the recognition of NAD⁺, and the A2 fragment, (constituted by residues 193(195)-240), responsible of the union to the B5 pentamer. During the first stage of the infection, the A1 fragment of CT is able to cross the cell membrane of the target cell and to catalyze the transfer of an ADP-ribose group from NAD⁺ to a arginine's guanidium group of the A subunit of the G protein (figure 31). Upon this transfer, the G protein is locked in its active state and, consequently, the cAMP levels increase, leading to the lost of salts and fluids from the infected cell, and thus the development of the disease. On the other hand, the five B subunits constitute the centers for the specific recognition of the GM1 receptor located on the membrane surface of the target cell. Upon binding, the B5 pentamer does not exhibit any dramatic changes and, above all, the GM1 binding site on the pentamer is relatively far with respect to the A subunit. The correlated disease evolves principally with an intestinal illness, which could manifest from a weak diarrhea to a severe dehydration, which could end with the death of the patient. Knowledge of the structure of this protein and its activity became very important in the development of possible drugs and vaccines against these diseases. This would explain why CT and LT are object of intense studies.

The membrane receptor: GM1 ganglioside

GM1 {Gal(β1-3)GalNAc(β1-4)[NeuAc(α2-3)]Gal(β1-4)Glc(β1-4)ceramide} is a ganglioside, whose terminal part is constituted by a pentasaccharide. This is the receptor for CT toxin (Figure 32).

GM1 has been found anchored in different human cell membranes, particularly of the nervous system, but its physiological function on the epithelial intestinal cells is still unknown. Even so, the interaction between GM1 and CT has been demonstrated to be very specific (De Wolf et al. 1981): those cells that do not present GM1 receptor on the surface do not recognize the toxin; and by adding GM1 those cells, they are sensitive respect to the toxin interaction. Furthermore, GM1 specifically binds to the AB₅ hexamer and to the B₅ pentamer, but not to the B monomer. In fact, the binding site for GM1 is found to be on the interface within two B monomers. As a consequence, the five binding sites on B₅ act, in a cooperative manner, for GM1 recognition.

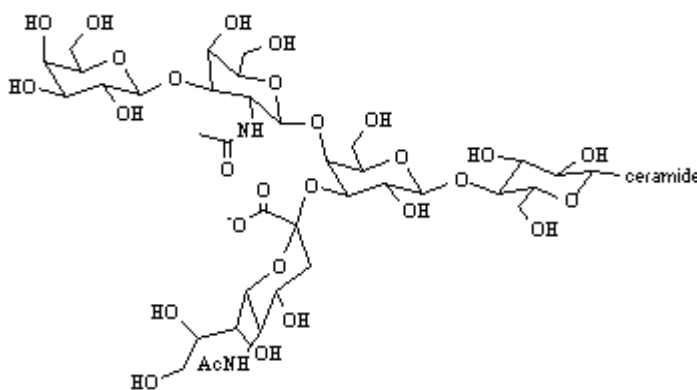


Figure 32. Saccharidic portion of GM1

The CT:GM1 complex

The first data about recognition of this type of toxins and sugars were obtained with the X-ray structure of the LT:lactose complex, while in 1994 the first structure of the CT:GM1 complex was solved (Merritt et al. 1994). As depicted in figure 33, the galactose and sialic acid moieties of GM1 are in contact with the protein, the N-Acetyl galactosamine residue has a minor contribution, whereas the other two GM1 residues seem to serve as a mere scaffold.

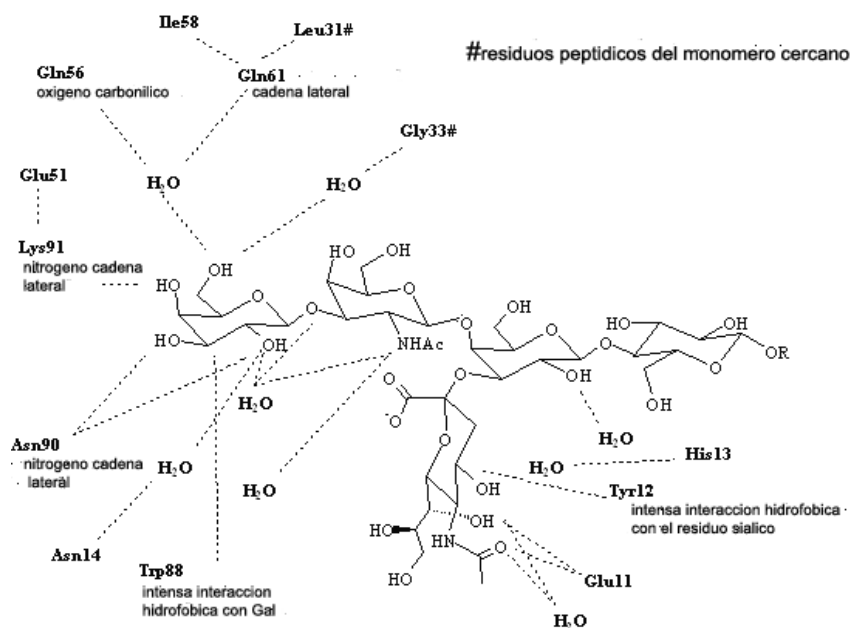


Figure 33. Scheme of the binding site of the complex CT:GM1

Approximately, 350\AA^2 of the GM1 surface, which, in principle, are accessible to the solvent are involved in the interaction with the protein. In particular, 43% are related to the sialic acid, 39% to the terminal galactose, and 17% to the N-Acetyl galactosamine residue. As already mentioned, the hydroxyl groups O-2, O-3 and O-4 of the galactose are hydrogen bond acceptors with respect to donors at the Asn90 and Lys91 lateral chains. Also, indirect hydrogen bonds, mediated by water, are found with residues Asn14 and Gly51. Finally, the galactose ring stacks with the aromatic portion of Trp88, which allows the hydrophobic interaction mentioned in the first part.

Regarding sialic acid, direct hydrogen bonding interactions are found with Tyr12, Glu11, His13, as well as with the N-Acetyl galactosamine moiety. Also, two indirect hydrogen bonds, mediated by water, are found with the terminal galactose. The N-Acetyl galactosamine does not seem to take part at any hydrogen bonding, only a contact between the C β of His13 and its methyl group is observed. No other interactions are seen between the lactose moiety and the protein.

Development of CT artificial ligands

The knowledge of the exact binding motif of GM1 with CT could allow a rational design, development and synthesis of artificial ligands for the toxin. These molecules could possible function as

specific drugs by inhibiting the coordination of the B₅ pentamer with the human intestinal epithelium. In principle, a good artificial ligand should show a 3D-structure close to GM1 for those parts that are fundamental for the binding. Moreover, it should be synthetically simple for those regions that are not in contact with the protein.

Here, I will present two papers and a manuscript that refer to the rational design of new CT binders, with a detailed study of their conformations both in free and bound state, using NMR and molecular modeling.

References

- De Wolf, M. J., M. Fridkin and L. D. Kohn (1981) "Tryptophan residues of cholera toxin and its A and B protomers. Intrinsic fluorescence and solute quenching upon interacting with the ganglioside GM1, oligo-GM1, or dansylated oligo-GM1" *J Biol Chem* **256**(11): 5489-96.
- Finkelstein, R. A. (1988) *Immunochemical and Molecular Genetic Analysis of Bacterial Pathogens*, New York, Elsevier (eds Owen, P. & Foster, T. J.).
- Jobling, M. G. and R. K. Holmes (1991) "Analysis of structure and function of the B subunit of cholera toxin by the use of site-directed mutagenesis" *Mol Microbiol* **5**(7): 1755-67.
- Merritt, E. A., S. E. Pronk, T. K. Sixma, K. H. Kalk, B. A. van Zanten and W. G. Hol (1994) "Structure of partially-activated E. coli heat-labile enterotoxin (LT) at 2.6 Å resolution" *FEBS Lett* **337**(1): 88-92.
- Merritt, E. A., S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial and W. G. Hol (1994) "Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide" *Protein Sci* **3**(2): 166-75.
- Merritt, E. A., T. K. Sixma, K. H. Kalk, B. A. van Zanten and W. G. Hol (1994) "Galactose-binding site in Escherichia coli heat-labile enterotoxin (LT) and cholera toxin (CT)" *Mol Microbiol* **13**(4): 745-53.
- Sixma, T. K., K. H. Kalk, B. A. van Zanten, Z. Dauter, J. Kingma, B. Witholt and W. G. Hol (1993) "Refined structure of Escherichia coli heat-labile enterotoxin, a close relative of cholera toxin" *J Mol Biol* **230**(3): 890-918.
- Sixma, T. K., S. E. Pronk, K. H. Kalk, B. A. van Zanten, A. M. Berghuis and W. G. Hol (1992) "Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography" *Nature* **355**(6360): 561-4.
- Sixma, T. K., S. E. Pronk, K. H. Kalk, E. S. Wartna, B. A. van Zanten, B. Witholt and W. G. Hol (1991) "Crystal structure of a cholera toxin-related heat-labile enterotoxin from E. coli" *Nature* **351**(6325): 371-7.

Mimics of ganglioside GM1 as cholera toxin ligands: replacement of the GalNAc residue †

Anna Bernardi,^{*a} Daniela Arosio,^a Leonardo Manzoni,^b Diego Monti,^b Helena Posteri,^a Donatella Potenza,^a Silvia Mari^a and Jesús Jiménez-Barbero^c

^a Università di Milano – Dipartimento di Chimica Organica e Industriale e Centro di Eccellenza CISI, via Venezian 21, 20133 Milano, Italy

^b CNR-Istituto di Scienze e Tecnologie Molecolari (ISTM), Via Golgi 19, 20133 Milano, Italy

^c Dept. de Estructura y Función de Proteínas, Centro de Investigaciones Biológicas CSIC, Velázquez 144, 28006 Madrid, Spain

Received 28th October 2002, Accepted 13th December 2002

First published as an Advance Article on the web 5th February 2003

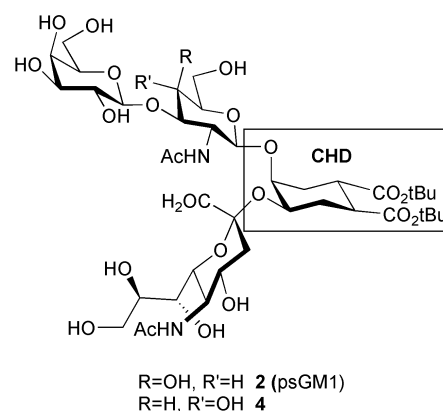
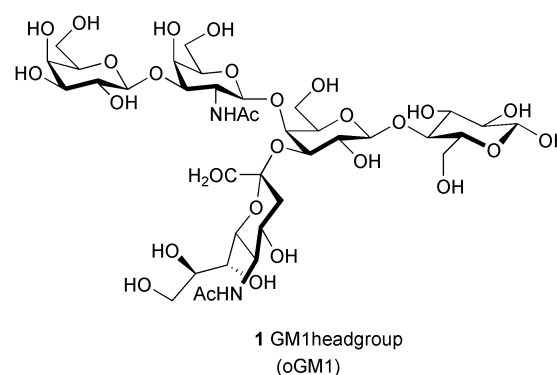
Two new cholera toxin (CT) ligands (**4** and **5**) are described. The new ligands were designed starting from the known GM1 mimics **2** and **3** by replacement of their GalNAc residue with the C4 isomer GlcNAc. As predicted by molecular modelling, the conformational properties of the equivalent pairs **2–4** and **3–5** are very similar and their affinity for CT is of the same order of magnitude. NMR experiments have also proved that **5** occupies the GM1-binding site of the toxin and have revealed its bound conformation.

Introduction

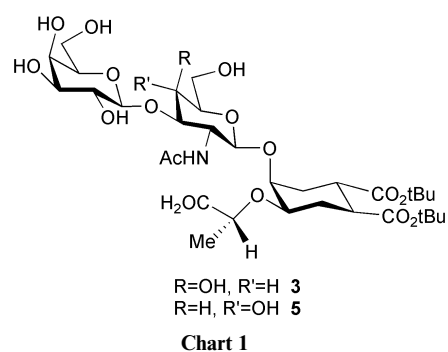
Carbohydrate–protein complexes are formed in the initial steps of a large number of physiological and pathological processes, which range from cell–pathogen interaction, to cell–cell recognition, to tumor metastasis, *etc.*¹ Interference with these recognition events could be used to modulate or alter signal transmission, or to prevent the onset of diseases. Therefore the synthesis of functional sugar mimics capable of antagonising oligosaccharides at the protein receptor level has attracted a great deal of attention as a way to develop drugs with good stability and synthetic availability.²

Our group is currently involved in an effort to devise a rational approach to the design and synthesis of glycomimetics by scaffold replacement.³ In brief, we have proposed the use of conformationally stable cyclic diols⁴ to replace non-pharmacophoric parts of bioactive oligosaccharides, while preserving the correct pharmacophore orientation. Computational tools can be used to predict the three-dimensional structures of the mimics and compare them with the structure of the natural ligand. When supported by adequate experimental work, molecular modelling also makes it possible to obtain at least qualitative predictions on the binding mode of new substrates, and to design further simplifications of the glycomimetic structures aimed at reducing the carbohydrate-likeness of the mimics and increasing their drug-like properties.

This approach has been validated⁵ using as a model system the recognition pair composed of the head-group of ganglioside GM1 (**1**) (Chart 1) and the two bacterial enterotoxins (cholera toxin (CT) and heat-labile toxin of *E. coli* (LT)) that use it as their target on cell membranes. GM1 interacts with CT and LT using the Gal and NeuAc residues at the oligosaccharide non-reducing end, as has been shown by biochemical⁶ and structural⁷ studies. Not long ago, we described the rational design and the synthesis of the pseudo-oligosaccharide **2**^{5,8} (Chart 1), a functional and structural mimic of ganglioside GM1 based on the use of a conformationally restricted cyclohexanediol (CHD)³ to replace the reducing end of the ganglioside head-group, which is not involved in toxin binding. Mimic



R=H, R'=OH **4**



R=H, R'=OH **5**

Chart 1

† Electronic supplementary information (ESI) available: synthetic details, product characterisations and full NOE contact list. See <http://www.rsc.org/suppdata/ob/b2/b210503a/>

2 was found to be as active as GM1 in binding to CT.⁵ More recently, we reported three second-generation mimics⁹ that were conceived by replacing the sialic acid (NeuAc) moiety of **2** with simple α -hydroxyacids. The most active compound of this series was the (*R*)-lactic acid-containing ligand **3** (Chart 1).

In an effort to further simplify the synthetic complexity of the structure we also examined the pseudo-tetrasaccharide **4** (Chart 1), which is derived from **2** by replacing the GalNAc residue with an *N*-acetylglucosamine (GlcNAc). Modification of the hexosamine was considered because it would simplify the synthesis of the artificial receptors (GalNAc is actually made from GlcNAc), while not negatively interfering with the formation of the toxin complex, since the GalNAc of GM1 interacts with the protein only *via* the *N*-acetyl group. Furthermore, the 4-hydroxy group of GalNAc in the experimental CT–GM1 complex⁷ and in the computational LT–GM1¹⁰ and LT–**2**⁵ models is located outside the protein binding pocket and fully exposed to the solvent. Assuming that **4** will bind to CT and/or LT with the same general mode of GM1 and **2**, it appeared likely that inversion of configuration at C4 of the hexosamine could allow new H-bond contacts to be formed between the toxin and the substrate. In principle, these may or may not compensate for the loss in complex solvation due to the burying of the hydroxy group in the binding cavity. To evaluate the potential of **4** as a GM1 antagonist, computer simulations were used to model the free ligand **4** and its LT complex, followed by comparison with the corresponding structures of **2**.¹¹

The following predictions were made:¹¹

1. Changing the hexosamine in the pseudo GM1 structure should not modify the overall molecular conformation, and, in particular, should not alter the relative position of the Gal and NeuAc binding determinants.

2. Compared to **2**, the new compound **4** will indeed insert one more hydroxy group within the protein binding site. Molecular dynamics simulations suggested that, in turn, this may trigger a series of rearrangements and reorientations of side chains and crystallographic water molecules in the toxin, leading to new H-bond contacts which may result in enhanced affinity of the new inhibitor.

3. Free Energy Perturbation (FEP) calculations performed by carrying out the **2** \rightarrow **4** mutation in solution and in the protein complex suggested that the GlcNAc mimic **4** should be a stronger binder than its parent compound **2**.

We now report the experimental data relative to the GlcNAc-containing mimics **4** and **5**, the latter being the GlcNAc-containing version of the second generation binder **3**. The ligands have been synthesised and their solution structure has been studied by NMR spectroscopy, confirming the computational predictions.¹¹ The complex formed by **5** with the recognition element of CT, CTB5, has been studied by TR-NOE NMR, and the bound conformation of the ligand has been determined. Competition experiments, carried out by adding oGM1 to a solution of the CTB5–**5** complex, could also be monitored by NMR, and have confirmed that the GlcNAc-containing ligand and the ganglioside are indeed competing for the same binding site. Affinity constants for CTB5 have been obtained for **4** and **5** by fluorescence titrations, and found to be similar to those of the parent compounds **2** and **3**, respectively. Thus, the computational predictions on the GlcNAc binders were revealed to be qualitatively correct, but the FEP calculations were quantitatively incorrect.

Results

Synthesis of the ligands

The GlcNAc-containing ligands **4** and **5** were prepared following the synthetic scheme adopted for the parent compounds **2**⁸ and **3**⁹ and reported in Scheme 1.

The full synthetic sequence and product characterisations are reported in the Supplementary Information. † Briefly, the known acceptors **6**⁸ and **7**⁹ (Scheme 1) were glycosylated at the axial hydroxy group using the trichloroacetimidate **8** as a Gal β -1,3-GlcNAc donor, and TfOH or TMSOTf as the promoter. The adducts were routinely deprotected to yield the free ligands.

NMR studies of the free ligands **4** and **5**

Two-dimensional 400-MHz NOESY and ROESY spectra of **4** and **5** were obtained in D₂O. The relevant inter-residual contacts are collected in Table 1, and compared with the data reported in the literature for **1**,¹² **2**⁵ and **3**¹³ in D₂O. Complete spectral assignments and full contact lists are collected in the Supplementary Information. †

The NMR data obtained for **4** (Table 1, column 4) show the characteristic set of NOE cross-peaks (Fig. 1b) which is also observed for **1**¹² and **2**⁵ (Table 1, columns 1 and 2; Fig. 1a). The data are consistent with one major conformation of the oligosaccharide framework featuring the Gal β -1,3-GlcNAc β -1,4-

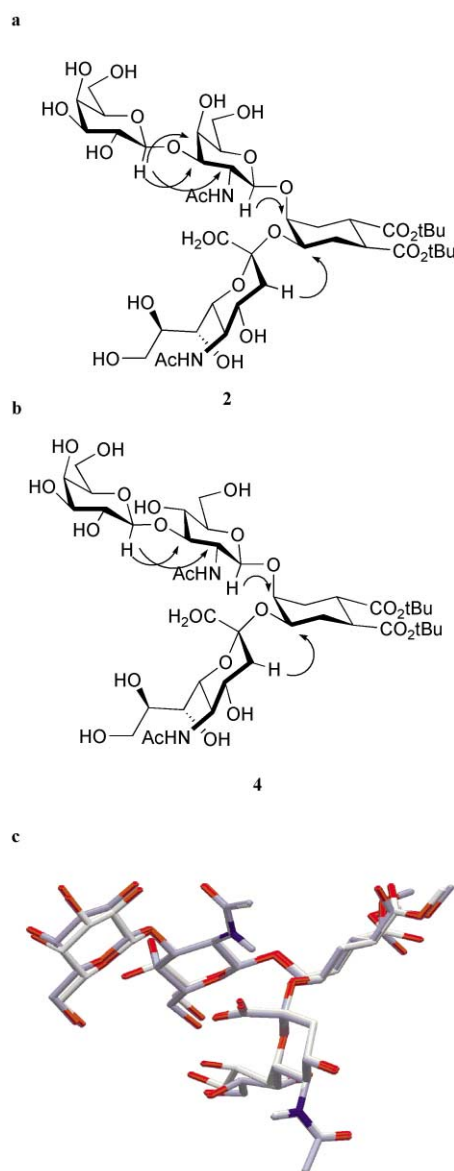
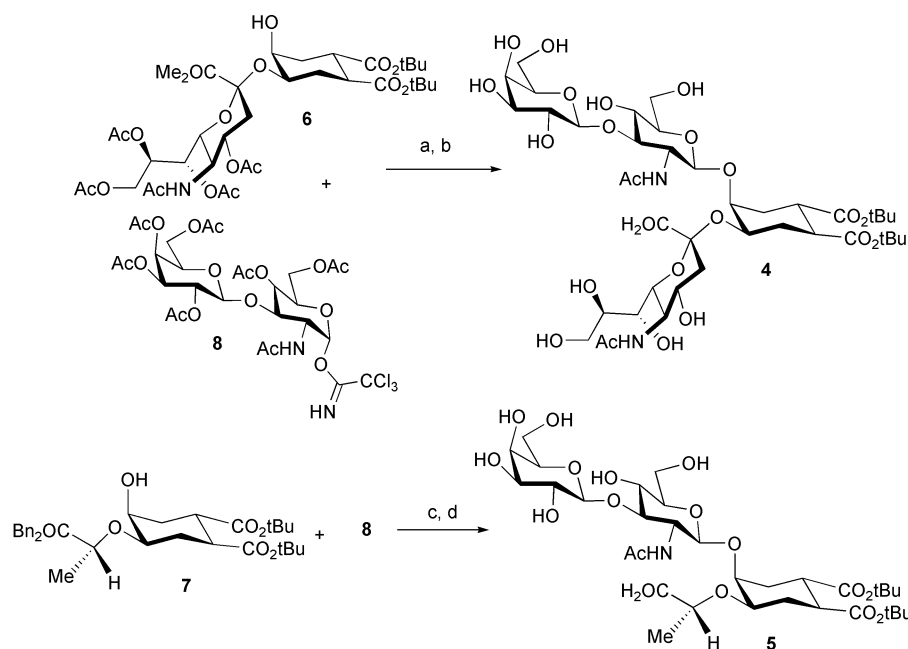


Fig. 1 a. Observed NOE contacts in the NMR spectra of **2**. b. Observed NOE contacts in the NMR spectra of **4**. c. Overlap of the calculated structures of **2** (light blue) and **4** (white). For the Gal β -1,3-GalNAc β -1,4-CHD pseudo-trisaccharide the glycosidic angles ϕ and ψ are defined as H1–C1–O–Cn and C1–O1–Cn–Hn, respectively. For the NeuAc α -2,3-CHD fragment ϕ and ψ are defined as (O=C1)C1–C2–O–Cn and C2–O–Cn–Hn, respectively.

Table 1 Relevant NOE contacts in ligands 1–5^a

| Residue | 1 ^b | 2 ^c | 3 ^d | 4 ^e | 5 ^e free state | 5 ^e CTB5 complex |
|--------------------------|----------------|---|--|--|------------------------------------|--|
| NeuAc [N] | N-3ax N8 | GII-3 (s) GN1 (m) ^f | CHD3 (s/m) ^g CHD4 (w) | — | CHD3 (s/m) ^g | — |
| (<i>R</i>)-lactic acid | H _L | — | — | CHD3 (s) CHD4 (s) CHD2eq (m) GN1 (w) CHD3 (s) GN1 (m) | — | CHD3 (s) CHD4 (s) CHD2eq (m/w) GN1 (w) CHD3 (s) GN1 (m) |
| | Me | — | — | — | — | — |
| GalNAc [GN] | GN1 | GII-4 (s) | CHD4 (s) | CHD4 (s) | — | — |
| GlcNAc [GN] | GN1 | — | — | — | CHD4 (s) | CHD4 (s) |
| Gal [G] | G1 | GN3 (s) ^f GN2 (w) GN4 (w) ^f | GN3 (s) GN2 (w) GN4 (w) | GN3 (s) GN2 (w) GN4 (w) | GN3 (s) ^h GN2 (w) | GN3 (s) — GN3 (s) |

^a ROESY cross-peaks in D₂O solution. In parentheses the observed intensities; s: strong, m: medium, w: weak. ^b From ref. 12. ^c From ref. 5. ^d From ref. 13. ^e This work. ^f Data from DMSO-d₆ solution (ref. 31); not seen in D₂O, due to signal overlap. ^g N8–GN1 not measurable, due to signal overlap. ^h GN2 overlaps with GN3.



a. TMSOTf (0.4 eq), CH₂Cl₂, reflux (13%); b. 0.002 M MeONa in MeOH, then H₂O (92%); c. TfOH (0.2 eq), CH₂Cl₂, r.t., then reflux (20%); d. H₂, Pd-C, then b (70% over the two steps).

Scheme 1 Synthesis of 4 and 5.

CHD pseudo-trisaccharide and the NeuAc α -2,3-CHD fragment exclusively in the *syn*, *syn* (ϕ , ψ 55°, 0°) and the *anti*, *syn* (ϕ , ψ -170°, -30°) conformations, respectively (for the definition of the ϕ , ψ inter-glycosidic torsion angles, see the caption of Fig. 1). A marked conformational restriction is a distinctive feature of the oGM1 pentasaccharide and its mimics. In general, all experiments find that the ganglioside head-group can be broken down into two areas: a so-called “core trisaccharide”, the GalNAc β -1,4(Neu5Ac α 2-3)Gal trisaccharide which is often described as “rigid”, and more mobile regions corresponding to the external sugars at both ends of the head-group. However, also the mobility of the Gal β -1,3GalNAc fragment at the non reducing end is actually limited to ample oscillations around well-defined average glycosidic torsion angles. Dynamic simulations of the oligosaccharide suggest that the NMR data showing two weak and equivalent NOE contacts between Gal-H1 and GalNAc-H2 and H4 (Table 1, column 1; see Fig. 1 for the

equivalent contacts in 2) are accounted for by a model whereby the Gal β -1,3GalNAc bond populates a single, broad energy minimum, rather than two individual conformations.¹⁴ The energy well is centered at ϕ , ψ 55°, -5°, which is also the conformation observed in the CTB5–oGM1 X-ray structure (ϕ , ψ 55° \pm 10°, 0° \pm 10°).⁷ Therefore it can safely be asserted that oGM1 exists largely as one main conformer. The same behaviour has been observed for 2, which assumes the conformation depicted in Fig. 1c (light blue structure).⁵ In 4, the flexibility of the Gal β -1,3-GlcNAc β -1,4-CHD fragment is further reduced by the H-bond interaction of GlcNAc-OH4 with the Gal oxygen, which constrains the amplitude of the oscillations closer to the observed bound conformation. In fact, the conformational analysis of 4¹¹ led to 18 conformers within 10 kJ mol⁻¹ from the global minimum that all shared a common conformation of the core fragment. This conformation is completely superimposable with the NMR solution structure of 2,

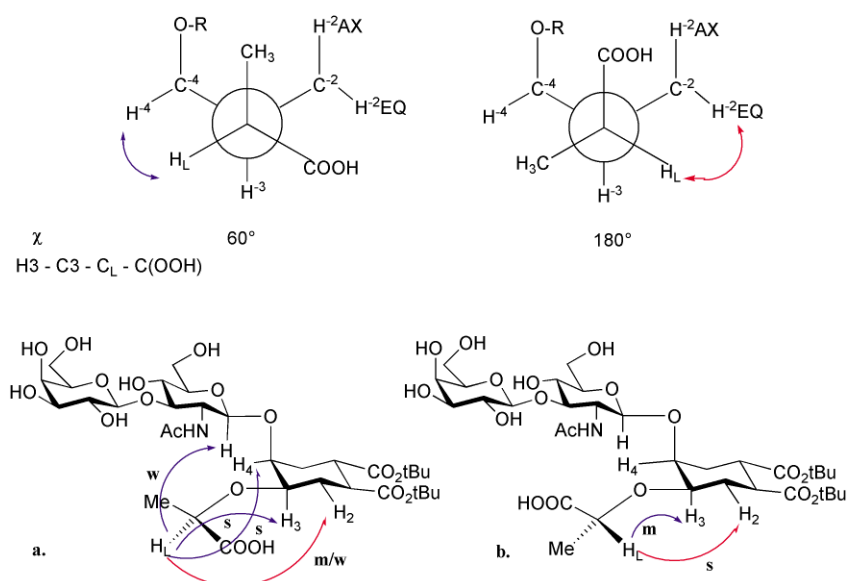


Fig. 2 Experimentally observed side-chain conformations in the (*R*)-lactic acid-containing ligand **5**. The intensity of the observed NOE cross-peaks is indicated: w:weak; s:strong; m/w:medium/weak. **a.** Free state. The mutually exclusive NOE contacts H_L/H_4 and H_L/H_{2eq} reveal the existence of two different conformations in water solution. The conformation depicted is the prevalent one. **b.** Bound state. The single conformation selected by the toxin is not the most abundant one in the free state.

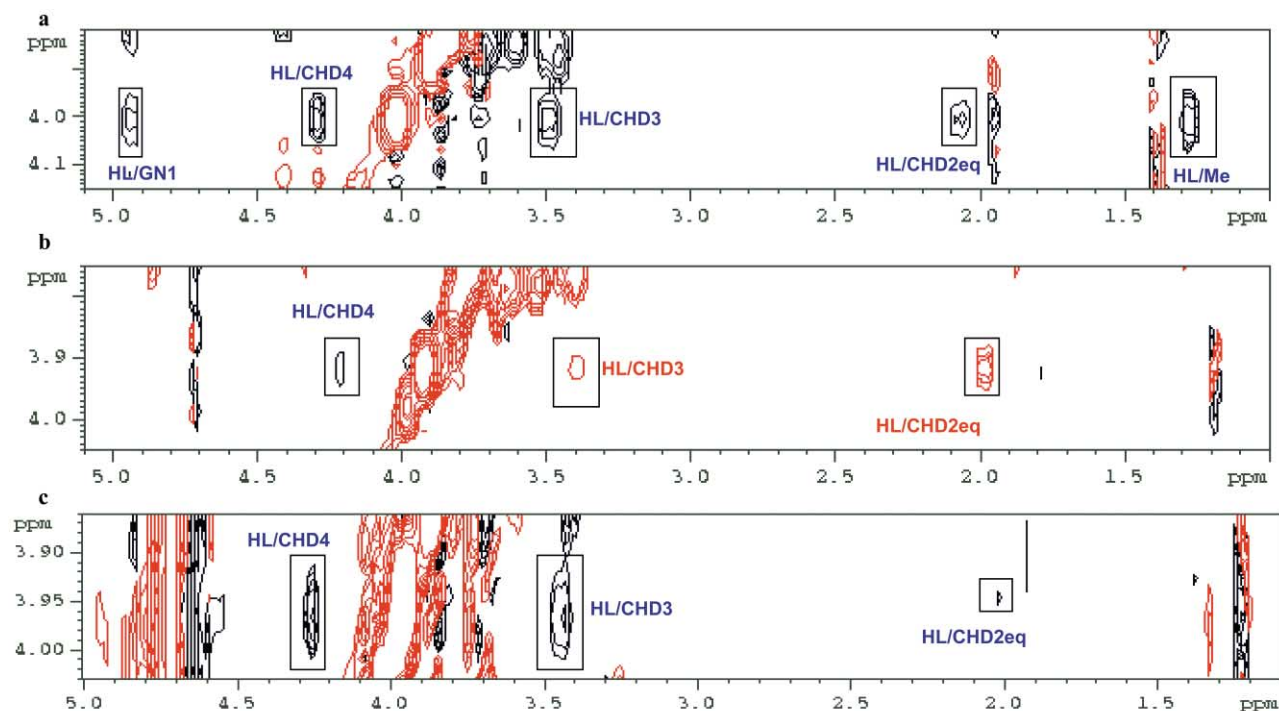


Fig. 3 NMR spectra of **5**. **a**) NOESY spectrum of **5** in the free state. **b**) TR-NOESY spectrum of **5** in the CTB5 complex. **c**) TR-NOESY spectrum of the CTB5-**5** complex after adding oGM1. Blue cross-peaks are due to the free state form, red cross-peaks are TR-NOE signals arising from the bound state of **5**.

which also corresponds to the bound structure observed in the CT-GM1 complex (Fig. 1c). The experimental data that are now available for **4** confirm the computational prediction by showing exclusively the expected pattern of NOE contacts (Fig. 1b).

For the (*R*)-lactic acid-containing ligand **5** the NMR data (Table 1, column **5**, free state) are once more consistent with a *syn, syn* disposition of the Gal β 1-3GlcNAc β -1,4-CHD pseudotrisaccharide, but show a substantial flexibility of the (*R*)-lactic acid side-chain. At least two conformations are represented with χ 60° (major, Fig. 2a) and χ 180° (minor), where χ is the improper dihedral angle H3-C3-C_L-C(=O) (see Fig. 2) defining the position of the carboxylate relative to the cyclohexanediol ring. This is inferred by the presence of two mutually exclusive

NOE contacts from H_L to CHD4 and CHD2eq (Fig. 3a), which should arise from the χ 60° and χ 180° conformations (Fig. 2), respectively. The same pattern of NOE contacts was also observed for the GalNAc derivative **3**¹³ (Table 1, column **3**).

Thus, the NMR experiments for the unbound ligands suggest that the GalNAc-containing and the GlcNAc-containing molecules are indeed adopting identical conformations, regardless of the nature of the hexosamine.

NMR studies of the CT-**5** complex

Information about the conformation of complexed ligands can be derived from transferred nuclear Overhauser effect (TR-NOE) studies,¹⁵ provided that the exchange between the

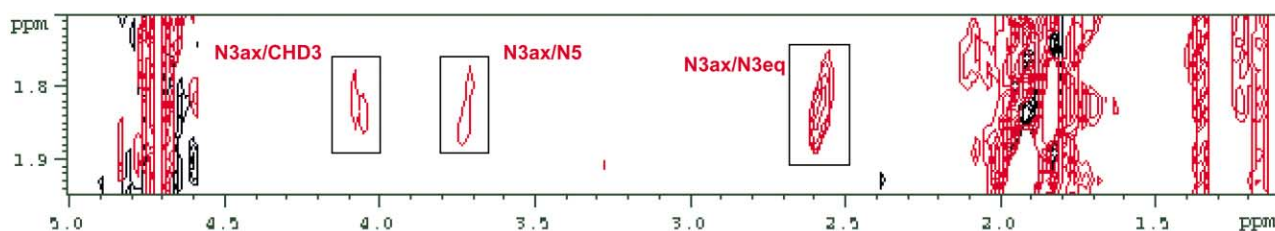


Fig. 4 TR-NOESY spectrum of the CTB5-5 complex after addition of the GM1 oligosaccharide **1** (oGM1). In-phase (red) TR-NOESY cross-peaks belonging to the H_{3ax} of the NeuAc residue in oGM1 (N3ax) show that oGM1 is bound to CT.

complexed and uncomplexed states is sufficiently fast.¹⁶ When a ligand interacts with a macromolecule, the cross-relaxation rates of its bound states become opposite in sign to those of the free ligand and generate negative NOEs. So, in the TR-NOESY experiment the cross-peaks which are due to bound species are in phase with the diagonal signals, whereas unbound species give rise to cross-peaks which are out of phase. The bound form in-phase signals can be exploited to assess the conformational properties of the bound state. The TR-NOE technique allows the study in solution of how a ligand fits into a protein binding site. The phase behaviour of NOESY cross-peaks can be used to analyze the affinity of individual ligands within mixtures of compounds,¹⁷ and can also be exploited for competition experiments.¹⁸ Our group has recently used it to study the complexes formed by CTB5 and a group of GM1 mimics, including **3**.¹³

The affinities of **1**,¹⁹ **2**⁵ and **4** (see below) for CTB5 are on the upper limit of what can be revealed by the TR-NOE technique,²⁰ but the study of the CTB5-5 complex yielded much important information. No difference is observed between the free-state and bound-state conformations of the Gal-GlcNAc and GlcNAc-CHD fragments of **5**, which are always found to be *syn* (ϕ , ψ 55°, 0°; see Table 1, column 5, CTB5 complex). In contrast, a clearly different set of cross-peaks was observed for the NOEs involving the hydroxyacid protons H_L and CH₃ (see Table 1) upon passing from the free to the bound state. In particular, the cross-peaks observed for H_L in both free and bound states are shown in Fig. 3a and 3b, respectively. The out-of-phase NOESY cross-peaks relative to the free state of **5** are depicted in blue, whereas the in-phase TR-NOESY peaks arising from its bound state are shown in red, like the diagonal signals. The free state spectrum (Fig. 3a) shows the set of NOE contacts (blue cross-peaks) discussed above, and sketched in Fig. 2a (CHD3 and CHD4 strong, CHD2eq m/w). The TR-NOESY spectrum obtained upon addition of CTB5 to the solution is shown in Fig. 3b (see also Table 1). The strong H_L/CHD-H4 cross-peak found in the free state NOESY of **5** (Fig. 3a) almost disappears in the TR-NOESY spectrum of the CTB5-5 complex, and the small residual signal is still out-of-phase (blue), therefore it still belongs to the free state of the ligand. The in-phase (red) cross-peaks belonging to the bound state of **5** show a close H_L/CHD-H2eq contact, much stronger than in the free state. This implies a conformational transition of the hydroxyacid side-chain, with the χ 180° conformer (Fig. 2b) being selected in the bound state preferentially with respect to the most abundant χ 60° conformer (Fig. 2a). The behaviour of **5** described here closely parallels the observations already reported for **3**,¹³ and is strongly suggestive of a common binding mode for the two ligands in the toxin.

Finally, a competition experiment was performed by adding oGM1 to the CTB5-5 complex (Fig. 3c). oGM1 has a much higher affinity for CT than **5** (see below), therefore in a competition experiment it is expected to bind selectively to the toxin and displace **5** from its binding site. Indeed, the cross-peaks of **5** which are in phase (red) in the TR-NOESY spectrum of its CT complex (e.g. H_L/CHD3 and H_L/CHD2eq in Fig. 3b) become out-of-phase (blue) upon addition of oGM1 to the mixture (Fig. 3c), and the relative intensity of the H_L/CHD4 and

H_L/CHD2eq cross-peaks typical of the free state of **5** (CHD4 s, CHD2eq m/w) is also restored. Thus, it clearly appears that **5** is no longer bound to the toxin. Furthermore, the NeuAc-H3eq/Gal-H3 NOE contact typical of GM1 (see Table 1) is observed (Fig. 4) as an in-phase (red) cross-peak in the same spectrum, proving that GM1 has indeed displaced **5** from the toxin binding site. This experiment clearly establishes that oGM1 and **5** are competing for the same binding site in CTB5, as expected on the basis of the computational results.

Binding studies

The interaction of cholera toxin B5 pentamer (CTB5) with Gal-containing ligands can be studied using the intrinsic fluorescence of the Trp88 residue in the toxin binding site.²¹ Binding of **1-3** to CTB5 induces bathochromic shifts in the emission spectrum whose extent depends on the ligand.⁹ oGM1 binding is also accompanied by an increase in the intensity of fluorescence emission, whereas a variable decrease in fluorescence is seen upon binding of **2** (small decrease) and **3** (larger decrease). Surprisingly, when 0.5 μ M CTB5 was titrated with the GlcNAc-containing ligands **4** and **5** no bathochromic shift occurred. However, a small decrease in fluorescence was observed with **4** and a larger one with **5** (see Supplementary Information, Figs. S1 and S2). † The binding isotherms of GalNAc-containing and GlcNAc-containing ligands were therefore compared by analysing the intensity data. This requires a relatively high CTB5 concentration (0.5 μ M). However, wavelength titrations performed for **1** and **2** at lower toxin concentrations (0.1 μ M) gave comparable results (See Supplementary Information, Fig. S3). †

Binding of **1** to CTB is known to occur cooperatively.¹⁹ The observed concentration of the ligand at 50% saturation (IC₅₀) in our titration (0.5 μ M CTB5, see Fig. 5) is 1.5 μ M. Under the same conditions, the intrinsic association constant, determined by calorimetric titration, is 1.05×10^6 M⁻¹.¹⁹ The first generation mimic **2** was also found to bind cooperatively and with comparable affinity (IC₅₀ 1.2 μ M, see Fig. 5).⁹ Binding of the

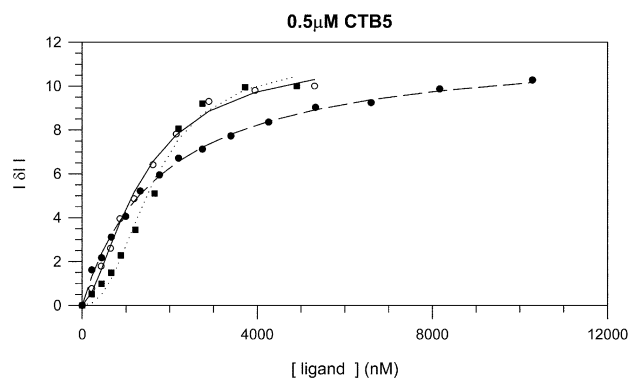


Fig. 5 Fluorescence intensity titrations of CTB5 (0.5 μ M) with o-GM1 **1** (black square, dotted line); ps-GM1 **2** (empty circles, solid line); the GlcNAc-containing analogue **4** (black circles, broken line). The absolute values of the variation of fluorescence intensity emission at 350 nm ($|\Delta I|$) have been normalised to 10, and plotted against the total concentration of the ligands (nM).

GlcNAc analogue **4** fits a simple 1 : 1 isotherm and doesn't show any cooperativity (Fig. 5), the IC_{50} is 1.3 μ M and the dissociation constant, determined by nonlinear regression analysis, is 1.8 μ M. Although quantitative comparison is complicated by the different cooperativity behaviour, the relative binding affinities of **1**, **2**, and **4** toward CTB5 can be appreciated by comparing the corresponding binding isotherms, collected in Fig. 5 and showing that the three ligands clearly appear to be of comparable potency.

The two lactic acid-containing compounds **3** and **5**, both binding CTB5 without any cooperativity effect, allow an easier estimation of relative affinity. The binding isotherms obtained in a titration of 0.5 μ M CTB5 are reported in Fig. 6. The dissociation constants calculated by non-linear regression analysis are 190 and 450 μ M for **3** and **5**, respectively.

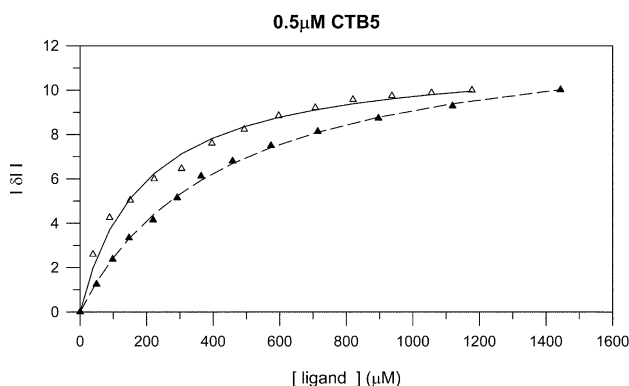


Fig. 6 Titrations of CTB5 0.5 μ M with **3** (empty triangles, solid line) and **5** (black triangles, broken line). The absolute values of the variation of fluorescence intensity emission at 350 nm have been normalised to 10, and plotted against the total concentration of the ligands (μ M). The dissociation constants calculated by non-linear regression analysis are 190 and 450 μ M, respectively.

Thus, it appears that, compared to the parent ganglioside and pseudo-gangliosides, the GlcNAc compounds **4** and **5** are indeed showing a comparable affinity for CT, as expected based on the conformational analysis and molecular dynamic simulation of the toxin–**4** complex.¹¹ However, contrary to the FEP predictions,¹¹ **4** is not more active than **2**, nor does **5** appear to be any more active than **3**.

Discussion

The complex formed by ganglioside GM1 [Gal β 1-3GalNAc β 1-4(NeuAc α 2-3)Gal β 1-4Glc β 1-1Cer] and the two AB5 enterotoxins CT and LT is one of the best characterised protein–carbohydrate pairs. Both toxins recognise the pentasaccharide head-group of ganglioside GM1 (o-GM1, **1**) on the host cell epithelial surface using the B5 pentamer. The pentamer has a characteristic, doughnut-like shape, with a central pore. Once the toxin is attached to the membrane, a fragment of the A monomer moves through the pore and is finally translocated through the host cell membrane into the cytosol. Biochemical⁶ and structural¹⁷ data have shown that there are five binding sites in the toxin, and that the ganglioside binds to them using the two terminal sugars at its non-reducing end, a galactose (Gal) and a sialic acid (NeuAc). Binding of oGM1 to CT displays positive cooperativity, as determined by calorimetric titrations,¹⁹ and in biological settings the interaction between the B5 pentamer and several membrane-bound GM1 molecules is a text-book case of high-affinity multivalent interaction.²²

We have shown that it is possible to replace the reducing end lactose of oGM1 with an appropriate diol as in psGM1 **2** with no loss of affinity or cooperativity effect.⁵ The sialic acid can also be replaced with various hydroxyacids, as in the lactic acid derivative **3**.⁹ In this case, the amount of affinity retained varies

with the nature of the hydroxyacid side-chain,²³ but in all the analogues we have studied so far cooperativity is lost.^{9,23}

Much less is known about the role played by the hexosamine. In the X-ray structure of the CTB5–oGM1 complex the GalNAc residue does not exhibit any directed or mediated H-bond interaction with the protein, but the acetamide methyl group makes a van der Waals contact with the C β of His13.⁷ There is no natural glycolipid which contains a Gal β 1,3-GlcNAc β 1,4-(NeuAc α 2,3-)Gal fragment, and, to the best of our knowledge, no binding data are available for this oligosaccharide or for larger entities that contain it. In a previous paper,¹¹ we evaluated the potential of **4**, a GlcNAc analogue of psGM1 **2**, to behave as a GM1 antagonist using computer simulations to model the free ligand **4** and its LT complex. On the basis of this computational work it was expected that exchanging GalNAc with GlcNAc would yield a new molecule with the same overall molecular conformation of **2**, capable of fitting the GM1-binding site of CT and displaying a higher affinity than psGM1.

The experimental results described above concerning **4** and its lactic acid analogue **5** have now shown that the prediction was qualitatively correct, in particular:

1. NMR experiments have shown that the conformation of GlcNAc-containing ligands **4** and **5** in the free state and in the CTB5–**5** complex do not differ from those observed for the GalNAc substrates **2**⁵ and **3**¹³ (free state) and CTB5–**3**¹³ complex.

2. Competition experiments using the TR-NOESY technique have unequivocally shown that **5** binds into the GM1-binding site of CTB5, from which it can be displaced by the higher-affinity natural ligand oGM1.

3. Fluorescence emission titrations of CTB5 have shown that the affinities of the GlcNAc-containing ligands **4** and **5** are of the same order of magnitude as those measured for the corresponding GalNAc-containing compounds **2** and **3**. However, in contrast with the behaviour of **2**, binding of **4** to CTB5 is not cooperative. Furthermore, the increase of affinity predicted by FEP calculations¹¹ in the **2** to **4** mutation is not borne out by experiment.

The structural reasons leading to the cooperative behaviour of GM1 binding to CT are unclear, thus it is very difficult to understand what leads to the loss of cooperativity in some of the psGM1 analogues. Computational analysis (let alone prediction) of cooperativity in systems of this size is currently unfeasible, so at this time we can only try a speculative interpretation of the available experimental data. Based on X-ray structures of bound and unbound CT and LT, the main structural effect that sugar binding has on these toxins is a tightening of the 51–60 loop of the B subunits, a loop that connects beta strand β 4 to helix α 2 of the B monomer.²⁴ The position of the α 2 helix controls the size of the pore in the B5 doughnut,²⁵ and thus it may regulate translocation of the A fragment through the pore and ultimately through the host cell membrane. Through the α 2 helix, carbohydrate binding to one site could in principle be signalled to the adjacent B monomer. However, this same tightening is elicited by sugars that are cooperative binders, such as GM1, and sugars that are non-cooperative binders, such as galactose.²⁴ When we first determined the affinity of **3** and other analogues obtained by sialic acid replacement,⁹ we found that none of them displayed cooperative binding and speculated that the communications between CTB monomers elicited by **1** and **2** may be mediated by the NeuAc residue. The NeuAc side-chain, in fact, is in contact with the Gly-33 residue of the B+1 monomer through one water molecule (W2) located at crystallographic site two in the CT–GM1 complex.⁷ This molecule is not seen in the known X-ray structures of isolated enterotoxins,²⁴ but it is present in many²⁴ (but not all)²⁶ of the known toxin–ligand structures. Neither of the new GlcNAc-containing mimics **4** and **5** appears to exhibit cooperative behaviour in binding to CTB5,

independent of the presence of the sialic acid moiety (see Fig. 6), apparently suggesting that NeuAc is not the cooperativity determinant. However, it should be noted that during the dynamics simulations of the LT–4 complex¹¹ a displacement of the water molecule W2 was observed from the position it occupies in CT–1⁷ and LT–2¹¹ and toward the 4-hydroxy group of the GlcNAc residue of 4. Thus, in the LT–4 complex W2 loses its interaction with Gly-33(B+1) and participates in a different H-bond network.¹¹ On the basis of this dynamics simulation and of the above speculation on inter-monomer communication, it is still possible that the cooperativity behaviour observed for 1 and 2 is indeed determined by the sialic acid side-chain *via* the W2 water molecule. The latter, however, is displaced from its normal position by the GlcNAc-containing ligands and cannot transmit its signal to the adjacent (B+1) monomer through the Gly-33 residue. If this is not the case, more subtle effects must be operating in defining the cooperativity behaviour of ganglioside binding to CT.

Free Energy Perturbation (FEP) calculations have recently been systematically applied to a series of antibody–sugar complexes using GLYCAM, TIP3P water, RESP charges and stepwise perturbations.²⁷ The simulations were shown to reproduce reasonably the known geometries of ligand–protein complexes, while the calculated values of $\Delta\Delta G$ of binding were found to be qualitatively reproduced and to depend heavily on the choices made about the protonation state of an His located in the vicinity of the binding site. Our estimate of the $\Delta\Delta G$ of binding between 2 and 4 was obtained¹¹ with a similar set of parameters (GLYCAM, TIP3P water, ESP-derived charges on the carbohydrate atoms) and the mutation was performed in the 4 to 2 sense using the slow-growth algorithm rather than stepwise perturbations of the ligand. The energetic gain calculated for 4 (3.8 ± 1.9 kcal mol⁻¹) resulted from both a better solvation of the GalNAc ligand 2 in the free state, and a more favourable interaction of 4 with the protein. Since, as we have discussed above, the pseudo-ganglioside oligosaccharides experimentally display a limited internal flexibility,^{5,13,14} incomplete sampling of the free ligands doesn't appear to be a likely explanation of the rather large error in the relative free energies computed. However this might not hold in the case of the protein-complex, where sampling of the protein's degrees of freedom is also important (*vide infra*). The use of LT rather than CT in the calculation was also taken into account as a possible source of error in the calculation. However, preliminary fluorescence titrations²⁸ obtained with a sample of the entire LT toxin²⁹ showed the same trends observed with CT, *i.e.* equivalent pairs of ligands (3 and 5, and 2 and 4) have similar affinity for LT as well as for CT.

The most likely explanation for the error in the computed relative free energies lies in the computational treatment of the pseudo-ganglioside–toxin complex part of the FEP. First, incomplete sampling of the complex simulation cannot be ruled out, particularly considering that a substantial amount of side-chain rearrangements occur in the protein binding site on passing from the LT–4 to the LT–2 complex. Furthermore, the state of protonation of His-57, one of the amino acids most involved in the rearrangement, was not addressed at all during the FEP simulation.

Although achieving a quantitative prediction of relative binding energies for oligosaccharide–protein complexes may well require a lot more effort, this work shows that computational tools can be used with success to design new inhibitors of carbohydrate–protein interaction, and that they yield qualitatively useful results. Indeed, at the start of this project, nothing could suggest that inversion of the hexosamine C4 in psGM1 2 would yield a ligand with good CT affinity. We have already noted that, to the best of our knowledge, the natural counterpart of 4 has never been described, and certainly has never been tested as a ligand for bacterial enterotoxins. The calculations correctly predicted that inversion of the hexo-

samine C4 in 2 would yield a new molecule with overall shape and conformational properties very similar to psGM1. It also allowed a prediction that the new molecule 4 would be able to interact with the cholera toxin in a way similar to its GalNAc analogue. The question remains open of how to reliably achieve quantitative prediction of relative binding affinities in this and similar systems.

In conclusion, two new artificial ligands of the cholera toxin, the pseudo sugars 4 and 5 are described. The new ligands were designed starting from the known GM1 mimics 2 and 3 by replacement of their GalNAc residue with the C4 isomer GlcNAc. Such substitution had been suggested by inspection of the CT–GM1 complex, and supported by computational predictions, which suggested that the three-dimensional shape of the new ligands and their mode of interaction with CT would be similar to those of the starting structures. These predictions are now confirmed by the experimental results showing that the conformational properties of the equivalent pairs 2–4 and 3–5 are indeed very similar and that their affinity for CT is of the same order of magnitude. NMR experiments have also allowed the gathering of information on the structure of the CT–5 complex showing that 5 occupies the GM1-binding site of the toxin and that it binds with a conformation similar to the one adopted by 3 in the CT–3 complex. Since GalNAc is normally synthesised by C4 inversion of GlcNAc, the GlcNAc ligands constitute a new class of GM1 mimics with improved synthetic accessibility.

Experimental section

Synthesis

The synthesis of compound 6 was described in ref. 8. The full synthetic sequence leading to 4 and 5 and the product characterisations are reported in the Supplementary Information.

NMR

NMR spectra of 4 and 5 were recorded at 25–30 °C in D₂O, on Varian Unity 500 MHz or Bruker AVANCE 400 MHz spectrometers. For the experiments with the free ligands, the corresponding compound (1–1.5 mg) was dissolved in D₂O and the solution was degassed by passing N₂. COSY, TOCSY and HSQC experiments were performed using standard sequences at temperatures between 298 and 310 K. NOESY experiments were performed with mixing times of 500, 700, 1000 ms (4) and 400, 600, 700, 800 ms (5). ROESY experiments³⁰ were performed with mixing times of 50, 100, 150, 200, 300 ms (4) and 100, 200, 250, 300 ms (5).

The cholera toxin CTB pentamer (CTB5) was purchased from List Biological Laboratories Inc. The commercial sample was ultrafiltered to remove EDTA and tris salt, redissolved in phosphate buffer and subjected to two cycles of freeze-drying with D₂O to remove traces of H₂O. The sample was then dissolved with D₂O, and the solution transferred to the NMR tube to give a final concentration *ca.* 0.1 mM. TR-NOESY experiments were performed with mixing times of 100, 200 and 300 ms, for a *ca.* 50 : 1 molar ratio of 5 : lectin. TR-ROESY experiments were also performed with mixing times of 100, 200, 250, 300 ms. In the competition experiment, 1.5 mg of oGM1 1 (gift from Professor Sandro Sonnino, Dipartimento di Chimica e Biochimica Medica, Università di Milano) were added to the same solution and the TR-NOESY spectra were recorded as described above.

Titration

Fluorescence titrations were performed with an LS50 Perkin Elmer fluorimeter, using a pH 7.5 tris buffer to dissolve CTB5 (0.5 μM) and ligands. Fluorescence from TRp-88 was measured with an excitation of 280 nm and an emission from 300 to 450

nm. The emission spectra are reported as Supplementary Information. The K_d reported in the text were determined by non-linear regression analysis, using Sigma Plot 2.0 (Jandel Corporation).

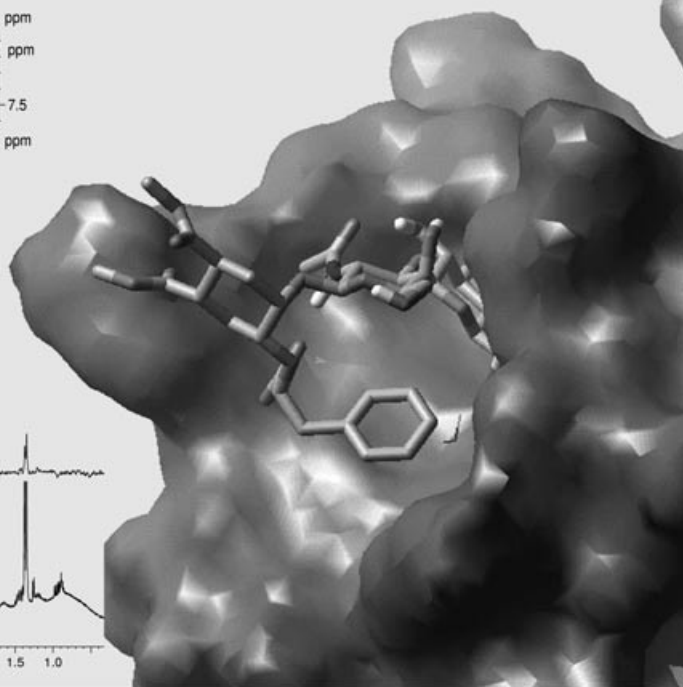
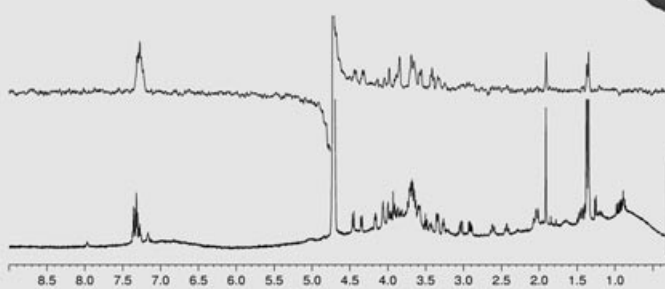
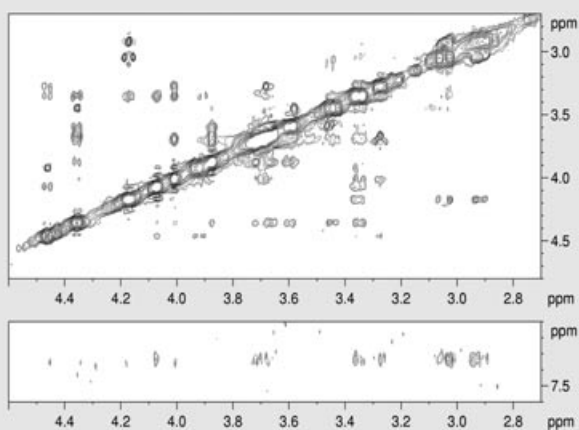
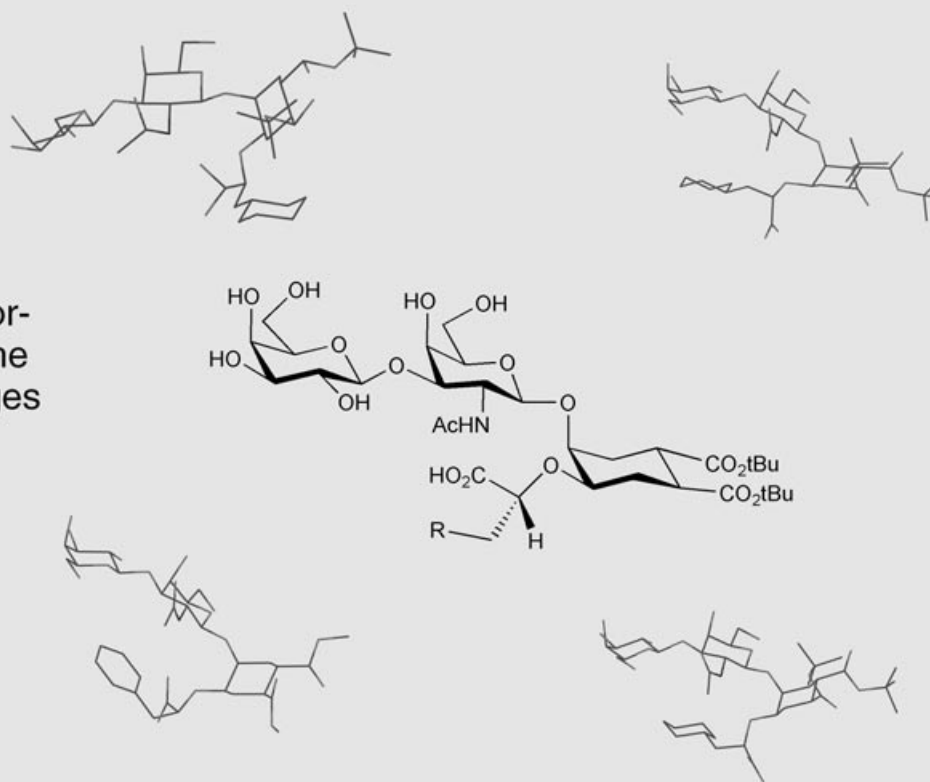
Acknowledgements

We thank Professor Sandro Sonnino (Dipartimento di Chimica e Biochimica Medica, Università di Milano) for a generous gift of oGM1, and for helpful discussions throughout this project. We thank Giorgio Colombo (IBMR, Milano) for carefully reading this manuscript and providing his comments. Funding was provided by MIUR, CNR, COST D-13/012, Azioni Integrate.

References

- 1 (a) H. Lis and N. Sharon, *Chem. Rev.*, 1998, **98**, 637–674; (b) R. A. Dwek, *Chem. Rev.*, 1996, **96**, 683–720; (c) A. Varki, *Glycobiology*, 1993, **3**, 97–130.
- 2 P. Sears and C. H. Wong, *Angew. Chem., Int. Ed.*, 1999, **111**, 2446–2471; P. Sears and C. H. Wong, *Angew. Chem., Int. Ed.*, 1999, **38**, 2300–2324 and references therein.
- 3 A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, S. Pasquarello and P. Seneci, *J. Org. Chem.*, 2001, **66**, 6209–6216.
- 4 *trans*-Cyclohexane-1,2-diol has been used by many authors as an effective replacement for 3,4-disubstituted *N*-acetylglucosamine in the synthesis of sialyl Lewis X mimics. For a review, see: E. E. Simanek, G. J. McGarvey, J. A. Jablonowski and C.-H. Wong, *Chem. Rev.*, 1998, **98**, 833–862.
- 5 A. Bernardi, A. Checchia, P. Brocca, S. Sonnino and F. Zuccotto, *J. Am. Chem. Soc.*, 1999, **121**, 2032–2036.
- 6 C.-L. Schengrund and N. J. Ringle, *J. Biol. Chem.*, 1989, **264**, 13233–13237 and references therein.
- 7 (a) E. A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial and W. G. J. Hol, *Protein Sci.*, 1994, **3**, 166–175; (b) E. A. Merritt, P. Kuhn, S. Sarfaty, J. L. Erbe, R. K. Holmes and W. G. J. Hol, *J. Mol. Biol.*, 1998, **282**, 1043–1059.
- 8 A. Bernardi, G. Boschini, A. Checchia, M. Lattanzio, L. Manzoni, D. Potenza and C. Scolastico, *Eur. J. Org. Chem.*, 1999, **6**, 1311–1317.
- 9 A. Bernardi, L. Carrettoni, A. Grosso Ciponte, D. Monti and S. Sonnino, *Bioorg. Med. Chem. Lett.*, 2000, **10**, 2197–2200.
- 10 A. Bernardi, L. Raimondi and F. Zuccotto, *J. Med. Chem.*, 1997, **40**, 1855–1862.
- 11 A. Bernardi, M. Galgano, L. Belvisi and G. Colombo, *J. Comput.-Aided Mol. Des.*, 2001, **15**, 117–128.
- 12 P. Brocca, P. Berthault and S. Sonnino, *Biophys. J.*, 1998, **74**, 309–318.
- 13 A. Bernardi, D. Potenza, A. M. Capelli, A. García-Herrero, F. J. Cañada and J. Jiménez-Barbero, *Chem.-Eur. J.*, 2002, **8**, 4597–4612.
- 14 P. Brocca, A. Bernardi, L. Raimondi and S. Sonnino, *Glycoconjugate J.*, 2000, **17**, 283–299.
- 15 A. A. Bothner-By and R. Gassend, *Ann. N. Y. Acad. Sci.*, 1973, **222**, 668–676.
- 16 P. L. Jackson, H. N. Moseley and N. R. Krishna, *J. Magn. Reson., Ser. B*, 1995, **107**, 289–292; V. L. Bevilacqua, D. S. Thomson and J. H. Prestegard, *Biochemistry*, 1990, **29**, 5529–5537; V. L. Bevilacqua, Y. Kim and J. H. Prestegard, *Biochemistry*, 1992, **31**, 9339–9349.
- 17 M. Vogtherr and T. Peters, *J. Am. Chem. Soc.*, 2000, **122**, 6093–6099.
- 18 M. Mayer and B. Meyer, *J. Med. Chem.*, 2000, **43**, 2093–2099.
- 19 A. Schoen and E. Freire, *Biochemistry*, 1989, **28**, 5019–5024.
- 20 An accurate NMR study of these complexes is in progress, and will be reported in due course.
- 21 J. A. Mertz, J. A. McCann and W. D. Picking, *Biochem. Biophys. Res. Commun.*, 1996, **226**, 140–144.
- 22 M. Mammen, S.-K. Chio and G. M. Whitesides, *Angew. Chem., Int. Ed.*, 1998, **37**, 2755–2794.
- 23 S. Cattaldo Tesi di Laurea, Università di Milano a.a. 2000–2001.
- 24 E. A. Merritt, T. K. Sixma, K. H. Kalk, B. A. M. van Zanten and W. G. J. Hol, *Mol. Microbiol.*, 1994, **13**, 745–753.
- 25 A. T. Amn, S. Fraser, E. A. Merritt, C. Rodighiero, M. Kenny, M. Ahn, W. G. J. Hol, N. A. Williams, W. I. Lencer and T. R. Hirst, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 8536–8541.
- 26 W. E. Minke, C. Roach, W. G. J. Hol and C. L. M. J. Verlinde, *Biochemistry*, 1999, **38**, 5684–5692.
- 27 A. Pathiaseril and R. J. Woods, *J. Am. Chem. Soc.*, 2000, **122**, 331–338.
- 28 D. Arosio, unpublished.
- 29 Generous gift of Mariagrazia Pizza at Chiron Vaccines, Siena.
- 30 J. Dabrowski, T. Kozár, H. Grosskurth and N. E. Nifant'ev, *J. Am. Chem. Soc.*, 1995, **117**, 5534–5539.
- 31 D. Acquotti, L. Poppe, J. Dabrowski, C.-W. V.d. Lieth, S. Sonnino and G. Tettamanti, *J. Am. Chem. Soc.*, 1990, **112**, 7772–7778.

For more information see the following pages



Intramolecular Carbohydrate–Aromatic Interactions and Intermolecular van der Waals Interactions Enhance the Molecular Recognition Ability of GM1 Glycomimetics for Cholera Toxin

Anna Bernardi,^{*,[a]} Daniela Arosio,^[a] Donatella Potenza,^[a]
Inmaculada Sánchez-Medina,^[a] Silvia Mari,^[b] F. Javier Cañada,^[b] and
Jesús Jiménez-Barbero^{*,[b]}

Dedicated to the memory of Dr. Juan Carlos del Amo, postdoctoral fellow, killed in Madrid, in the terrorist attack of March 11th

Abstract: The design and synthesis of two GM1 glycomimetics, **6** and **7**, and analysis of their conformation in the free state and when complexed to cholera toxin is described. These compounds, which include an (*R*)-cyclohexyllactic acid and an (*R*)-phenyllactic acid fragment, respectively, display significant affinity for cholera toxin. A detailed NMR spectroscopy study of the toxin/glycomimetic complexes, assisted

by molecular modeling techniques, has allowed their interactions with the toxin to be explained at the atomic level. It is shown that intramolecular van der Waals and CH– π carbohy-

drate–aromatic interactions define the conformational properties of **7**, which adopts a three-dimensional structure significantly preorganized for proper interaction with the toxin. The exploitation of this kind of sugar–aromatic interaction, which is very well described in the context of carbohydrate/protein complexes, may open new avenues for the rational design of sugar mimics.

Keywords: carbohydrate–aromatic interactions • carbohydrate–protein interactions • cholera toxin • glycomimetics • oligosaccharides

Introduction

Many biologically significant processes are known to be controlled in their early stages by the interaction of proteins with oligosaccharides, often in the form of glyconjugates. Thus, oligosaccharide mimics that can antagonize oligosaccharides at the protein receptor level are receiving much attention, both as tools to modulate or alter signal transmission and to be developed into drugs.^[1] Our groups have been exploring this field by using the GM1 ganglioside **1a** as a model system (Scheme 1). GM1 is a membrane glycolipid which functions as the cellular receptor of two related

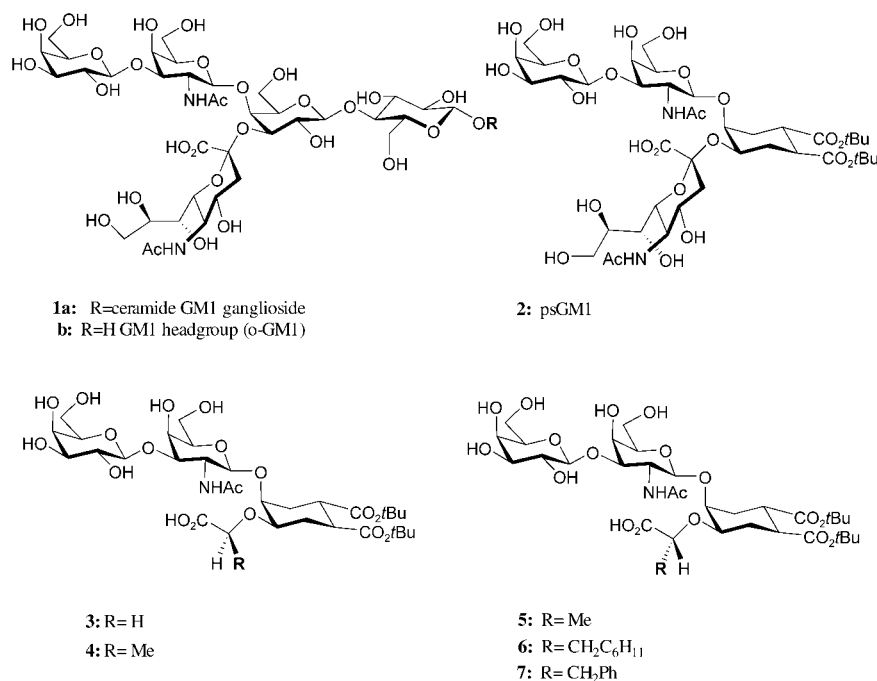
bacterial enterotoxins, the cholera toxin (CT) and the heat-labile toxin of *Escherichia coli* (LT). Both toxins are hexameric AB₅ proteins and use the GM1 headgroup pentasaccharide (o-GM1, **1b**, Scheme 1) as their molecular target to attack and penetrate the host cells. The recognition pair composed of GM1 and these two bacterial enterotoxins has been particularly well studied, both from a biochemical and a structural point of view.^[2,3] This information served as the basis for the rational design of the pseudooligosaccharide **2** (Scheme 1),^[4,5] which was found to be as active as GM1 in binding to CT.^[4] More recently, we have reported on a group of second-generation mimics, **3–7** (Scheme 1),^[6,7] obtained by replacing the sialic acid (NeuAc) moiety of **2** with simple α -hydroxy acids. All of these compounds show moderate to good affinity for CT (Table 1); this affinity depends critically on the configuration of the hydroxy acid stereocenter and on the nature of the substituent R (Table 1).

The design process of this series of mimics was supported by extensive NMR spectroscopy studies.^[8] After the first group of three ligands (**3–5**, Scheme 1) was synthesized and tested, their conformation was investigated first in solution and then upon binding to the cholera toxin by using transferred nuclear Overhauser effect (TR-NOE)^[9,10] measurements. It was found that CT selects a conformation similar to the global minimum of the free pseudosaccharides from

[a] Prof. Dr. A. Bernardi, Dr. D. Arosio, Dr. D. Potenza, I. Sánchez-Medina
Università di Milano—Dipartimento di Chimica Organica e Industriale e Centro di Eccellenza CISI
via Venezian 21, 20133 Milano (Italy)
Fax: (+39)02-50314072
E-mail: anna.bernardi@unimi.it

[b] S. Mari, Dr. F. J. Cañada, Prof. Dr. J. Jiménez-Barbero
Dept. de Estructura y Función de Proteínas
Centro de Investigaciones Biológicas, C.S.I.C.
Ramiro de Maeztu 9, 28040 Madrid (Spain)
Fax: (+34)91-5360432
E-mail: jjbarbero@cib.csic.es

Supporting information for this article is available on the WWW under <http://www.chemeurj.org/> or from the author.



Scheme 1. Ganglioside GM1 and its mimics.

Table 1. CT affinity constants of the second-generation GM1 mimics 3–7 as determined by fluorescence titration experiments.^[a]

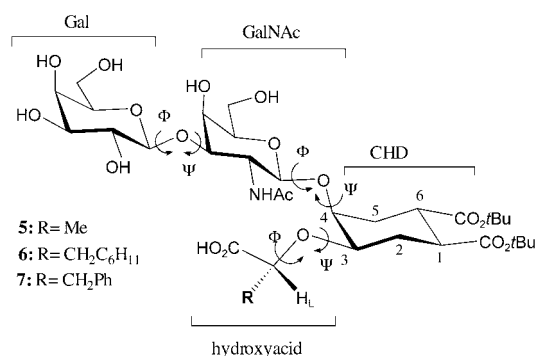
| Compound | R (configuration) | K_d [μM] | Ref. |
|----------|---|-------------------------|------|
| 3 | H | 750 | [6] |
| 4 | Me (<i>S</i>) | 1000 | [6] |
| 5 | Me (<i>R</i>) | 190 | [6] |
| 6 | CH ₂ C ₆ H ₁₁ (<i>R</i>) | 45 | [7] |
| 7 | CH ₂ Ph (<i>R</i>) | 10 | [7] |

[a] K_d values were obtained from fluorescence intensity titrations of 0.5 μM CT solutions (in tris(hydroxymethyl)aminomethane (Tris) buffer, pH 7, room temperature) and nonlinear fitting (SigmaPlot) of data obtained in duplicate runs.

the ensemble of the presented conformations, and no evidence of major conformational distortions was obtained. In the free state the three molecules were shown to be rather flexible in the hydroxy acid region. In the bound state^[8] the protein appeared to select for binding one or two side-chain conformations that could reproduce the orientation of the NeuAc carboxy group in GM1. The NMR data were interpreted with the aid of molecular-modeling techniques, which allowed workable models for the ligand:toxin complexes to be derived. These models suggested that the higher affinity of the (*R*)-lactic acid derivative **5** relative to **3** and **4** (Table 1) could result from van der Waals interactions established between its side-chain methyl group and a hydrophobic area in the toxin binding site near the sialic acid side-chain binding region of the CT/GM1 complex. This information in turn suggested that the affinity of the pseudo-GM1 binders could be improved by adding appropriate hydrophobic fragments to the framework of the (*R*)-lactic acid GM1 mimic **5**.^[8] We now report on the pseudo-GM1 ligands **6** and **7**, which include a cyclohexyl group and a phenyl group, respectively. Their synthesis is described (see the Supporting Information) and detailed NMR spectroscopy and computa-

tional studies in the free state and in the complex with CT are discussed. A preliminary communication has been reported.^[7]

Definitions and abbreviations: Residues of the pseudosugars **5–7** are defined as indicated in Scheme 2. The CHD residue is numbered as depicted in Scheme 2, to help comparison with the branching galactose unit of GM1. Glycosidic angles are defined as follows: Gal β (1 \rightarrow 3)GalNAc: Φ = GalH1-GalC1-O1-GalNAcC3, Ψ = GalC1-O1-GalNAcC3-GalNAcH3; Gal β (1 \rightarrow 4)CHD: Φ = GalH1-GalC1-O1-CHDC4, Ψ = GalC1-O1-CHDC4-CHDH4; hydroxy acid-CHD: Φ = C(O)-C $_{\alpha}$ -O-CHDC3, Ψ = C $_{\alpha}$ -O-CHDC3-CHDH3. The improper-

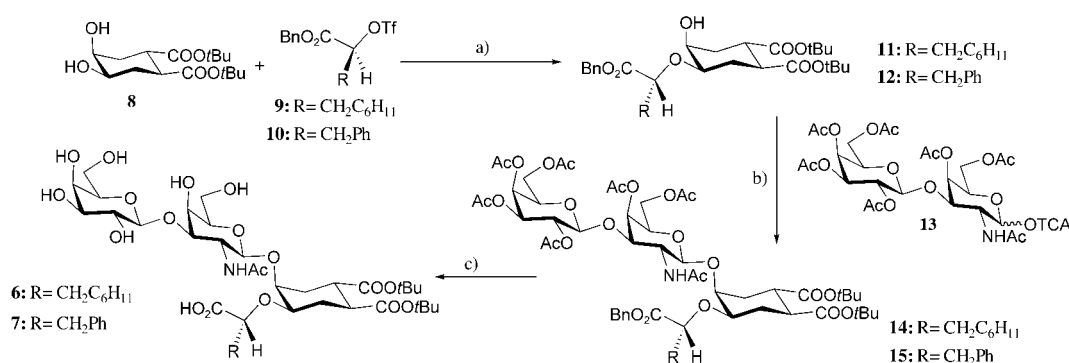
Scheme 2. GM1 mimics nomenclature, abbreviations, and definitions. Gal = galactose, GalNAc = *N*-acetyl galactosamine, CHD = cyclohexanediol, H_L = H-L in the text and other figures.

er dihedral angle χ (see Figure 3) describes the relative orientation of the carboxy group and the CHD ring and is defined as follows: χ = C(O)-C $_{\alpha}$ -CHDC3-CHDH3. The dihedral angle θ (see Figure 4) describes the orientation of the cyclohexyl group in the hydroxy acid side chain and is defined as θ = C(O)-C $_{\alpha}$ -CH₂-CC _{γ} .

Results

Synthesis of the ligands: The synthetic pathway followed for the preparation of ligands **6** and **7** is shown in Scheme 3. The full synthetic sequence and product characterizations are reported in the Supporting Information for this paper.

Binding-affinity determination: Dissociation constants (K_d) were obtained from fluorescence intensity titrations of 0.5 μM CT solutions (in Tris buffer, pH 7, room temperature)



Scheme 3. Synthesis of **6** and **7**. a) Bu_3SnO , benzene, reflux; then CsF , DME, and **9** or **10**; b) **13** (0.5 equiv) and TfOH (0.05 equiv) in CH_2Cl_2 , RT \rightarrow reflux; c) $\text{H}_2/\text{Pd-C}$, MeOH; then cat. MeONa in MeOH. Bn = benzyl, DME = 1,2-dimethoxyethane, TCA = trichloroacetic acid, Tf = triflate = trifluoromethanesulfonyl.

as previously described.^[7] In particular, the K_d values shown in Table 1 were obtained by nonlinear fitting (SigmaPlot) of data obtained in duplicate runs. Remarkably, the simple (*R*)-2-hydroxy-3-phenyl propionic acid (phenyllactic acid) derivative **7** showed a dissociation constant of $10\ \mu\text{M}$, which is only one order of magnitude less potent than the monovalent association of the natural ligand o-GM1 against the cholera toxin.^[11] The cyclohexyl derivative **6** showed a somewhat lower affinity, with $K_d = 45\ \mu\text{M}$; this is still better than the value obtained for the simple (*R*)-lactic derivative **5** ($K_d = 190\ \mu\text{M}$). This experimental fact enforces the hypothesis drawn from the previous modeling studies^[8] and emphasizes the importance of setting a nonpolar group bulkier than a methyl group at the proper position. The fine details of the differences between **6** and **7** will be given below.

NMR studies of the free ligands 6 and 7: NMR spectroscopy experiments were carried out at 400 and 500 MHz at temperatures of 293–300 K. A complete assignment of the ^1H and ^{13}C NMR signals of **6** and **7** was achieved on the basis of COSY, TOCSY, HSQC, and NOESY experiments. Chemical shifts and coupling constants are reported in the Supporting Information. No changes in the NOESY spectra were noticed upon addition of calcium ions (up to 5 mM). It has been reported^[12] that GM1 and other gangliosides might form complexes with calcium ions. Probably, the lack of a lipid chain, which might also affect local concentrations, may somehow influence this type of interaction in this case. For both compounds, the analysis of the vicinal proton–proton coupling constants for the six-membered rings indicates that the Gal and GalNAc chairs are in the usual $^4\text{C}_1$ conformation. The diol moiety also adopts a chair conformation with the ester groups in the equatorial orientations.

The intraresidue NOE cross-peaks also support the theory that all the six-membered rings of the molecules are in chair conformations (see below and Tables 2 and 3).

Table 2. Principal NOE contacts for compound **6** in the free state and bound to cholera toxin. The distances r ($\pm 10\%$) are estimated according to a full-matrix relaxation approach.^[32] The intraresidue Gal and GalNAc H-1/H-3 and H-1/H-5 contacts were taken as internal references.^[a]

| Proton pair | | Observed intensity free state | Deduced r [Å] | Observed intensity bound state | Deduced r [Å] |
|-------------|------------------------------|-------------------------------|-----------------|--------------------------------|-----------------|
| H-1 GalNAc | H-4 CHD | strong | 2.4 | strong | 2.4 |
| | H-L | n.o. ^[a] | > 3.5 | n.o. ^[a] | > 3.5 |
| H-1 Gal | H-3 GalNAc | strong | 2.4 | strong | 2.4 |
| | H-3 CHD | medium | 2.7 | strong | 2.5 |
| H-L | CH_2 ^[b] | strong | | strong | |
| | H-2ax CHD | n.o. ^[a] | | very weak | |
| | H-2eq CHD | weak | 2.9 | medium-strong | 2.7 |
| | H-3 CHD | medium | 2.6 | strong | 2.5 |
| | H-4 CHD | very weak | 3.1 | very weak | 3.2 |
| H-4 CHD | H-2ax CHD | n.o. ^[a] | > 3.5 | very weak | 3.3 |
| | H-1 GalNAc | strong | 2.5 | strong | 2.5 |
| | H-L | n.o. | > 3.5 | weak | 3.2 |

[a] ax = axial, eq = equatorial, n.o. = no observable NOE contact. [b] CH_2 and H-5ax CHD are isochronous.

Vicinal $J_{5,6}$ couplings for the hydroxymethyl groups could be measured for **6** and **7**. The coupling values were between 7.5–9 Hz for both the Gal and GalNAc residues, in agreement with a *gt:tg* equilibrium of the ω torsion angle, as is usually the case for these Gal-type sugars.^[13]

The conformation of oligosaccharides is defined by the Φ and Ψ torsion angles around the glycosidic linkages (Scheme 2). For **6** and **7**, the rotation around the ether linkage in the hydroxy acid side chain must also be described. Φ, Ψ definitions can be extrapolated from the glycosidic bond convention and defined as described in Scheme 2. Experimental information on these geometrical parameters can be gathered by using NOE measurements.^[14] Specifically, selective 1D NOESY and 2D T-ROESY^[14] experiments were carried out that provided experimental interproton distances.

These, in turn, allowed the conformation equilibria present in solution to be profiled and average molecular conformations to be derived (Tables 2 and 3).

The relationship between NOE signals and proton–proton distances is well established^[15,16] and can be worked out at

Table 3. Principal NOE contacts for compound **7** in the free state and when bound to cholera toxin. The distances r ($\pm 15\%$) are estimated according to a full-matrix relaxation approach.^[132] The intrasidue Gal and GalNAc H-1/H-3 and H-1/H-5 contacts were taken as internal references.

| Proton pair | | Observed intensity free state | Deduced r [Å] | Observed intensity bound state | Deduced r [Å] |
|-------------|---------------------------|-------------------------------|-----------------|--------------------------------|-----------------|
| H-1 GalNAc | H-4 CHD | strong | 2.4 | medium-strong | 2.5 |
| | H-L | n.o. | >3.5 | n.o. | >4.0 |
| H-1 Gal | H-3 GalNAc | strong | 2.4 | strong | 2.4 |
| | H-1 GalNAc | medium | 2.8 | medium | 2.8 |
| Ph | H-L | medium | 2.7 | medium | 2.7 |
| | H-4 CHD | medium | 2.8 | medium | 2.8 |
| | H-6 GalNAc | weak | 3.1 | medium | 2.8 |
| | H-3 GalNAc ^[a] | medium | 2.8 | medium | 2.8 |
| | H-5 GalNAc | medium | 2.8 | medium | 2.8 |
| | H-3 CHD ^[a] | strong | 2.3 | strong | 2.3 |
| | H-2eq CHD | strong | 2.4 | strong | 2.4 |
| H-L | H-1 CHD | weak | 3.0 | weak | 3.2 |
| | H-1 GalNAc | strong | 2.5 | strong | 2.5 |
| | H-L | strong | >3.5 | weak | 3.2 |
| H-4 CHD | H-1 GalNAc | strong | 2.5 | strong | 2.5 |
| | H-L | n.o. | >3.5 | weak | 3.2 |
| | H-L | n.o. | >3.5 | weak | 3.2 |

[a] H-3 GalNAc and H-3 CHD are isochronous.

least semiquantitatively by using a relaxation matrix.^[16] The NOE intensities reflect the conformer populations, and therefore information on the population distributions in free solution can be obtained by focusing on the key, mutually exclusive NOE interactions that characterize the different possible conformations.^[17]

At 400 MHz all the cross-peaks observed in the NOE spectra of **6** and **7** were very weak. The $\omega\tau_c$ value is close to 1.1, which provides an almost zero longitudinal NOE contact (τ_c is the overall correlation time). Although some initial information could be derived from T-ROESY spectra, many crucial proton signals (for example, H-L and H-2 GalNAc) were basically isochronous; this prompted us to a 500 MHz analysis. At 293 K, the corresponding NOESY spectra (Figures 1 and 2) were good enough to extract the observed cross-peaks, which are reported in Tables 2 and 3.

The orientation around the Gal β (1 \rightarrow 3)GalNAc linkage can be defined^[8,18] by the NOE contact observed between the anomeric proton of galactose, H-1 Gal, and the protons on the GalNAc moiety, especially H-3, H-4, and H-2. The H-1 Gal/H-3 GalNAc cross-peak is very strong for both **6** and **7** (see Tables 2 and 3), with an interaction similar to or even stronger than the corresponding intrasidue H-1 Gal/H-3 Gal and H-1 Gal/H-5 Gal cross-peaks. As we have already described for **3–5**,^[8] the H-1 Gal/H-3 GalNAc distance obtained from the NMR spectroscopy data (approximately 2.4 Å) can

be correlated with one major orientation around the Gal β (1 \rightarrow 3)GalNAc Φ/Ψ torsion angles. This region corresponds to the global minimum *syn* conformation, characterized by Φ/Ψ values of around 60°/–20°.

A similar situation occurs for the GalNAc β (1 \rightarrow 4)CHD moiety, with a similar NOE pattern.^[8,18] In fact, only the very strong interresidue H-1 GalNAc/H-4 CHD cross-peak is observed; this cross-peak corresponds to a H-1 GalNAc/H-4 CHD distance of 2.5 ± 0.2 Å. The average conformation in solution for this linkage corresponds to the global minimum *syn* conformer, characterized by

Φ/Ψ values of around 50°/20°. Conformers in the *anti-Φ* or *anti-Ψ* regions of both glycosidic linkages with populations above 5% would give rise to exclusive H-1 Gal/H-2, H-4 GalNAc and/or H-1 GalNAc/H-5ax, H-5ax CHD NOE contacts that are not observed in these cases.

Regarding the orientation of the acetamide moiety^[18] of both **6** and **7**, the cross-peak of the methyl group with the corresponding H-2 GalNAc proton is a very weak, and no cross-peaks to either H-1 or H-3 GalNAc are observed. Therefore, these observations are in agreement with a major orientation in solution with the methyl group pointing out of the pyranoid ring with an *anti-like* relationship with the C-2 atom.

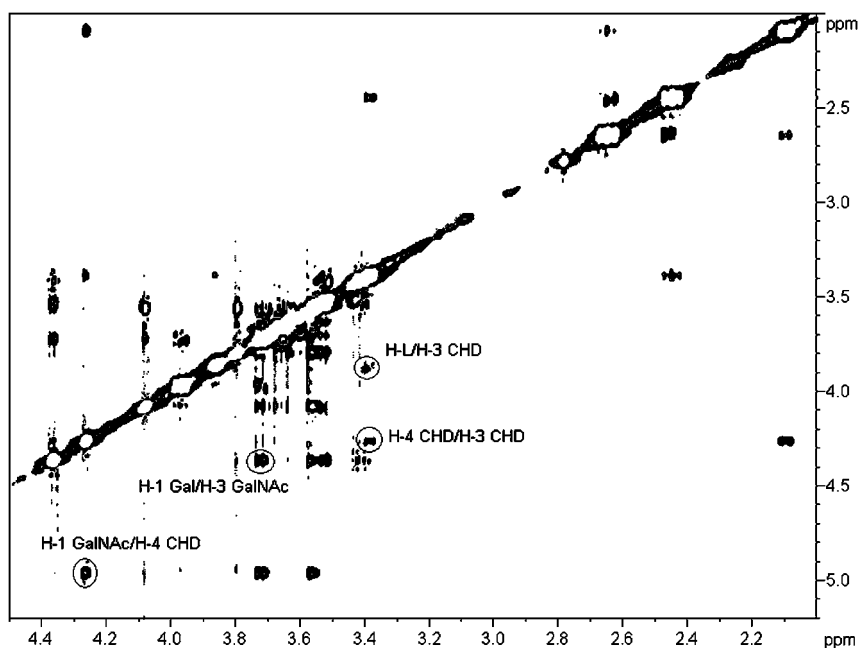


Figure 1. 500 MHz NOESY spectrum of **6** in D₂O at 293 K.

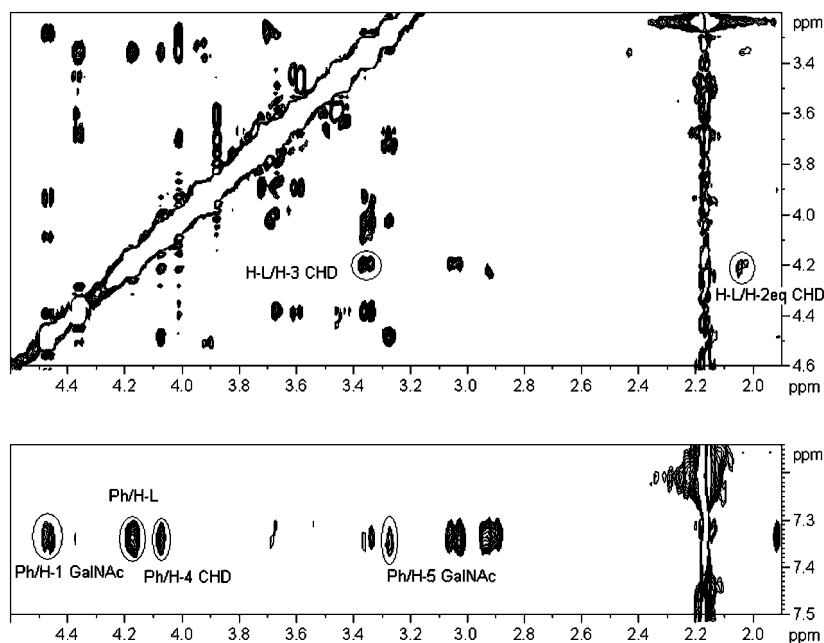


Figure 2. Fragments of the 500 MHz NOESY spectra of **7** in D₂O at 300 K, from which the orientation of the side chain is deduced. The NOE contacts for H-L (top) and for the aromatic protons (bottom) are indicated.

Thus, all data so far concur to describe the “upper” portion of both **6** and **7** as a pseudotrisaccharide with relatively little flexibility, which can be described by oscillations around one single major conformation. This is the same situation observed for all the GM1 mimics synthesized so far,^[8] and for the Gal β (1 \rightarrow 3)GalNAc β (1 \rightarrow 4)Gal fragment of o-GM1 itself.^[19]

The additional torsional degrees of freedom available for **6** and **7**, correspond to the hydroxy acid side-chain bonds. Due to the absence of anomeric effects, the ether linkage that connects the CHD ring to the hydroxy acid moiety is more flexible than the interglycosidic linkages described so far. The NOE contacts observed in this region can arise from multiple combinations of Φ, Ψ values that happen to

determine similar through-space proton–proton interactions for the reporter H-L. Thus, the experimental inter-proton distances generated by NMR spectroscopy analysis can be analyzed more conveniently by using the improper dihedral angle descriptor χ (C(O)-C $_{\alpha}$ -CHDC3-CHDH3), which is defined in Figure 3, and univocally describes the orientation of H-L (and of the hydroxy acid carboxy group) relative to the CHD ring. In principle, three idealized staggered orientations can be drawn across the ether connector (Figure 3); these can be identified as *anti*- χ , (*-g*)- χ , and (*+g*)- χ and they give rise to specific interactions between H-L and the protons on the C-3, C-2 and C-4 carbon atoms of the CHD ring. The major orientation(s) of the hydroxy acid relative to the CHD ring can

thus be defined by focusing on the NOE cross-peaks for the H-L proton to H-3, H-4, H-2eq, and H-2ax of CHD (Tables 2 and 3). Further information on the mobility of the side chain and on its conformation(s) can be gathered from the NOE contacts observed for the R side-chain substituent.

For the cyclohexyl derivative **6**, medium and weak NOE contacts are observed between H-L and the key CHD protons (see Table 2). Nevertheless, the NOE intensity of the H-L/H-3 CHD cross-peak is higher than that of the H-L/H-2eq CHD connectivity and also stronger than the H-L/H-4 CHD one. The H-L/H-1 GalNAc and H-L/H-2ax NOE cross-peaks are also negligible. In turn, the cyclohexyl moiety of the side chain only displays weak cross-peaks to its vicinal H-3 CHD proton. No cross-peaks are observed

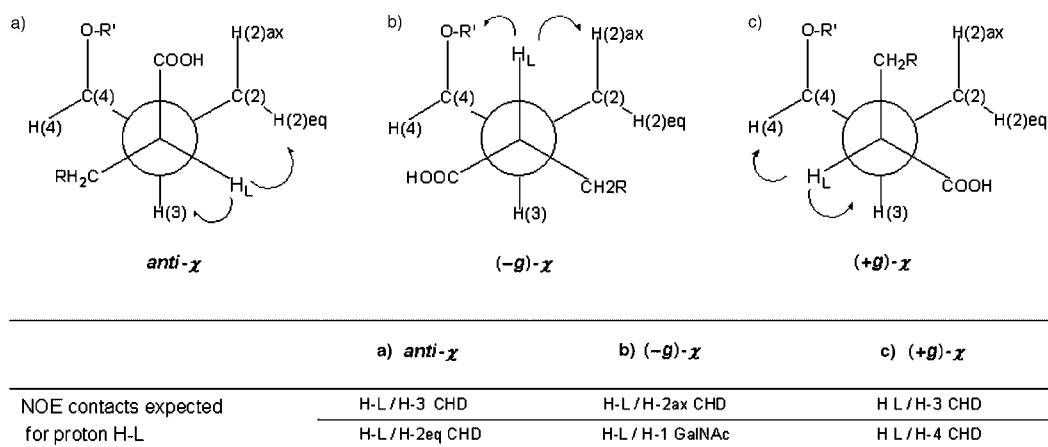


Figure 3. Newman projections along the ether linkage between the CHD ring and the hydroxy acid moiety. The exclusive NOE contacts expected for H-L in each rotamer are shown. Each of these projections may correspond to multiple Φ, Ψ combinations. χ = the improper dihedral angle. See text for further details.

between the cyclohexyl side chain and H-2eq or H-4 CHD. The absence of cross-peaks between H-L and H-1 GalNAc and the very weak intensity of those between H-L and H-4 CHD and between H-L and H-2 CHD are consistent with a very flexible side chain and indicate that probably all three staggered rotamers depicted in Figure 3 are contributing to the equilibrium.

For the conformation analysis of **7**, the key NOE contacts that allow the conformation equilibrium present in solution to be described were obtained from NOESY and T-ROESY experiments carried out at 500 MHz and 300 K. Several interresidual cross-peaks were observed. As described above for **6**, the major arrangement of the hydroxy acid side chain can be defined by focusing on the NOE contacts between H-L and the CHD protons. In this case, important information is also obtained from the through-space interactions between the aromatic protons and the CHD and GalNAc residues (Figure 2, Table 3). The H-L proton shows NOE cross-peaks of strong intensity with H-2eq and H-3 CHD (with corresponding distances of 2.4 and 2.3 ± 0.1 Å, respectively), while no cross-peaks are observed between H-L and H-4 CHD (Figure 2a, Table 3). This pattern, which is markedly different from what was measured for **6**, is in agreement with one major *anti*- χ orientation (Figure 3) between the carboxyl group and the H-3 CHD proton. In addition, the aromatic protons also show clear cross-peaks to a variety of protons (Figure 2b, Table 3) located not only on the CHD moiety (H-4), but also on the α face of the GalNAc residue (H-1, H-3, and H-5, Figure 2b). Thus, it appears that in the major solution conformation of **7** the phenyl moiety stacks below the GalNAc residue.

Collectively, these experimental data allow us to conclude that the phenyllactic derivative **7** is significantly less flexible in the hydroxy acid fragment than the cyclohexyllactic analogue **6** and suggest that a single conformation of the hydroxy acid chain can describe the conformational properties

of **7** in aqueous solution. Importantly, the origin of this conformation lock appears to be intramolecular van der Waals and CH- π interactions between the aromatic ring and the GalNAc residue; this may be described as hydrophobic packing. These interactions are not so favorable in **6** where the cyclohexyl ring obviously differs from the phenyl moiety in its electronic and geometrical features. The importance of carbohydrate–aromatic interactions for the molecular recognition of oligosaccharides by the binding sites of proteins has been well documented.^[20–24] Here, we present clear evidence of the importance of this type of interaction for favoring a given type of three-dimensional structure.

Computational models of **6** and **7** were generated by using previously established protocols (MC/EM conformation searches and MC/SD molecular-dynamics simulations, AMBER* force field augmented by the Kolb parameters^[25] for hydroxy acids, GB/SA water solvation) and compared to the experimental results. The calculations showed that the Gal–GalNAc–CHD fragment of both molecules is populating the *syn* conformation for both glycosidic linkages. Minima were located at φ/ψ values of $50^\circ/0^\circ$ for the Gal–GalNAc linkage and $25^\circ/30^\circ$ for the GalNAc–CHD linkage. By contrast, and in agreement with the experimental data discussed above, the computational description of the hydroxy acid side chain differs significantly for the two molecules. For **6**, MC/EM calculations located three conformers within 1 kcal mol^{-1} of the global minimum, at $\chi = 150, 65,$ and 39° (Figure 4a and Table 4). MC/SD dynamics showed that the molecule is highly flexible along the variable χ angle and continuously populates an ample region with χ varying from 40 – 160° plus a secondary low-energy region at $\chi = -60^\circ$ (Figure 4b). The presence of this extensive conformation equilibrium agrees with the experimental observation.

MC/EM calculations also showed that the side chain of **7** is significantly less flexible than that of **6**, although computa-

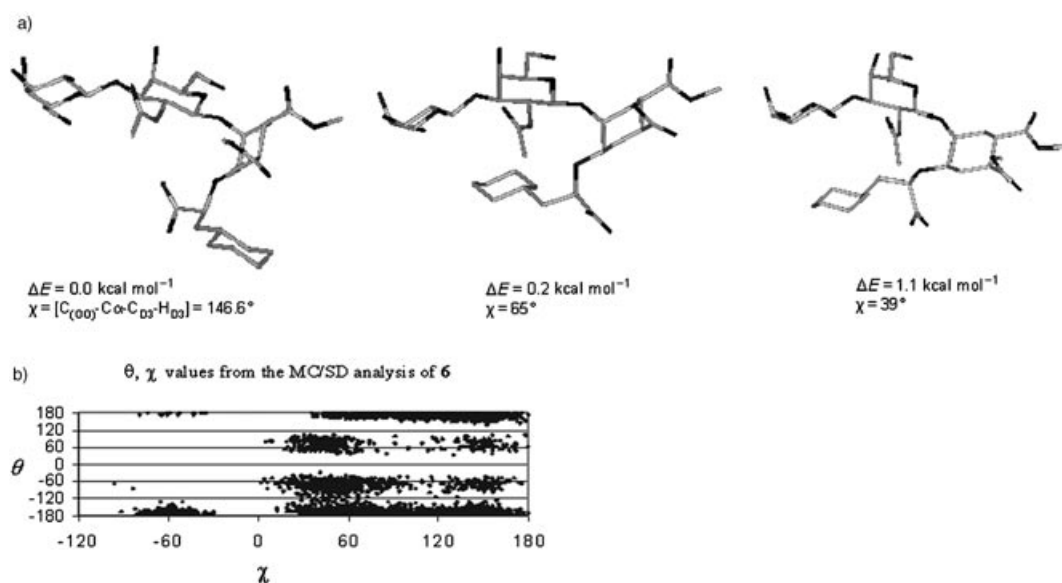


Figure 4. Conformation analysis of **6**. a) Low-energy conformations within $1.2 \text{ kcal mol}^{-1}$ of the global minimum. b) MC/SD dynamics simulations (5 ns). The improper dihedral angle, χ , is plotted against the dihedral angle, $\theta\text{C}(\text{O})\text{-C}_\alpha\text{-CH}_2\text{-CC}_\gamma$.

Table 4. Three different conformation (conf.) families were found by an MC/EM^[a] conformation search, each with a different value of the improper angle, χ ($\Phi = C_{(OO)}-C_{\alpha}-O-C_{D3}$; $\chi = C_{(OO)}-C_{\alpha}-C_{D3}-H_{D3}$). The key interatomic proton–proton distances are outlined.

| | Distances [Å] | | |
|--------------------|---|--|--|
| | conf. 1 ($\Delta E = 0.0 \text{ kcal mol}^{-1}$, $\chi = 146.6^\circ$) | conf. 2 ($\Delta E = 0.2 \text{ kcal mol}^{-1}$, $\chi = 65^\circ$) | conf. 3 ($\Delta E = 1.1 \text{ kcal mol}^{-1}$, $\chi = 39^\circ$) |
| H-L/H-3 CHD | 2.51 | 2.81 | 2.42 |
| H-L/H-2eq CHD | 2.25 | 3.92 | 4.24 |
| H-L/H2ax CHD | 3.39 | 4.49 | 4.47 |
| H-L/H-4 CHD | 4.39 | 3.22 | 2.47 |
| H-L/H-1 GalNAc | 4.36 | 3.71 | 2.81 |
| H-1 GalNAc/H-4 CHD | 2.69 | 2.24 | 2.31 |

[a] 10000 steps of MC/EM with the H₂O GB/SA solvent model and the Amber* force field were employed.

tionally the restriction of mobility is mostly confined to the orientation of the phenyl ring, which is always calculated to stack either with the Gal or the GalNAc ring (Figure 5). Two energetically equivalent orientations of the carboxyl group were calculated, with $\chi \approx 150^\circ$ (*anti*- χ) and $\chi \approx 60^\circ$ ((+)-*g*- χ ; Figure 5 a and b). However, in this case, the NMR experimental data (see above) allows us to conclude that only one major orientation of the side chain, corresponding to the *anti*- χ conformation, is in fact present. The best fit with the NOE data is represented by a conformation found near the global minimum, as shown in Figure 5 c. Based on these data, ligand **7** appears to be preorganized for toxin binding, in contrast with the conformational behavior of its aliphatic analogue **6**, which is fairly flexible.

It is interesting to note that the stacking effect between the aromatic ring and the sugars is somewhat “understood” by the molecular mechanics computations, which most likely “see” it as a solvation effect (the hydrophobicity of the ring) combined with a positive van der Waals interactions between the phenyl group and the α face of the sugars. This latter interaction appears to be precluded to the cyclohexyl substituent for steric reasons.

NMR studies of the CT complexes: NMR experiments, including TR-NOESY^[9,10] and saturation transfer difference (STD),^[26–28] were performed to deduce the bound conforma-

tions of **6** and **7** to the CT B pentamer. As previously shown, for ligands which are not bound tightly and exchange between the free and bound states at a reasonably fast rate, the TR-NOESY experiment provides an adequate means for determining the conformation of the bound ligand.^[9,10] As mentioned above (Table 1), the present systems bind in the micromolar range to CT and thus can properly be studied by this technique. The addition of cholera toxin to a D₂O solution of **6** and **7** induced broadening of

the resonance signals in the ¹H NMR spectrum, a result indicating that binding occurs (Figures S1 and S2 in the Supporting Information). TR-NOESY experiments were performed on the ligand/CT samples at different mixing times and with different ligand to toxin molar ratios (25:1 to 50:1). Negative cross-peaks were clearly observed at 300 K, as expected for ligand binding.

For the phenyllactic acid based ligand **7**, the cross-peak pattern is similar to the one described above for free **7** in aqueous solution. Indeed, the same set of interresidue cross-peaks with basically identical intensities (relative to the intraresidual contacts) were observed (Table 3). The orientation around the Gal β (1 \rightarrow 3)GalNAc and GalNAc β (1 \rightarrow 4)CHD fragments is unchanged upon binding, as deduced by the corresponding H-1 Gal/H-3 GalNAc and H-1 GalNAc/H-4 CHD cross-peaks. The orientation of the phenyllactic acid moiety relative to the GalNAc and CHD residues is also almost identical to that observed in the free state, as shown by the H-L/H-2eq CHD, H-L/H3 CHD, and aromatic/H-1, H-3, H-5 GalNAc short contacts (Figure 6, Table 3). TR-ROESY experiments allowed spin-diffusion effects to be excluded for these key cross-peaks.^[29]

For **6**, TR-NOESY experiments were also performed at a 32:1 ligand/receptor molar ratio. Negative cross-peaks were clearly observed at 300 K, both at 400 and 500 MHz, a result indicating ligand binding (Table 2). In this case, the analysis

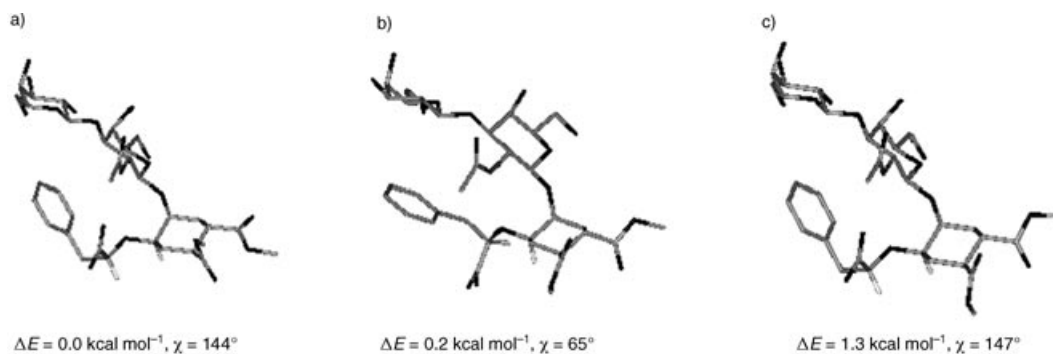


Figure 5. Two major conformation families, a and b, were found for the side chain of **7** by the MC/EM conformation search. They are characterized by a different value of the improper torsional angle, χ . According to the experimental results, only the *anti*- χ conformation ($\chi \approx 150^\circ$) is found in solution. The best fit with the NOE data is obtained with conformer c, which belongs to the a family.

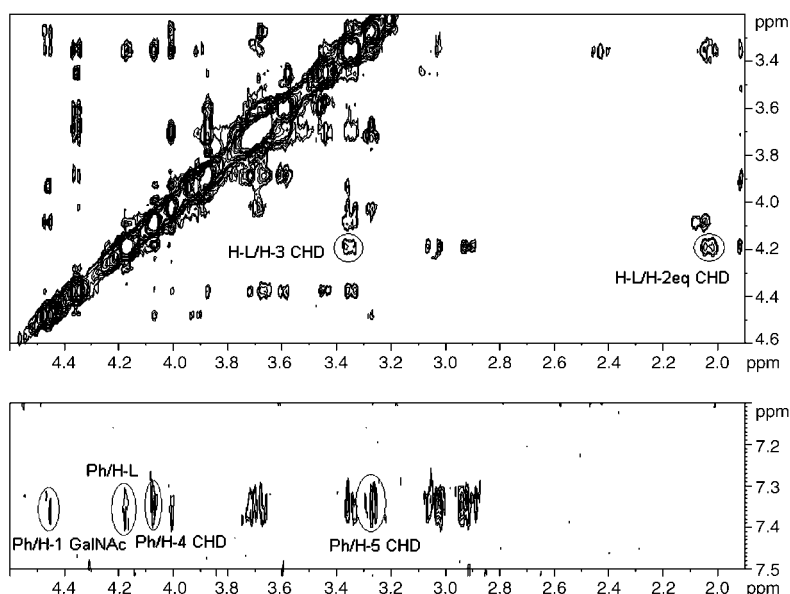


Figure 6. TR-NOESY spectrum of **7** bound to CT (solvent D₂O, 293 K).

of the cross-peaks shows that, out of the conformation equilibrium observed in the free state, one main conformation, which features an *anti*-type relationship of the carboxyl group to the H-3 CHD proton (*anti*- χ in Figure 3), is selected for binding. In fact, cross-peaks between H-L/H-2eq CHD and H-L/H-3 CHD were observed (Figure 7) that agreed with the conformation selection of the *anti*- χ rotamer ($\chi \approx 150^\circ$). A minor proportion of a *gauche*- χ rotamer cannot be discarded, as indicated by the presence of a very weak H-L/H-4 CHD TR-NOESY cross-peak.

Finally, in order to map the binding epitope of the glycomimetics bound to the cholera toxin, 1D STD NMR experiments were also performed.^[26–28] For **7**, the difference spec-

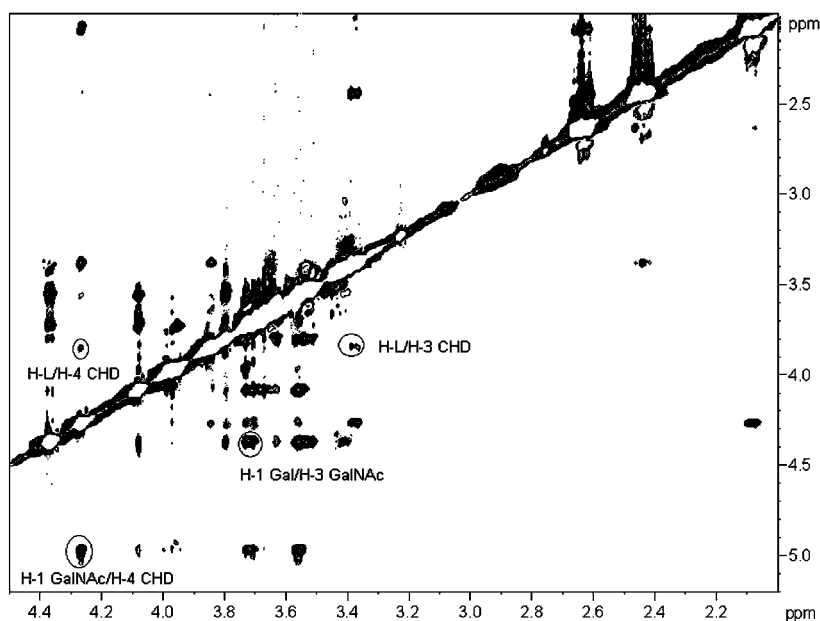


Figure 7. TR-NOESY spectrum of **6** bound to CT (solvent D₂O, 300 K).

trum showed that the aromatic protons, together with H-4 Gal and H-6 Gal, give rise to the most prominent STD signals (Figure 8). Thus, it can be inferred that this region (aromatic ring and Gal moiety) of the glycomimetic is in more intimate contact with the toxin binding site. Additionally, the resonance signals of H-3 and H-1 Gal, H-1 GalNAc, and H-4 CHD also appeared in the difference spectrum but with smaller intensities. The weakest STD signals were observed for the CHD moiety protons, except for H-4, a result indicating that this part of the molecule is not involved in the molecular recognition process (Figure 8) to a significant extent.

The STD spectrum of the CT/**6** mixture (Figure 9) also allowed the binding epitope to be deduced. Resonance signals belonging to the Gal moiety (H-1, H-2, H-4, H-5, H-6) were evident in the difference spectrum. Additional peaks for H-1, H-4, and H-5 GalNAc were also clearly observed. However, those belonging to the CHD ring and to the pendant cyclohexyl moiety were rather weak, or basically nonexistent, including the H-3 CHD and H-4 CHD protons (see Figure 9). This indicates that these moieties establish only marginal interactions with the protein, in contrast to the aromatic ring of the phenyllactic derivative **7**.

As a final step, a three-dimensional model of the complex structures was obtained by performing MC/EM calculations within the binding site of the LT toxin, according to the procedure already described for the lactic acid analogues.^[8] The corresponding views are depicted in Figure 10. It clearly appears that the phenyl ring in the LT/**7** complex (Figure 10a) is calculated to be in close proximity to the protein binding site. On the contrary, and in accordance with experimental observations, the cyclohexyl ring in LT:**6** projects away from the protein surface (Figure 10b).

Dynamics simulations (MC/SD) also showed that the phenyl ring of **7** is locked between the protein and the GalNAc ring (see Figure S3 (animated gif file) in the Supporting Information). A clear intramolecular carbohydrate–ar-

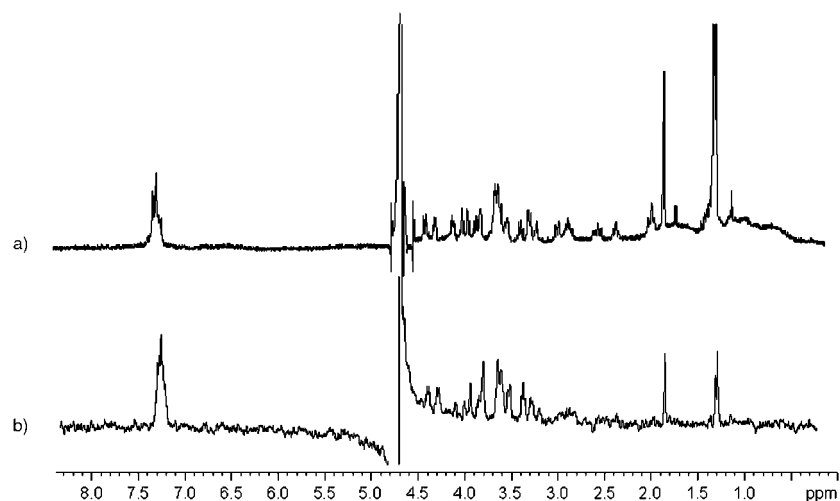


Figure 8. a) Reference 1D ^1H NMR spectrum of **7** with CTB5 (ratio 25:1) in D_2O at 300 K. b) STD spectrum with 1 s presaturation of the protein envelope protons. Ligand proton signals are evident in the aromatic region. Protons are those belonging to the Gal moiety (H-4 and H-6). H-1 GalNAc, H-1 Gal, H-4 CHD, and H-3 GalNAc are also clearly visible, while the other protons of the CHD unit do not appear in the STD spectrum.

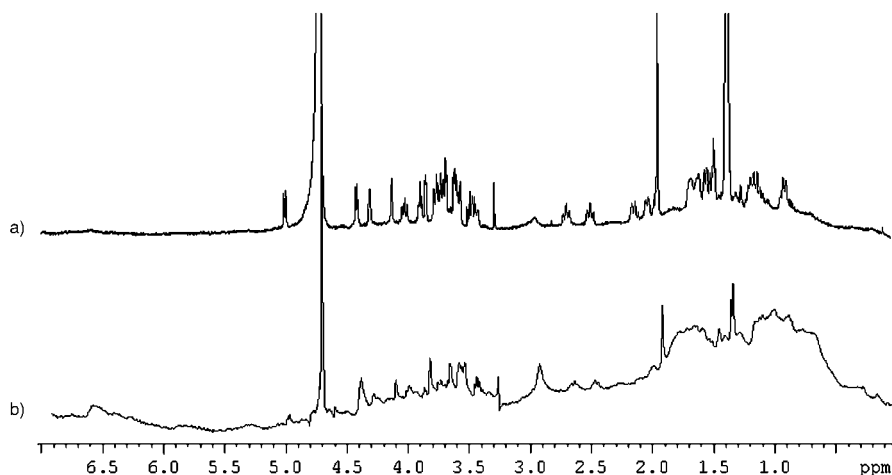


Figure 9. a) Reference 1D ^1H NMR spectrum of **6** with CTB5 (ratio 32:1) in D_2O at 298 K. b) STD spectrum with 2 s presaturation of the protein envelope protons. Protons belonging to the GalNAc and Gal moiety (H-1 GalNAc, H-1 Gal, H-4 GalNAc, H-2 GalNAc, H-4 Gal, H-6 GalNAc, H-6 Gal, H-5 GalNAc, H-5 Gal, and H-2 Gal) are clearly visible.

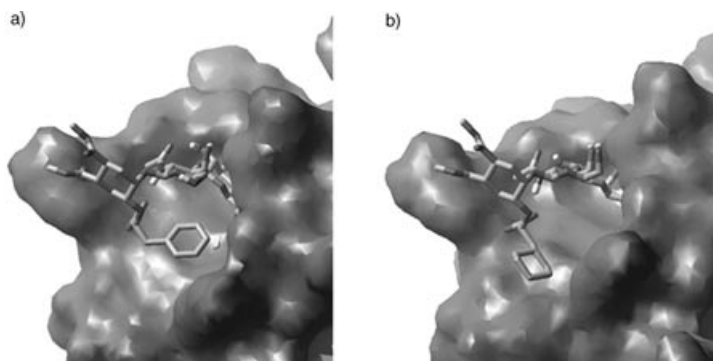


Figure 10. a) Molecular model (MC/EM, Amber*, H_2O , GB/SA, best fit with experimental data for **7**). b) Molecular model (MC/EM, Amber*, H_2O , GB/SA, lowest energy conformer for **6**).

omatic stacking interaction takes place and locates the phenyl ring in the proper position to establish additional contacts with the hydrophobic patch in the protein. The typical Gal–toxin interaction, which defines the structure of cholera toxin complexes, is also evident.

In contrast, in the LT/6 complex the cyclohexyl ring appears to display more flexibility even in the bound state, with the Gal moiety establishing the major contacts with the toxin. No stacking of the GalNAc moiety with the aliphatic cyclohexyl side chain takes place. The cyclohexyl ring appears to move in and out of the toxin binding side and to sample at least two different orientations relative to the protein cavity (see Figure S4 (animated gif file) in the Supporting Information).

Discussion

All the known structural data on the complexes formed between CT and GM1 or the pseudo-GM1 mimics **2–5** show that the principal interactions between the protein and the (pseudo)carbohydrate ligands are established through the nonreducing end galactose unit and the carboxy group of the NeuAc residue (in **1** and **2**) or of the surrogate hydroxy acid (in **3–5**). NMR and computational studies performed on the complexation of **3–5** by CT

have suggested that additional van der Waals interactions can be gained by lipophilic substituents on the (*R*)-hydroxy acid side chain,^[8] which, in the bound conformation of the ligand, can extend towards a hydrophobic area of the toxin binding site close to the region that accommodates the sialic acid side chain in the CT/GM1 complex.^[3] The present work shows that the pseudo-GM1 ligands **6** and **7**, which include an (*R*)-cyclohexyllactic acid fragment and an (*R*)-phenyllactic acid fragment, respectively, do indeed display stronger affinity for CT than **5**.

Although the two ligands described in this paper are similar in nature and activity, a detailed analysis of their behavior in solution and in the binding site of the cholera toxin reveals striking differences, both in terms of conformational flexibility and of binding mode. Intra- and intermolecular in-

teractions among the different residues strongly modulate the conformational features of these molecules in solution and when bound to the toxin. The major differences between **6** and **7** may be ascribed to the presence of an intramolecular aromatic–carbohydrate interaction in the phenyl-lactic acid derivative **7** that strongly biases its conformational behavior by severely restricting its conformational freedom. This conformational constraint is lacking in the cyclohexyllactic acid derivative **6**, which behaves similarly to the lactic acid derivative **5** in terms of flexibility and preferred orientations of the hydroxy acid side chain. As a result of the conformation lock, the side chain of **7** is preorganized in an *anti- χ* conformation that allows optimal interaction of the carboxy group in the carboxylate binding region of CT. The same conformation appears to be attained by **6** in its bound state, but it has to be selected from a pool of different rotamers that are simultaneously present in solution. NMR analysis of the bound state of the ligands also reveals that the phenyl ring of **7** is in close contact with the protein, as revealed by the intense signal of the aromatic protons in the STD spectrum of the CT:**7** complex. In contrast, the cyclohexyl group of **6** appears to make a much looser contact with the toxin. Thus, the preorganization effect and a more efficient van der Waals interaction between the side-chain substituent and the protein appear to concur in determining the high affinity of **7** for CT.

The phenyl ring–GalNAc stacking interaction that we have observed in **7** has been described here for the first time as an element of a conformation lock in the structure of a sugar mimic. This kind of sugar–aromatic interaction, which is very well described in the context of carbohydrate/protein complexes^[20–24] may have a broader application, which we intend to explore, as an element in conformation-based design of sugar mimics. Additionally, it has been reported that the presence of the hydrophobic ceramide moiety may influence the head-group presentation of the glycan^[30] and may even influence the binding to cholera toxin.^[31] Thus, although this report has focused on the importance of the oligosaccharidic part, further studies in our groups are paying attention to the importance of the lipid chain for the interaction under different experimental conditions.

Experimental Section

NMR spectroscopy experiments: NMR spectra were recorded at 25–30 °C in D₂O on Varian Unity 500 MHz and Bruker Avance 400 and 500 MHz spectrometers. For the experiments with the free ligand, the compound was dissolved in D₂O and the solution was degassed with a stream of argon. COSY, TOCSY, and HSQC experiments were performed by using the standard sequences. 2D T-ROESY experiments were performed with mixing times of 300 and 500 ms. The strength of the 180 pulses during the spin-lock period was attenuated 4 times with respect to that of the 90 hard pulses (between 7.2 and 7.5 μ s). In order to deduce the interproton distances, relaxation matrix calculations were performed by using software written inhouse which is available from the authors upon request.^[32] For the bound ligands, STD and TR-NOE experiments were performed as previously described.^[8] First, the B chain of cholera toxin was subjected to 2 cycles of freeze–drying with D₂O to remove traces of H₂O. It was then transferred in solution to the NMR

spectroscopy tube to give a final concentration of approximately 0.1–0.2 mM. TR-NOESY experiments were performed with mixing times of 100, 200, and 300 ms, for molar ratios of ligand/protein between 15:1 and 50:1. No purging spin-lock period to remove the protein background signals was employed. First, in all cases, line broadening of the sugar protons was monitored after the addition of the ligand. STD experiments were carried out by using the method proposed by Meyer, Peters, and their respective co-workers.^[26–28] No saturation of the residual HDO signal was employed and, again, no spin-lock pulse was employed to remove the protein background signals. In our hands, the use of a spin-lock period induced artifacts in the difference spectrum. The theoretical analysis of the TR-NOE contacts of the sugar protons was performed according to the protocol employed by London, with a relaxation matrix with exchange as described.^[8] Different exchange-rate constants, k , defined as $\text{pf} \cdot k = K_{-1}$ (where pf is the fraction of the free ligand), and leakage relaxation times were employed to obtain the optimal match between the experimental and theoretical results for the intraresidue H-1/H-3 and H-1/H-5 cross-peaks of the Gal and GalNAc moieties for the given protein/ligand ratio. Normalized intensity values were used since they allow correcting for spin-relaxation effects. The overall correlation time, τ_c , for the free state was always set to 0.15 ns and the τ_c for the bound state was estimated as 15 ns according to the molecular weight of the toxin ($\tau_c = 10\text{--}12 \times M_w$). To fit the experimental TR-NOE intensities, exchange-rate constants, k , between 100–1000 s⁻¹ and external relaxation times, ρ^* , for the bound state of 0.5, 1, and 2 s were tested. Optimal agreement was achieved when $k = 150 \text{ s}^{-1}$ and $\rho^* = 1 \text{ s}$.

TR-ROESY experiments were also carried out to exclude spin-diffusion effects. A continuous-wave spin-lock pulse was used during the 250 ms mixing time. Key NOE contacts were shown to be direct cross-peaks, since they showed different signs to diagonal peaks. In some cases, they allowed the detection of intra-Gal and intra-GalNAc H-1/H-4 and H-1/H-6 cross-peaks that were due to spin diffusion.

Computational methods

Conformation search and dynamics of isolated **6 and **7**:** The calculations were performed by using the MacroModel/Batchmin^[33] package (version 7.0) and the AMBER* force field. Kolb's parameters were used for the hydroxy acid moiety.^[25] Bulk water solvation was simulated by using MacroModel's generalized Born GB/SA continuum solvent model.^[34] The conformation searches were carried out by using 20,000 steps of the usage-directed MC/EM procedure according to previously established protocols.^[35,19] Extended nonbonded cut-off distances (a van der Waals cut-off of 8.0 Å and an electrostatic cut-off of 20.0 Å) were used.

For the MC/SD^[36] dynamic simulations, van der Waals and electrostatic cut-offs of 25 Å, together with a hydrogen-bond cut-off of 15 Å, were used. The dynamic simulations were run with the AMBER* all-atom force field. Charges were taken from the force field (all-atom option). The same degrees of freedom of the MC/EM searches were used in the MC/SD runs. All simulations were performed at 300 K, with a dynamic time step of 1 fs and a frictional coefficient of 0.1 ps⁻¹. Typically, 2 runs of 5 ns each were performed, with 2 conformations of the substrates as the starting point; these conformations differed at the hydroxy acid linkage and were selected from the MC/EM outputs. The Monte Carlo acceptance ratio was about 4%, and each accepted MC step was followed by an SD step. Structures were sampled every 1 ps and saved for later evaluation. Convergence was checked by monitoring both energetic and geometrical parameters.

Conformation search (MC/EM) of the LT complexes: The conformation searches were carried out by using the usage-directed MC/EM procedure, with slight variations of the protocol used in the study of the LT/psGM1 complex.^[4] In brief, the starting structure for LT/**7** was obtained by superimposing the conformation of **7** that exhibits the best fit with the NOE data (conformer *c* in Figure 5) on the psGM1 ligand in our model of the LT:psGM1 complex by using the galactose coordinates. This “docking” step was followed by substructure energy minimization. The starting structure of the LT/**6** complex was obtained from the LT/**7** complex by graphically converting the phenyl ring into a cyclohexyl group.

The explicit torsional variables were the same as those used for the free-state calculations, that is, the Gal–GalNAc and Gal–CHD anomeric linkages, the C-5–C-6 bonds of Gal and GalNAc, and the five bonds of the hydroxy acid moiety. Furthermore, the ligand was allowed to rotate (max

180°) and translate (max. 1 Å) within the binding site (MOLS command of Batchmin). 10000 MC/EM steps were performed. Bulk water solvation was simulated by using the GB/SA model. Five crystallographic water molecules were retained, as previously described.^[4,35] All calculations were carried out on a B2 (B+B(+1)) dimer. Only the ligand and a shell of residues surrounding the binding site of LT were subjected to energy minimization. All the residues within 6 Å of the sugars were completely included in the shell. An all-atom treatment was used for the ligand and for the aromatic residues of the protein. The rest of the toxin was treated with a united-atom model. The ligand and all of the binding site polar hydroxy and amino hydrogen atoms were unconstrained during energy minimization. All other atoms that belonged to the substructure being minimized were constrained to their crystallographic coordinates by parabolic restraining potentials that increased with the distance from the sugar substrate. The following force constants were used: 100 kJ Å⁻² for atoms within 0–3 Å of any atom of the ligand; 200 kJ Å⁻² for atoms within 3–4 Å; 400 kJ Å⁻² for atoms within 4–5 Å. The periphery of the restrained structure was checked with the EdgeD command of MacroModel, and isolated atoms were included to avoid incomplete functional groups. All other atoms were ignored.

MC/SD dynamics of the LT:ligand complexes: The simulations were carried out by using the same substructure and explicit MC variables described above, but with the lowest energy conformation from the MC/EM search as the starting point. Extended nonbonded cut-offs were employed (van der Waals and electrostatic cut-off of 25 Å, hydrogen-bond cut-off of 15 Å). The simulations were performed for 1 ns, at 300 K, with a dynamic time step of 1 fs and a frictional coefficient of 0.1 ps⁻¹. Structures were sampled every 2 ps and saved for later evaluation.

Acknowledgments

The project was supported by the Azioni Integrate program between Italy and Spain, and the European programs COST-D13 and "Improving Human Potential" under contract HPRNCT-2002-00173 (Glycidic Scaffolds Network). I.S.-M. acknowledges a Marie Curie Fellowship from the European Community program 'Improving Human Potential' under contract HPMT-CT-2001-00293. The Madrid group also acknowledges the Ministry of Science and Technology of Spain for funding (Grant BQU2003-03550-C01).

- [1] P. Sears, C. H. Wong, *Angew. Chem.* **1999**, *111*, 2446–2471; *Angew. Chem. Int. Ed.* **1999**, *38*, 2300–2324, and references therein.
- [2] C.-L. Schengrund, N. J. Ringler, *J. Biol. Chem.* **1989**, *264*, 13233–13237.
- [3] E. A. Merritt, S. Sarfaty, F. van den Akker, C. L'Hoir, J. A. Martial, W. G. J. Hol, *Protein Sci.* **1994**, *3*, 166–175; E. A. Merritt, T. K. Sixma, K. H. Kalk, B. A. M. Van Zanten, W. G. J. Hol, *Mol. Microbiol.* **1994**, *13*, 745–753; E. A. Merritt, S. Sarfaty, M. G. Jobling, T. Chang, R. K. Holmes, W. G. J. Hol, *Protein Sci.* **1997**, *6*, 1516–1528.
- [4] A. Bernardi, A. Checchia, P. Brocca, S. Sonnino, F. Zuccotto, *J. Am. Chem. Soc.* **1999**, *121*, 2032–2036.
- [5] A. Bernardi, G. Boschin, A. Checchia, M. Lattanzio, L. Manzoni, D. Potenza, C. Scolastico, *Eur. J. Org. Chem.* **1999**, 1311–1317.
- [6] A. Bernardi, L. Carrettoni, A. Grosso Ciponte, D. Monti, S. Sonnino, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2197–2200.
- [7] D. Arosio, S. Baretta, S. Cattaldo, S. Potenza, A. Bernardi, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3831–3834.
- [8] A. Bernardi, D. Potenza, A. M. Capelli, A. García-Herrero, F. J. Cañada, J. Jiménez-Barbero, *Chem. Eur. J.* **2002**, *8*, 4597–4612.
- [9] A. A. Bothner-By, R. Gassend, *Ann. N.Y. Acad. Sci.* **1973**, *222*, 668–676.
- [10] P. L. Jackson, H. N. Moseley, N. R. Krishna, *J. Magn. Reson. Ser. B* **1995**, *107*, 289–292; V. L. Bevilacqua, D. S. Thomson, J. H. Prestegard, *Biochemistry* **1990**, *29*, 5529–5537; V. L. Bevilacqua, Y. Kim, J. H. Prestegard, *Biochemistry* **1992**, *31*, 9339–9349; H. Kogelberg, D. Solís, J. Jiménez-Barbero, *Curr. Opin. Struct. Biol.* **2003**, *13*, 646–653;
- [11] A. Schon, E. Freire, *Biochemistry* **1989**, *28*, 5019–5024.
- [12] M. B. Khalil, M. Kates, D. Carrier, *Biochemistry* **2000**, *39*, 2980–2988.
- [13] K. Bock, J. O. Dues, *J. Carbohydr. Chem.* **1994**, *13*, 513–543.
- [14] D. Neuhaus, M. P. Williamson, *The NOE Effect in Structural and Conformational Analysis*, VCH, New York, **1989**.
- [15] T. L. Hwang, A. J. Shaka, *J. Am. Chem. Soc.* **1992**, *114*, 3157–3158.
- [16] *NMR Spectroscopy of Glycoconjugates* (Eds.: J. Jiménez-Barbero, T. Peters), VCH, Weinheim, **2002**.
- [17] J. Dabrowski, T. Kozar, H. Grosskurth, N. E. Nifant'ev, *J. Am. Chem. Soc.* **1995**, *117*, 5534–5539.
- [18] A. Bernardi, D. Arosio, L. Manzoni, D. Monti, H. Posteri, D. Potenza, S. Mari, J. J. Barbero, *Org. Biomol. Chem.* **2003**, *1*, 785–792.
- [19] P. Brocca, A. Bernardi, L. Raimondi, S. Sonnino, *Glycoconjugate J.* **2000**, *17*, 283–299.
- [20] N. K. Vyas, *Curr. Opin. Struct. Biol.* **1991**, *1*, 732–740.
- [21] F. A. Quijcho, *Biochem. Soc. Trans.* **1993**, *21*, 442–448.
- [22] W. I. Weis, K. Drickamer, *Annu. Rev. Biochem.* **1996**, *65*, 441–473.
- [23] S. Elgavish, B. Shaanan, *J. Mol. Biol.* **1998**, *277*, 917–932.
- [24] J. Jimenez-Barbero, J. L. Asensio, F. J. Cañada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–555.
- [25] H. C. Kolb, B. Ernst, *Chem. Eur. J.* **1997**, *3*, 1571–1578.
- [26] M. Meyer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902–1906; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
- [27] J. Klein, R. Meinecke, M. Meyer, B. Meyer, *J. Am. Chem. Soc.* **1999**, *121*, 5336–5337.
- [28] M. Vogtherr, T. Peters, *J. Am. Chem. Soc.* **2000**, *122*, 6093–6099.
- [29] J. L. Asensio, F. J. Cañada, J. Jimenez-Barbero, *Eur. J. Biochem.* **1995**, *233*, 618–630.
- [30] D. Roy, C. Mukhopadhyay, *J. Biomol. Struct. Dyn.* **2002**, *19*, 1121–1132.
- [31] B. Lanne, B. Schierbeck, J. Angstrom, *J. Biochem.* **1999**, *126*, 226–234.
- [32] A. Poveda, J. L. Asensio, M. Martín-Pastor, J. Jimenez-Barbero, *J. Biomol. NMR* **1997**, *10*, 29–43.
- [33] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
- [34] W. C. Still, A. Tempzyk, R. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- [35] A. Bernardi, L. Raimondi, F. Zuccotto, *J. Med. Chem.* **1997**, *40*, 1855–1862.
- [36] F. Guarnieri, W. C. Still, *J. Comput. Chem.* **1994**, *15*, 1302–1310.

Received: January 26, 2004

Revised: May 3, 2004

Published online: July 27, 2004

Synthesis and conformational analysis of galactose-derived bicyclic scaffolds.

Submitted for publication

Silvia Mari,^c Ingrid Velter,^a Francesco Nicotra,^{a*} Barbara Laferla,^a Anna Bernardi,^{b*} Gilles Marcou,^b Ilaria Motto,^b Francisco J. Cañada,^c Jesus Jiménez-Barbero^{c*}

- a- Università degli Studi di Milano, Milan
- b- Università Bicocca, Milan
- c- Centro de Investigaciones Biológicas, CSIC, Madrid

Introduction

Carbohydrates can be regarded as ideal scaffolds for library generation:ⁱ they are mainly polyhydroxylated (but they also bear amino and carboxy functionalities) and each group has a well defined orientation, enforced by the conformational constraints of the cyclic structure. Natural carbohydrates, however, are non-original scaffolds, and their chemical and metabolic stability is not very high. The chemical modification of carbohydrates can afford original, stable, and possibly more rigid compounds. Such structures can be decorated by appending diverse pharmacophores to their functional groups. For practical purposes it is important that the new scaffolds be readily accessible on a large scale at a limited cost. Information about their structure and dynamics in water solution is also very valuable for structure-based (drug)design projects and for the rational design of individual ligands and libraries thereof. In this context, where computational tools are required, their ability to render the three-dimensional structure of the scaffolds should be tested against the available experimental data for validation, before embarking in subsequent molecular design projects.

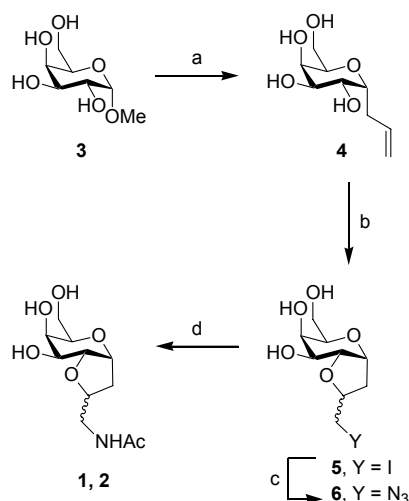
In the course of a collaboration within an European Research Network (<http://www.altaweb.it/glycidic/index.html>) we have developed a short synthesis of the functionally rich bicyclic structures **1** and **2** (Chart 1). These molecules are easily obtained from galactose in a few steps, without the need of protecting group chemistry (see below) and can be variously functionalized at the appended side-chain. Both characteristics make them particularly attractive for further (combinatorial) development. To fully exploit their potential, a thorough conformational study by NMR spectroscopy supported by molecular modeling has been performed and has allowed to establish their 3D structure. The results are reported in this paper.

Results

Synthesis

The preparation of the bicyclic acetamides **1** and **2** started from the commercially available methyl α -D-galactopyranoside **3** (Scheme 1), according to a procedure described by Gray and coworkers.ⁱⁱ Silylation of **3** using bis(trimethylsilyl)trifluoroacetamide (BTSFA) was followed by in situ

treatment with allyltrimethylsilane and catalytic trimethylsilyl trifluoromethane,ⁱⁱⁱ and afforded the α -C-allyl galactoside **4** as a single diastereomer. The stereochemistry of compound **4** was assigned by evaluation of the coupling constant J of the H(1) and H(2) hydrogen atoms in its acetylated derivative, $J_{H1-H2} = 4.8$ Hz, revealing thus their *cis* relationship.



Scheme 1. a) i. BTSFA, CH₃CN, reflux, 1 h; ii. allyltrimethyl silane, TMSOTf (0.5 eq), 0 °C - rt, 18 h, 85%; b) I₂, NaHCO₃, DMF, rt, 12 h, 84%; c) *n*Bu₄NN₃, DMF, rt, 14 h, 75%; d) H₂, Pd(OH)₂/C, Ac₂O, MeOH, rt, 24 h 70%.

Iodocyclization of compound **4** provided a 2.6:1 inseparable mixture of diastereomeric iodides **5**,^{iv} which was then submitted to azide displacement to afford a the bicyclic azides **6**. Hydrogenation of this mixture, over 20% Pd(OH)₂/C, in methanol and in the presence of acetic anhydride, at atmospheric pressure, afforded a mixture of bicyclic acetamides **1** and **2**, which were separated by HPLC, using a Waters high performance carbohydrate 60Å column. The synthesis of compounds **1** and **2** was thus performed without the use of protection/deprotection chemistry, in four steps and 37% overall yield.

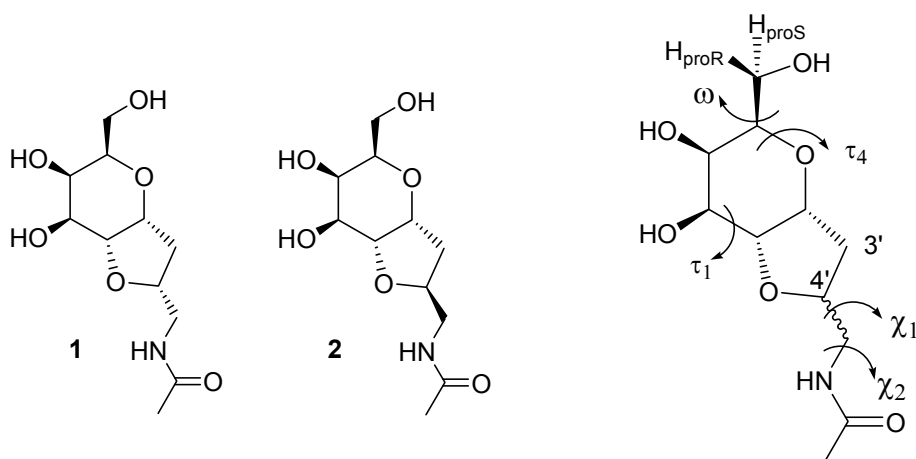


Chart 1. Structure of compounds **1** and **2** and definitions of dihedral angles ω , τ_4 , τ_1 , χ_1 and χ_2 as OC₅C₆O₆, C₁OC₅C₄, C₁C₂C₃C₄, O⁺C₄⁻C₅⁻N, C₄⁻C₅⁻NC(O) respectively.

Conformational studies: NMR data

In order to define their three-dimensional structure and dynamics **1** and **2** were subjected to NMR studies in D₂O solution at 400 MHz and 500 MHz. Both coupling constant analysis (Table 1) and NOE data (Table 2) suggest that the bicyclic cores of both compounds explore multiple conformations. In particular, the data obtained for the 6-membered ring protons are clearly not consistent with the single ⁴C₁ chair that is usually adopted by *galacto*-pyranoses: the observed J_{2,3} values are too small, and the J_{4,5} values too large (Table 1). Consistently, the H3-H5 NOE crosspeaks are weaker than expected for a typical ⁴C₁ pyranose ring, and the H1-H6*proR* proton pair gives rise to medium intensity NOE contacts which are not compatible with a ⁴C₁ ring conformation (Table 2). Hence, the 1,2-*cis*-fused furane ring appears to distort the six-membered ring of both **1** and **2** out of the usual *galacto*-pyranose conformation by a measurable extent.

The orientation of the hydroxyl methyl group tethered to the pyrane ring was also investigated: as expected for a galactose residue the measured coupling constants were consistent with a preferred *gt* rotamer ($\omega = +60^\circ$).^v

Table 1. Experimental ³J_{HH} values for compounds **1** and **2**.

| Protons pairs | Exp. ³ J _{H-H} (Hz) for 1 (R) | Exp. ³ J _{H-H} (Hz) for 2 (S) |
|---------------------|---|---|
| H1-H2 | 6.6 | 5.9 |
| H2-H3 | 6.4 | 6.7 |
| H3-H4 | 2.8 | 3.0 |
| H4-H5 | 2.8 | 3.2 |
| H1-H3' <i>proS</i> | 6.6 | 6.9; 5.3 ^(c) |
| H1-H3' <i>proR</i> | 6.6 | |
| H4'-H3' <i>proS</i> | 6.8; 7.8 ^(c) | 7,4; ? ^(c) |
| H4'-H3' <i>proR</i> | | |

c) The scalar coupling between these proton pairs could not be assigned: the corresponding values were not used in the following fitting procedure.

Table 2. Experimentally observed Noe contacts

| Compound | Proton pair | Experimental nOe (%) ^a |
|----------|---------------------|-----------------------------------|
| 1 | H1-H4' | 2.30 |
| | H3' <i>proS</i> -H3 | 3.70 |
| | H3' <i>proS</i> -H5 | 7.09 |
| | H3-H5 | 3.24 |
| | H1-H6 <i>proR</i> | 3.01 |

| | H4'-H2 | 2.72 |
|----------|------------|------|
| 2 | H1-H3'proR | 3.94 |
| | H1-H3'proS | 0.43 |
| | H1-H6proR | 2.89 |
| | H5-H3 | 3.43 |
| | H4'-H3 | 2.14 |
| | H6proS-H4 | 2.61 |
| | H3'proS-H3 | 1.93 |
| | H3'proS-H5 | 5.10 |
| | H3'proR-H2 | 1.28 |

a) % of nOe cross peak relative to its corresponding diagonal peak

Modeling results

In order to identify the other accessible conformations, a conformational analysis of **1** and **2** was performed by molecular mechanics, using three different force fields (AMBER*, MM3* and OPLSAA) and the implicit water salvation model GB/SA^{vi} of Macromodel.^{vii}

A conformational search was performed using the MC/EM protocol,^{viii} and the resulting conformations within 3 kcal/mol from the global minimum were clustered for each molecule based on the values of two dihedral angles of the ring τ_1 and τ_4 (C1-C2-C3-C4 and C1-O-C5-C4, respectively (See Chart 1). Structures were classified as skew-boat (S), chair (⁴C₁), or inverted chair (¹C₄) conformations of the pyrane moiety.^{ix} The bicycle conformations of the cluster leaders were found to be independent of the configuration of the stereocenter at C-4', and hence basically identical for both molecules. The results turned out to be highly force field dependent (for details, see the Supp. Info) and are illustrated in Figure 1, which shows the leading members of the clusters found by the three different force fields, their relative energy and their Boltzmann distribution at room temperature.

For each force field one different conformation should represent at least 92% of the global population, as based on the Boltzmann distribution at room temperature (Figure 1, last column). AMBER* favors a ⁴C₁ chair, MM3* the ¹S₃ skew-boat and OPLSAA a ¹C₄ chair. Each conformation of the pyrane ring is associated with multiple folds of the 5-membered ring and multiple conformations of the pending side chain, which are described in detail in the Supp.Info. section. Here we'll only note that for **1** the ¹S₃ cluster structures calculated by AMBER* and those calculated by MM3* are associated with different folds of the 5-membered ring. From Figure 1 it can also be noted that the *gt* orientation of the hydroxyl methyl group is correctly reproduced by AMBER* and OPLSAA, but conspicuously not by MM3* (see the cluster leaders in Figures 1).

None of the force field representations can reproduce the experimental observations. For instance, based on the population distributions of Figure 1 the diagnostic J_{4,5} coupling constants for **1** would be 0.9 Hz (AMBER*), 4.1 Hz (MM3*) and 6.0 Hz (OPLSAA), respectively for the three force fields, none of which agrees with the measured value of 2.8 Hz (see Table 1). Analogous considerations can be extended to **2**. Similarly, the NoE data estimated by a full matrix relaxation approach^x using the

Noeprom software^{xi} from the calculated structures and relative energies did not reproduce the experimentally observed cross peaks intensities. The use of dynamics simulations did not improve the fitting, suggesting that the problem is not caused by improper simulation of the dynamic behavior of the bicycles, but rather by poor representation of the actual potential energy surface by all the force fields examined. We therefore decided to use the calculated geometries as templates and to extract their relative populations from the experimental data by a fitting procedure based on both the coupling constant values and the intensity of the NoE cross peaks.

Coupling Constant Fitting

As discussed above, depending on the force field (AMBER*, MM3, or OPLSAA) different conformations of the 6 and the 5-membered rings were predicted. Furthermore, the fold of the fused five-membered ring appears to depend on the conformation of the pyrane moiety. With AMBER*, MM3 and OPLSAA geometries at hand, averaged $^3J_{H-H}$ were calculated for each cluster of conformers by applying the empirical generalization of Karplus equation proposed by Haasnoot et al.^{xii} to the corresponding proton-proton torsion angles. The cluster-averaged J values were found to be somewhat force field dependent, particularly for the 5-membered ring protons, reflecting the different 5-membered ring folds located by each force field.

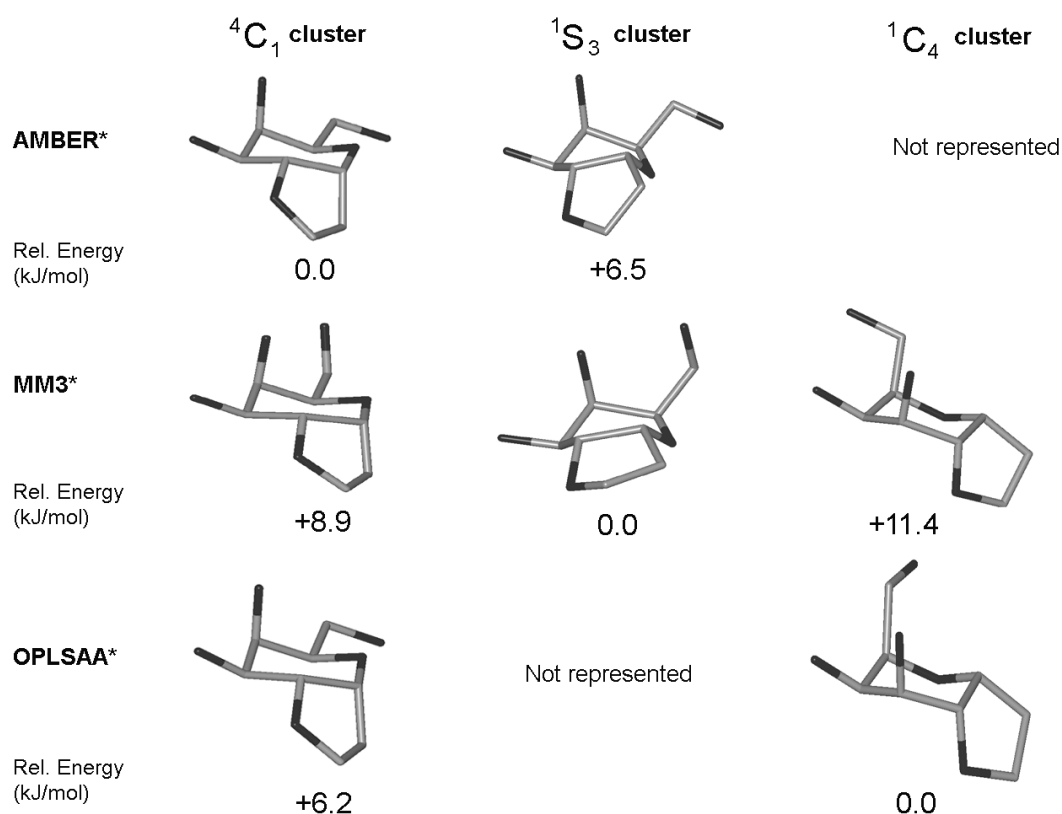


Figure 1. MC/EM search of **1** and **2**: clustering of the pyrane ring. Leading members of the clusters found by the three different force fields are shown, together with their relative energy (kJ/mol).

1: 4C_1 : 1S_3 97:3 (AMBER*), 4C_1 : 1S_3 : 1C_4 4:95:1 (MM3*) and 4C_1 : 1C_4 1:99 (OPLSAA)

2: 4C_1 : 1S_3 98:2 (AMBER*), 4C_1 : 1S_3 : 1C_4 3:92:5 (MM3*) and 4C_1 : 1C_4 1:99 (OPLSAA).

These values were used to obtain the relative population of each cluster by fitting to the experimental $^3J_{HH}$ values. This procedure allowed to account for the different 5-membered ring puckering represented in each cluster. We felt that this procedure was more robust than the alternative possibility of using only the cluster leader geometry as representative of each cluster in the fitting process. An initial value was assigned to the weight of each cluster and an estimation of the scalar couplings calculated; then, the deviation between the calculated and experimental scalar couplings was iteratively minimized (up to 10.000 steps), thus obtaining the best-fitting population weights. The relative populations of the clusters obtained by this fitting procedure (J-fitted populations) are reported in Table 3 (J-fitted ratio row) together with their associated scalar couplings constants values, that can be compared to the experimental values.

According to AMBER*, the experimental J values measured for **1** were best fitted by a 60:40 ratio of 4C_1 : 1S_3 conformations. The sum of all the deviations for the 8 measured J-couplings is 4.13 Hz. In contrast, fitting based on the structures calculated by MM3* led to a relative populations 48: 25: 27 for the 4C_1 : 1S_3 : 1C_4 conformers. In this case the deviations' sum over the eight coupling values is 3.46 Hz. Finally for the OPLSAA structures the best fit of the experimental values is obtained by 60:40 ratio of the 4C_1 and 1C_4 clusters. The sum of the deviations in this case is the highest value found and corresponds to 5.64 Hz.

Similar results were obtained for compound **2**. With Amber*, the fitted 4C_1 : 1S_3 ratio is 45:55 and the corresponding deviation is 4.1 Hz; the MM3* fitted ratio is 60:24:16 4C_1 : 1S_3 : 1C_4 with a deviation of 3.0 Hz while with OPLSAA the fitting gave 80:20 of 4C_1 : 1C_4 (deviation is 4.4 Hz).

Thus, the experimental results can be reproduced with different sets of conformations. In all cases the 4C_1 chair is found to be the major isomer and to represent 45% - 80% of the global conformer population. However, the J analysis alone does not allow to discriminate among force fields.

Table 3. Experimental and estimated $^3J_{HH}$ values for compounds **1** and **2**.

| Comp. | Protons pairs | Exp. $^3J_{H-H}$ (Hz) | Amber* $^3J_{H-H}$ (Hz) ^(a) | | MM3 $^3J_{H-H}$ (Hz) ^(a) | | | OPLSAA $^3J_{H-H}$ (Hz) ^(a) | | | OPLSAA fitting $^3J_{H-H}$ (Hz) ^(b) | | |
|-----------|---------------|-----------------------|--|---------|-------------------------------------|---------|---------|--|--|---------|--|---------|--|
| | | | 4C_1 | 1S_3 | 4C_1 | 1S_3 | 1C_4 | 4C_1 | E_3' | 1C_4 | 4C_1 | 1C_4 | |
| | | | 4C_1 | 1S_3 | 4C_1 : 1S_3 60:40 | 4C_1 | 1S_3 | 1C_4 | 4C_1 : 1S_3 : 1C_4 48:25:27 | 4C_1 | E_3' | 1C_4 | 4C_1 : E_3' : 1C_4 60:0:40 |
| 1R | H1-H2 | 6.6 | 6.4 | 4.3 | 5.4 | 6.9 | 5.8 | 2.1 | 5.3 | 7.1 | 5.1 | 2.1 | 5.1 |
| | H2-H3 | 6.4 | 8.3 | 5.9 | 7.2 | 7.8 | 7.3 | 2.7 | 6.3 | 7.0 | 0.7 | 2.4 | 5.2 |
| | H3-H4 | 2.8 | 2.9 | 2.2 | 2.6 | 2.7 | 1.8 | 3.4 | 2.7 | 3.3 | 6.7 | 4.2 | 3.7 |
| | H4-H5 | 2.8 | 0.8 | 3.7 | 2.1 | 0.7 | 4.3 | 5.2 | 2.9 | 0.7 | 0.2 | 6.0 | 2.8 |
| | H1-H3'proS | 6.6 | 6.5 | 6.0 | 6.3 | 6.3 | 9.4 | 5.1 | 6.8 | 6.4 | 9.7 | 6.3 | 6.4 |

| | | | | | | | | | | | | | |
|-----------------------------|-------------|--------------------|-----------------------------|-----------------------------|---|-----------------------------|-----------------------------|-----------------------------|--|-----------------------------|------------------------------------|-----------------------------|---|
| | H1-H3'proR | 6.6 | 10.7 | 1.2 | 6.3 | 10.7 | 4.8 | 1.5 | 6.7 | 10.7 | 4.8 | 1.3 | 7.0 |
| | H4'-H3'proS | 6.8; | 5.7 | 9.0 | 7.2 | 4.8 | 5.5 | 9.4 | 8.8 | 4.5 | 5.4 | 9.4 | 6.4 |
| | H4'-H3'proR | 7.8 ^(c) | 11.0 | 2.1 | 6.9 | 11.3 | 10.5 | 2.7 | 6.2 | 11.4 | 10.9 | 5.1 | 8.9 |
| Conf. J-fitted ratio | | | ⁴ C ₁ | ¹ S ₃ | ⁴ C ₁ : ¹ S ₃ 45:55 | ⁴ C ₁ | ¹ S ₃ | ¹ C ₄ | ⁴ C ₁ : ¹ S ₃ : ¹ C ₄ 60:24:16 | ⁴ C ₁ | ¹ S ₃ (0) | ¹ C ₄ | ⁴ C ₁ : ¹ C ₄ 80:20 |
| 2S | H1-H2 | 5.9 | 6.3 | 4.4 | 5.3 | 6.8 | 5.8 | 2.7 | 5.9 | 6.8 | - | 2.2 | 5.9 |
| | H2-H3 | 6.7 | 8.3 | 6.0 | 7.0 | 7.3 | 7.3 | 2.4 | 6.5 | 7.3 | - | 2.3 | 6.3 |
| | H3-H4 | 3.0 | 2.9 | 2.2 | 2.5 | 3.4 | 1.8 | 3.3 | 3.0 | 3.4 | - | 4.2 | 3.6 |
| | H4-H5 | 3.2 | 0.9 | 3.7 | 2.4 | 0.7 | 4.2 | 5.0 | 1.9 | 0.7 | - | 6.1 | 1.8 |
| | H1-H3'proS | 6.9; | 7.2 | 4.8 | 5.9 | 7.2 | 5.0 | 4.3 | 6.2 | 7.4 | - | 4.7 | 6.9 |
| | H1-H3'proR | 5.3 ^(c) | 10.0 | 1.4 | 5.2 | 8.8 | 1.5 | 1.8 | 5.9 | 8.7 | - | 1.6 | 7.3 |
| | H4'-H3'proS | 7.4; | 2.4 | 11.4 | 7.4 | 7.1 | 11.4 | 10.5 | 8.7 | 6.8 | - | 10.0 | 7.4 |
| | H4'-H3'proR | ? ^(c) | 8.5 | 5.1 | 6.6 | 8.6 | 4.2 | 6.2 | 7.2 | 7.5 | -- | 6.7 | 7.3 |

a) cluster-averaged J-values. b) Scalar coupling estimated for the J-fitted cluster ratio shown below c) The scalar coupling between these proton pairs could not be assigned: the corresponding values were not used in the fitting procedure.

NOE contact analysis

As discussed in the preceding sections, the three force fields have predicted strikingly different mixtures of ⁴C₁, ¹S₃ or ⁴C₁ conformations. As an additional evidence for the presence and characterization of these equilibria, a NOE-based analysis was also performed using the set of nOe contacts collected in Table 2. These contacts were selected because they were identified as exclusive for specific clusters, as shown in Tables 4-6. The force field-predicted interproton distances and NOE intensities (%) for each cluster were derived using a full matrix relaxation calculation⁸ extended to all the members of each cluster (Table 4-6 Noeprom columns). Again these values are somewhat force field dependent, however some generalizations can be drawn. For the (4'-R) isomer **1** the H1-H4', H3-H3'proS and H5-H3'proS NOE contacts are exclusive for the ⁴C₁ cluster, whereas the H1-H6proR contact is exclusive for either the skew boat or the inverted chair families. Contact H2-H4' is highly enhanced in the ¹S₃ cluster relative to the ⁴C₁, while the H3-H5 distance is similar in both the chair and skew boat conformations (around 2.5 and 2.7 Å respectively), but is much longer for the inverted chair ¹C₄ cluster (> 4.2 Å). In the case of the (4'-S) compound **2** the exclusive nOe contacts for the ⁴C₁ family are H4'-H3, H3-H3'proS and H5-H3'proS, while for the ¹S₃ cluster they are H1-H6proR and H2-H3'proR. Again, the H3-H5 distance is similar in both the chair and skew boat conformations (around 2.5 and 2.7 Å respectively), but is much longer for the inverted chair ¹C₄ cluster (> 4.2 Å).

Finally, the conformer distribution was estimated by fitting the Noeprom-based % nOe intensities to the experimental ones. As described for the J-fitting procedure, an initial value was assigned to the weight of each cluster and an estimation of the total % nOe intensity calculated; then, the deviation between the calculated and experimental values was iteratively minimized (up to 10.000 steps) and

thus, the best-fitting population weights were obtained. As observed above for the J-fitting process, a simpler procedure could have used the leader geometry as representative of each cluster in the fitting process.

Table 4. Use of the experimental nOe intensities to estimate the relative population of the AMBER* conformations

| Comp. | Proton pair | Noeprom Amber* | Noeprom Amber* | Exp. nOe (%) | Amber* |
|--|-------------|---|---|-----------------|---------------------------------|
| | | ⁴ C ₁ nOe (%) ^a (distance, Å) ^b | ¹ S ₃ nOe (%) ^c (distance, Å) ^b | | noe- fitted nOe(%) ^e |
| ⁴ C ₁ : ¹ S ₃ noe-fitted ratio 57:43 (J-fitted 60:40) ^d | | | | | |
| <u>1</u> | H1-H4' | 3.48 (2.62) | 0.00 (4.01) | 2.30 | 1.97 |
| | H3'proS-H3 | 8.55 (2.42) | 0.00 (4.63) | 3.70 | 4.85 |
| | H3'proS-H5 | 12.49 (2.25) | 0.00 (4.54) | 7.09 | 7.09 |
| | H3-H5 | 4.20 (2.50) | 2.63 (2.71) | 3.24 | 3.52 |
| | H1-H6proR | 0.00 (4.69) | 3.30 (2.57) | 3.01 | 1.43 |
| | H4'-H2 | 0.58 (3.51) | 0.94 (3.32) | 2.72 | 0.74 |
| ⁴ C ₁ : ¹ S ₃ noe-fitted ratio 40:60 (J-fitted 45:55) ^d | | | | | |
| <u>2</u> | H1-H3'proR | 4.31 (2.40) | 4.69 (2.38) | 3.94 | 4.54 |
| | H1-H3'proS | 0.46 (2.95) | 1.57 (2.70) | 0.43 | 1.13 |
| | H1-H6proR | 0.00 (4.69) | 3.92 (2.51) | 2.89 | 2.36 |
| | H5-H3 | 4.88 (2.52) | 3.12 (2.71) | 3.43 | 3.82 |
| | H4'-H3 | 1.90 (2.92) | 0.94 (3.34) | 2.14 | 1.32 |
| | H6proS-H4 | 3.44 (2.71) | 2.37 (2.88) | 2.61 | 2.80 |
| | H3'proS-H3 | 7.21 (2.50) | 0.00 (4.65) | 1.93 | 2.86 |
| | H3'proS-H5 | 12.84 (2.24) | 0.00 (4.57) | 5.10 | 5.10 |
| | H3'proR-H2 | 0.00 (4.12) | 3.36 (2.80) | 1.28 | 2.02 |

a) Noeprom value from a full-matrix cross peak intensity calculation extended to the AMBER* ⁴C₁ cluster. b) Distances between proton pairs Boltzmann-averaged over the cluster. c) Noeprom value from a full-matrix cross peak intensity calculation extended to the AMBER* ¹S₃ cluster. d) J-fitted relative populations from Table 3 e) NoeProm calculations for a cluster ratio corresponding to the fitted population in the preceding column. The sum of all deviations of the nine fitted NOE values relative to the experimental values are of 5.31% for **1** and 4.90% for **2**.

However, using cluster-averaged nOe intensities as a representation of each cluster allows to account for the different 5-membered ring puckerings represented in each cluster. The population thus estimated are reported in Tables 4, 5 and 6 (noe-fitted ratio columns) together with the nOe intensity values they originate (noe-fitted % nOe columns). J-fitted population estimates are also included in Table 4-6, for comparison. The noe data can be reproduced by all sets of fitted conformer populations with average deviations going from 2% to 13% (see Tables' legends). Populations obtained by J-fitting or nOe-fitting don't differ significantly when the AMBER* (Table 4) or OPLSAA (Table 6) geometries are used .

For instance, using AMBER* geometries both fitting procedures describe the system as an almost equimolar mixture of 4C_1 and 1S_3 conformations. On the contrary, when the MM3* geometries were used to fit the nOe intensity data, the best fit was obtained by excluding the 1C_4 inverted chair, that was still accepted (as 15 % or 27 % of the conformer population for **1** and **2**, respectively) by the J-fitting procedure. Thus the analysis of the nOe contacts, which is more sensitive than the scalar coupling constant analysis to small changes in the three-dimensional structure of the systems examined, appears to suggest that a conformer distribution including 4C_1 and 1S_3 geometries only reproduces the experimental data better than a distribution which also includes the inverted 1C_4 chair.

Table 5. Use of the experimental nOe intensities to estimate the relative population of the MM3* conformations

| Comp. | Proton pair | Noeprom MM3* 4C_1 nOe (%) ^a (distance, Å) ^b | Noeprom MM3* 1S_3 nOe (%) ^c (distance, Å) ^b | Noeprom MM3* 1C_4 nOe (%) ^d (distance, Å) ^b | Exp. nOe (%) | MM3* Noe- fitted nOe(%) ^f 4C_1 : 1S_3 : 1C_4 noe-fitted ratio 50:50:0 (J-fitted 48: 25:27) ^e |
|----------|-------------|--|--|--|-----------------|---|
| | | | | | | |
| 1 | H1-H4' | 3.79 (2.60) | 0.78 (3.22) | 0.00 (4.07) | 2.30 | 2.30 |
| | H3'proS-H3 | 7.57 (2.47) | 0.00 (3.89) | 0.00 (5.40) | 3.70 | 3.82 |
| | H3'proS-H5 | 13.14 (2.22) | 0.00 (4.28) | 0.00 (4.51) | 7.09 | 6.63 |
| | H3-H5 | 4.21 (2.50) | 2.25 (2.78) | 0.00 (4.31) | 3.24 | 3.24 |
| | H1-H6proR | 0.00 (4.91) | 2.33 (2.70) | 4.51 (2.39) | 3.01 | 1.15 |
| | H4'-H2 | 0.67 (3.44) | 4.47 (2.57) | 1.70 (3.00) | 2.72 | 2.55 |
| 2 | H1-H3'proR | 2.57 (2.51) | 4.31 (2.42) | 3.96 (2.45) | 3.94 | 3.94 |
| | H1-H3'proS | 3.51 (2.42) | 1.53 (2.74) | 1.86 (2.70) | 0.43 | 1.95 |
| | H1-H6proR | 0.00 (4.87) | 2.51 (2.67) | 6.76 (2.28) | 2.88 | 1.98 |
| | H5-H3 | 3.71 (2.63) | 2.76 (2.77) | 0.00 (4.31) | 3.43 | 2.96 |
| | H4'-H3 | 7.88 (2.34) | 2.28 (2.87) | 0.00 (4.43) | 2.14 | 3.46 |
| | H6proS-H4 | 2.25 (2.87) | 1.05 (3.20) | 0.00 (4.02) | 2.61 | 1.30 |
| | H3'proS-H3 | 1.15 (3.19) | 0.00 (4.43) | 0.00 (5.37) | 1.93 | 0.24 |
| | H3'proS-H5 | 5.35 (2.51) | 0.00 (4.52) | 0.00 (4.53) | 5.10 | 1.13 |
| | H3'proR-H2 | 0.00 (4.08) | 1.91 (3.04) | 4.83 (2.64) | 1.28 | 1.51 |
| | | | | | | 4C_1 : 1S_3 : 1C_4 noe-fitted ratio 21:79:0 (J-fitted 60:24:16) ^e |

a) Noeprom value from a full-matrix cross peak intensity calculation extended to the MM3* 4C_1 cluster. b) Boltzmann-averaged distances between proton pairs. c) Noeprom value from a full-matrix cross peak intensity calculation extended to the MM3* 1S_3 cluster. d) Noeprom value from a full-matrix cross peak intensity calculation extended to the MM3* 1C_4 cluster. e) J-fitted relative populations from Table 3. f) NoeProm calculations for a cluster ratio corresponding to the fitted population in the preceding column. The sum of all deviations of the nine fitted NOE values relative to the experimental values are of 2.59% for **1** and 11.39% for **2**.

Table 6. Use of the experimental nOe intensities to estimate the relative population of the OPLSAA conformations

| Comp. | Proton pair | Noeprom OPLSAA ⁴ C ₁ nOe (%) ^a (distance, Å) ^b | Noeprom OPLSAA ¹ C ₄ nOe (%) ^c (distance, Å) ^b | Exp. nOe (%) | OPLSAA Noe- fitted nOe(%) ^e |
|----------|-------------|--|--|-----------------|---|
| | | | | | ⁴ C ₁ : ¹ C ₄ fitted ratio 55:45 (J-fitted 60:40) ^d |
| 1 | H1-H4' | 4.27 (2.56) | 0.00 (3.70) | 2.30 | 2.33 |
| | H3'proS-H3 | 5.04 (2.66) | 0.00 (4.96) | 3.70 | 2.75 |
| | H3'proS-H5 | 12.98 (2.25) | 0.00 (4.06) | 7.09 | 7.09 |
| | H3-H5 | 3.20 (2.62) | 0.00 (4.14) | 3.24 | 1.75 |
| | H1-H6proR | 0.00 (4.54) | 6.05 (2.32) | 3.01 | 2.75 |
| | H4'-H2 | 0.81 (3.31) | 3.31 (2.69) | 2.72 | 1.95 |
| | | | | | ⁴ C ₁ : ¹ C ₄ fitted ratio 76:24 (J-fitted 80:20) ^d |
| 2 | H1-H3'proR | 2.15 (2.57) | 4.27 (2.39) | 3.94 | 2.65 |
| | H1-H3'proS | 3.59 (2.42) | 1.62 (2.62) | 0.43 | 3.13 |
| | H1-H6proR | 0.00 (5.02) | 3.01 (2.54) | 2.88 | 0.71 |
| | H5-H3 | 3.76 (2.63) | 0.00 (4.25) | 3.43 | 2.88 |
| | H4'-H3 | 6.20 (2.43) | 0.00 (4.48) | 2.14 | 4.74 |
| | H6proS-H4 | 1.70 (2.96) | 0.00 (3.86) | 2.61 | 1.30 |
| | H3'proS-H3 | 1.39 (3.12) | 0.00 (5.35) | 1.93 | 1.06 |
| | H3'proS-H5 | 4.48 (2.55) | 0.00 (4.47) | 5.10 | 3.43 |
| | H3'proR-H2 | 0.00 (4.05) | 5.46 (2.61) | 1.28 | 1.28 |

a) Noeprom value from a full-matrix cross peak intensity calculation extended to the OPLSAA ⁴C₁ cluster. b) Boltzmann-averaged distances between proton pairs. c) Noeprom value from a full-matrix cross peak intensity calculation extended to the OPLSAA ¹C₄ cluster. d) J-fitted relative populations from Table 3 e) NoeProm calculations for a cluster ratio corresponding to the fitted population in the preceding column. The sum of all deviations of the fitted values relative to the experimental values are of 3.51% for **1** and 13.17% for **2**.

Conclusions

The structure and dynamics of **1** and **2** have proven hard to predict using computational methods. The experimental data could not be interpreted based on a single conformation of the bicyclic structures. However, a combination of NMR spectroscopy and molecular modelling was finally viable to define the three-dimensional structures and dynamics of these fluxional molecules in water solution. We could also determine which of three well-known force fields are most reliable in reproducing the experimental observations, and should therefore be used in design studies.

The NMR data suggest that the pyrane ring in both molecules rapidly (on the NMR time scale) interconverts between two (4C_1 , 1S_3) or three (4C_1 , 1S_3 and 4C_1) conformations, and data-fitting indicates that the canonical pyranose 4C_1 chair represents about 50-60% of the total conformer population. Both NOE and J analysis concur in describing a fluxional system that samples at least two conformations for the pyrane ring, a 4C_1 chair and a 1S_3 skew-boat, with a possible contribution by a 1C_4 inverted chair. However, a large presence of this latter conformer, which is predicted as the global minimum by the OPLSAA force field, is unlikely, on the basis of the medium-strong H3-H5 nOe crosspeaks that could be observed for both **1** and **2**. Indeed, although the overall NOE fitting obtained using OPLSAA conformations is comparable with those obtained with MM3 or AMBER*, the intensity of the H3-H5 crosspeak is very poorly reproduced using the two chairs only (Table 6, OPLSAA force field), and much better reproduced using conformers of the 4C_1 and 1S_3 type (Table 4 and 5, AMBER and MM3). Indeed, noe-fitted populations tend to exclude the 1C_4 inverted chair also when it is included in the initial population, as in the MM3* case (Table 5). Hence, OPLSAA, which yields an inverted chair as its global minimum, appears to be the least predictive of the force fields tested.

Both the other force fields examined, AMBER* and MM3*, can account for the experimentally observed NMR data, but MM3* is less accurate in the description of the C5-C6 extra-annular torsion. Neither AMBER* or MM3* potential energy surfaces directly yield an accurate representation of conformer distribution: AMBER* tends to over-represent the 4C_1 chair, and MM3* the 1S_3 skew-boat. These errors are reflected in the poor reproduction of experimental results for both 1,2-scalar couplings (J) and nOe intensities calculated directly from the modeling-derived populations using full-matrix approximations. On the contrary, conformer distributions compatible with experiment can be back-calculated by fitting the J and nOe experimental values with scalar couplings and (full-matrix approximated) nOe intensities calculated for single clusters of conformations originating by molecular modeling. It was shown that the populations calculated with the two fitting procedures are in full agreement starting from the AMBER* geometries, but nOe analysis tends to discount the 1C_4 inverted chair found by MM3* and retained by the J-fitting procedure.

The stereoisomeric bicyclic structures described in this paper largely retain the structural characteristics of natural galactosides, and yet, as C-glycosides, they are stabilized against chemical and enzymatic hydrolysis. Hence, they could be developed as drug-like ligands and inhibitors of galactose-binding proteins. Indeed, cholera toxin B5 pentamer, a paradigmatic example of a galactose receptor, binds to **1** and to galactose with similar affinity, as measured by fluorescence spectroscopy.^{xiii} With the same approach, similar scaffolds could be obtained starting from different monosaccharides. The synthesis of these structures, which can be achieved in just four high-yield steps, without the need for tedious protection-deprotection sequences, makes them very attractive for library design, or diversity-oriented synthetic projects. Further studies are in progress to develop the synthetic potential of these structures.

Experimental section

NMR Methods

NMR spectra were recorded in D₂O using Bruker AVANCE 400 and 500 spectrometers, at 293K. 1D, as well as 2D-, COSY, HSQC and NOESY experiments were recorded using the standard pulse sequences. NOESY experiments were carried out using mixing times of 600ms and 800ms. Exclusive nOe contacts for both compounds were identified and integrated from the NOESY spectrum relative to the diagonal peak of the corresponding rows.

Experimental scalar coupling constants, $^3J_{H-H}$, were extracted from signal splittings and compared with those estimated for the 4C_1 , 1C_4 , and skew boat geometries calculated from the molecular mechanics calculations.

Cross-peaks from NOESY experiments were integrated using the MestReC software.^{xiv} Their volumes were compared with those estimated for Boltzmann-averaging by NOEprom^{xi} using a full-matrix relaxation approach, as described below.

Computational methods

All structures were investigated with MacroModel 8.1,^{vii} using GB/SA implicit water solvent,^{vi} constant dielectric and extended electrostatic treatment (8 Å for Van der Waals cut-off of interactions, and 20 Å as cut-off of electrostatic interactions). Calculations were performed using the AMBER*, OPLS-AA and MM3* force fields.

A 10000 step Monte Carlo/Multiple Minimization^{viii} was performed using 500 steps Polak-Ribiere Conjugate Gradient algorithm for minimization. All found minima converged to 0.05 kJ·A⁻¹ for the gradient value. The pyrane ring was allowed to open and a chirality check was added to avoid configuration changes. Other variables for the Monte Carlo search were all the C-C, C-O and C-N single bonds. The orientation of the hydroxyl groups was not included in the list of explicit variables. The results were analyzed taking into account all conformations found by each force field within the first 3 kcal/mol from the respective global minimum. The structures were clustered based on the conformation of the pyrane ring (4C_1 chair, 1C_4 inverted chair and 1S_3 skew-boat). Each cluster contains multiple conformations, which differ mainly by the puckering of the 5-membered ring and by the orientation of the amide side-chain (see Sup. Info.).

The nOe volumes were generated from the computational models using the NOEprom software based on a full matrix relaxation approach and using a rigid isotropic model.^{xv} The nOe volumes reported in the Tables correspond to a 400 MHz spectrometer, using 0.8 second mixing time and a correlation time of 60 ps. The fitting for the 500 MHz data and other mixing times produced basically the same conclusions as those reported here. The force-field predictions were obtained as Boltzmann-weighted averages of all the conformations calculated by each force field, within 3 kcal/mol from the global minimum. The cluster-averaged nOe values of Tables 4-6 were calculated as Boltzmann

distributions over the conformers belonging to individual clusters. These cluster-averaged nOe volumes were then used as input for the fitting processes.

Similarly, cluster averaged values of the scalar couplings ${}^3J_{H-H}$ were estimated using the empirical generalization of Karplus equation proposed by Haasnoot et al.^{xii} as Boltzmann averages over the members of individual clusters. These cluster-averaged J values were used in the following fitting procedure to obtain J-fitted cluster populations.

Molecular dynamics calculations were performed with AMBER* and MM3* using stochastic dynamics (SD).^{xvi} (Mixed-Mode Monte Carlo- stochastic dynamics^{xvii} can not be used for cyclic structures). Simulations were run for 1ns, using the same force field, solvent and non-bonded interactions treatment used for the MC/MM search. The time step was set to 1 fs, the temperature to 300 ° K and the frictional coefficient to 0.5 ps⁻¹. Conformational sampling was done every pico-second. Under these conditions, SD simulations are not likely to have reached convergence, therefore, they were not used to extract the expected nOe intensities, but only used to assess the dynamic flexibility of the ring system (see Suppl. Info)

Fitting procedures

Fitting of the experimental ${}^3J_{HH}$ values was obtained by using the Simplex algorithm.^{xviii} An initial value was assigned to the weight of a given cluster and an estimation of the average scalar coupling calculated; then, the deviation between the calculated and experimental values was iteratively minimized (up to 10.000 steps) and thus, the best-fitting population weights were obtained. The same procedure was applied to estimate the fitted conformational distributions from NOE intensities.

ⁱ a) Cipolla, L.; Peri, F.; La Ferla, B.; Redaelli C.; Nicotra, F. *Curr. Org. Synthesis* **2005**, 2, 153; b) Gruner, S. A. W.; Locardi, E.; Lohof, E.; Kessler, H. *Chem. Rev.* **2002**, 102, 491; c) Le, G. T.; Abbenante, G.; Becker, B.; Grathwohl, M.; Halliday, J.; Tometzki, G.; Zuegg, J.; Meutermans, W. *DDT* **2003**, 8, 701; d) Oliver, S. F.; Abell, C. *Curr. Opin. Chem. Biol.* **1999**, 3, 291

ⁱⁱ Bennek, J. A.; Gray, G. A., *J. Org. Chem.* **1987**, 52, 892-897. See also Broxterman, H. J. G.; Kooreman, P. A.; Elst, H. van den; Roelen, H. C. P. F.; Marel, G. A. van der; Boom, J. H. van, *Recl.Trav.Chim.Pays-Bas* **1990**, 109, 583-590.

ⁱⁱⁱ In the original paper, 5 eq of TMSOTf and 5 eq of TMAH were used.

^{iv} a) Cipolla, L.; Lay, L.; Nicotra, F. *J. Org. Chem.* **1998**, 62, 6678-6681; b) Lay, L.; Cipolla, L.; La Ferla, B.; Peri, F.; Nicotra, F. *Eur. J. Org. Chem.* **1999**, 3437-3440

^v Asensio, J.; Jimenez-Barbero, J., *Biopolymers*, **1995**, 35, 55-73

^{vi} Still, W. C.; Tempzyk, A.; Hawley, R.; Hendrickson, T., *J. Am. Chem. Soc.* **1990**, 112, 6127-6129.

^{vii} Mohamadi, F.; Richards, N.G.J.; Guida, W.C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; W. C. Still, *J. Comput. Chem.* **1990**, 11, 440-467.

^{viii} Kolb, H.C.; Ernst, B.; *Chem. Eur. J.* **1997**, 3, 1571-1578.

^{ix} Béerces, A., Whitfield, D.M. and Nukada, T. *Tetrahedron*, **2001**, 57, 477-491

^x a) Borgias, B.A.; James, T.L. *Methods Enzymol.* **1989**, 176, 169-183; b) B. A. Borgias, P. D. Thomas, T. L. James, "Complete Relaxation Matrix Analysis (CORMA)", University of California, San Francisco, 1987, 1989.

^{xi} Noeprom software is free downloadable at <http://desoft03.usc.es/mmartin/software.html>

^{xii} Haasnoot CAG, De Leeuw FAAM, Altona C, *Tetrahedron*, **1980**, 36, 2783-2792

^{xiii} Ingrid Velter, unpublished results

^{xiv} www.mestrec.com

^{xv} Lipari, G.; and Szabo, A., *J. Am. Chem. Soc.*, **1982**, 104, 4546-4559

- ^{xvi} Guarnieri, F.; Still, W. C.; *J. Comput. Chem.* **1994**, 15, 1302 –1310.
^{xvii} Kolb, H.C.; Ernst, B.; *Chem. Eur. J.* **1997**, 3, 1571 –1578.
^{xviii} George B. Dantzig “Linear Programming and Extension” (1998) Princeton University Press [Chapter 5: The Simplex Method]

The DC-SIGN[#] receptor

Dendritic cells (DC) are important sentinels of our immune system that defend against invading pathogens. These professional antigen-presenting cells are seeded throughout peripheral tissues to monitor for pathogens, which they capture and process to antigenic fragments (Banchereau et al. 1998). After microbial or inflammatory stimuli, immature DC undergoes a process of maturation and migrate to secondary lymphoid organs to present processed antigens to naïve T cells. Antigen-specific T cells differentiate into effector T cells that are instrumental in combating infections. DC express on its surface diverse C-type lectins, such as the mannose receptor (MR) (Sallusto et al. 1995), DEC-205 (Kato et al. 2000), BCDA-2 (Dzionic et al. 2001), DC-ASGPR (Valladeau et al. 2001) and, particularly for this case, DC-SIGN receptor.

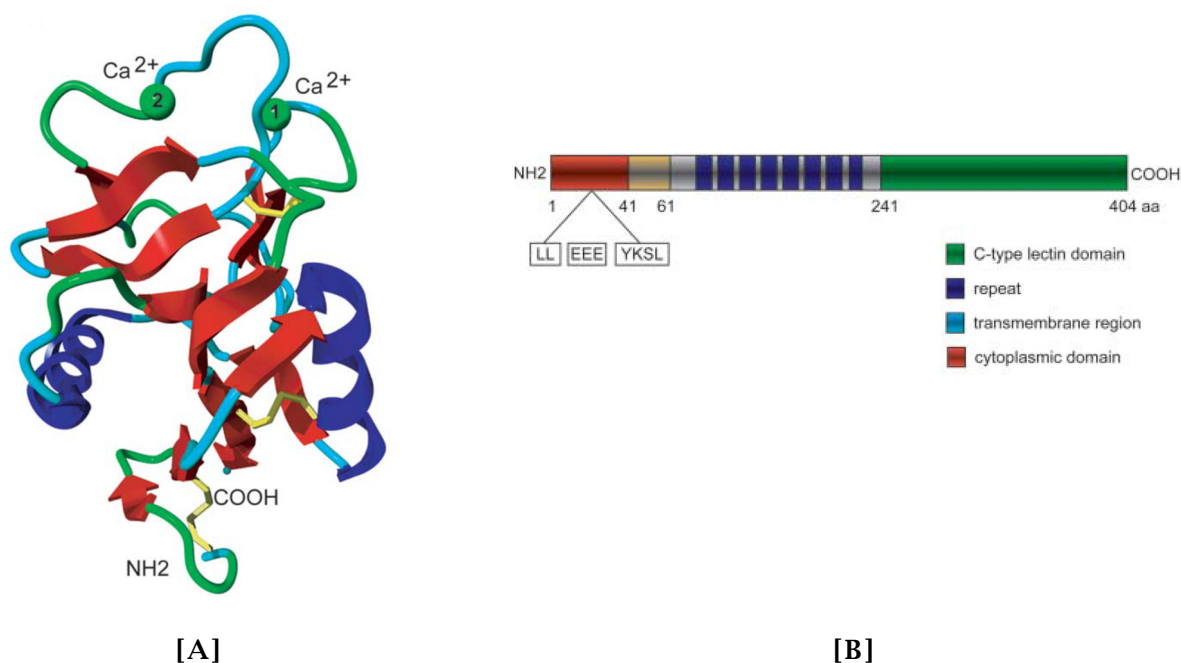


Figure 34. Structure of the C-type lectin DC-SIGN. [A] Ribbon diagram of the C-type lectin domain of DC-SIGN. The two α -helices are shown in blue, the β -strands in red, the calcium ions at site 1 and 2 in green, and the three disulphide bridges in yellow. Both the N- and the C-termini are at the bottom of the image. [B] Schematic structure of DC-SIGN. DC-SIGN is a type II trans-membrane C-type lectin; the extracellular N-terminus contains the C-type lectin domain. The intracellular C-terminus contains three internalization motifs.

DC-SIGN is a type II transmembrane protein that, based on its structure, belongs to the C-type lectin family (Fig. 34) (Alvarez et al. 2002; Colmenares et al. 2002; Curtis et al. 1992; Geijtenbeek et al. 2000; Geijtenbeek et al. 2003; Halary et al. 2002). DCSIGN contains a short cytoplasmic N-terminal domain with several intracellular sorting motifs, an extracellular stalk of seven complete and one partial tandem repeat, and a C-terminal lectin domain with affinity for mannose-containing carbohydrates (Fig. 34).

[#] The acronym stays for “**D**endritic **C**ell **s**pecific **i**ntercellular adhesion molecule (ICAM)-3 **g**rabbing **n**on-integrin”

As far as we know, DC-SIGN has two major roles. It can act as adhesion receptor during the initial recognition of T cells with DC, which is an antigen-independent interaction (Fig. 35A); but it can also function as an antigen receptor for pathogenic antigens, similarly to other C-type lectins (Fig 35B) (Geijtenbeek et al. 2003).

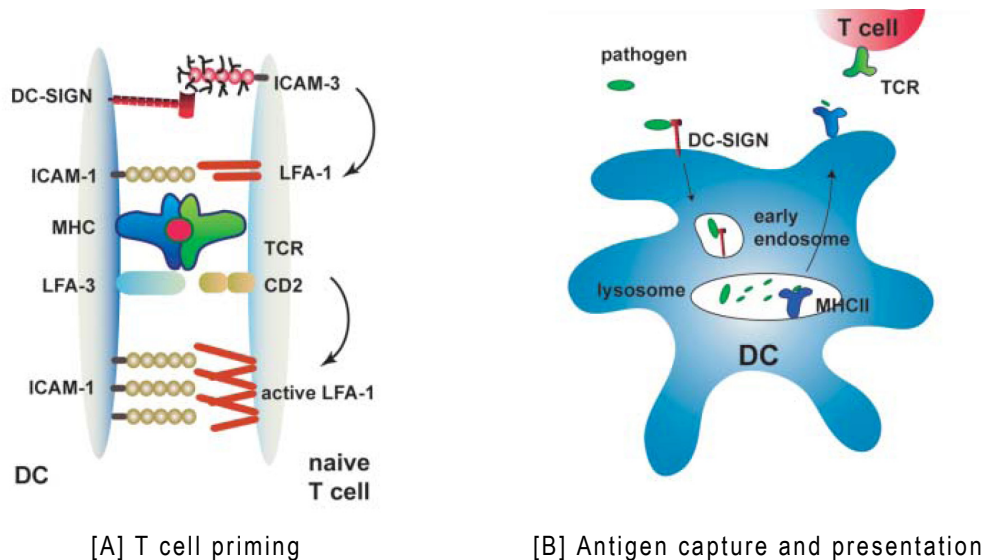


Figure 35. [A] Initial DC-T cell clustering, necessary for an efficient immune response, is mediated by transient interactions between DC-SIGN and ICAM-3. This interaction facilitates the formation of low-avidity LFA-1/ICAM-1 interaction and scanning of the antigen-MHC repertoire. [B] DC-SIGN also functions as an antigen receptor. DC-SIGN rapidly internalizes upon binding soluble ligand and is targeted to late endosomes/lysosomes, where antigens are processed and presented by MHC class II molecules.

Although the first mentioned role involves binding to intercellular adhesion molecule 3 (ICAM-3), which originally led to receptor's designation; the second function, pathogen's recognition, is believed to involve binding of surface carbohydrates and it is hence interesting for our purposes.

The presence of the C-type carbohydrate recognition domain (CRDs) in DC-SIGN is consistent with the receptor's role in the recognition of carbohydrate ligands. Binding to high mannose oligosaccharide underlies the interaction of DC-SIGN with enveloped viruses (Cambi et al. 2003), including the human immunodeficiency virus (HIV), hepatitis C virus and Ebola virus (van Kooyk et al. 2003; van Kooyk et al. 2003). Binding of HIV to DC-SIGN on dendritic cells enhances infection on T cells, but in some cases, DC-SIGN can serve directly as a route for infection of cells (Alvarez et al. 2002; Simmons et al. 2003). For instance, the interaction of DC-SIGN with mannose-containing glycans underlies the dendritic cell binding of *Mycobacterium tuberculosis*, yeast *Candida albicans* and *Leishmania* parasites (van Kooyk et al. 2003; van Kooyk et al. 2003). DC-SIGN also binds to fucose-containing glycans (Frison et al. 2003); including those found on schistosomes *Helicobacter pylori* (Appelmek et al. 2003). The carbohydrate specificity has recently been screened by Drickamer and co-workers, who analyzed in depth the structure complexes of DC-SIGN bound to the two classes of ligands (Guo et al. 2004). In particular, it is known that a mannan ligand is able to block HIV's gp120 interaction with DC-SIGN (Curtis et al. 1992; Geijtenbeek et al. 2000) and from this background, we thought on the

possibility of detecting the mannan-DC-SIGN interaction by NMR directly employing dendritic cells. In the following paper a novel living-cell NMR technique is discussed.

References

- Alvarez, C. P., F. Lasala, J. Carrillo, O. Muniz, A. L. Corbi and R. Delgado (2002) "C-type lectins DC-SIGN and L-SIGN mediate cellular entry by Ebola virus in cis and in trans" *J Virol* **76**(13): 6841-4.
- Appelmelk, B. J., I. van Die, S. J. van Vliet, C. M. Vandenbroucke-Grauls, T. B. Geijtenbeek and Y. van Kooyk (2003) "Cutting edge: carbohydrate profiling identifies new pathogens that interact with dendritic cell-specific ICAM-3-grabbing nonintegrin on dendritic cells" *J Immunol* **170**(4): 1635-9.
- Banchereau, J. and R. M. Steinman (1998) "Dendritic cells and the control of immunity" *Nature* **392**(6673): 245-52.
- Cambi, A. and C. G. Figdor (2003) "Dual function of C-type lectin-like receptors in the immune system" *Curr Opin Cell Biol* **15**(5): 539-46.
- Colmenares, M., A. Puig-Kroger, O. M. Pello, A. L. Corbi and L. Rivas (2002) "Dendritic cell (DC)-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin (DC-SIGN, CD209), a C-type surface lectin in human DCs, is a receptor for Leishmania amastigotes" *J Biol Chem* **277**(39): 36766-9.
- Curtis, B. M., S. Scharnowske and A. J. Watson (1992) "Sequence and expression of a membrane-associated C-type lectin that exhibits CD4-independent binding of human immunodeficiency virus envelope glycoprotein gp120" *Proc Natl Acad Sci U S A* **89**(17): 8356-60.
- Dzionic, A., Y. Sohma, J. Nagafune, M. Cella, M. Colonna, F. Facchetti, G. Gunther, I. Johnston, A. Lanzavecchia, T. Nagasaka, T. Okada, W. Vermi, G. Winkels, T. Yamamoto, M. Zysk, Y. Yamaguchi and J. Schmitz (2001) "BDCA-2, a novel plasmacytoid dendritic cell-specific type II C-type lectin, mediates antigen capture and is a potent inhibitor of interferon alpha/beta induction" *J Exp Med* **194**(12): 1823-34.
- Frison, N., M. E. Taylor, E. Soilleux, M. T. Bousser, R. Mayer, M. Monsigny, K. Drickamer and A. C. Roche (2003) "Oligolysine-based oligosaccharide clusters: selective recognition and endocytosis by the mannose receptor and dendritic cell-specific intercellular adhesion molecule 3 (ICAM-3)-grabbing nonintegrin" *J Biol Chem* **278**(26): 23922-9.
- Geijtenbeek, T. B., R. Torensma, S. J. van Vliet, G. C. van Duijnhoven, G. J. Adema, Y. van Kooyk and C. G. Figdor (2000) "Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses" *Cell* **100**(5): 575-85.
- Geijtenbeek, T. B., S. J. Van Vliet, E. A. Koppel, M. Sanchez-Hernandez, C. M. Vandenbroucke-Grauls, B. Appelmelk and Y. Van Kooyk (2003) "Mycobacteria target DC-SIGN to suppress dendritic cell function" *J Exp Med* **197**(1): 7-17.
- Geijtenbeek, T. B. H. and Y. van Kooyk (2003) "Pathogens target DC-SIGN to influence their fate. DC-SIGN functions as a pathogen receptor with broad specificity" *APMIS* **111**: 698-714.
- Guo, Y., H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis and K. Drickamer (2004) "Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR" *Nat Struct Mol Biol* **11**(7): 591-8.
- Halary, F., A. Amara, H. Lortat-Jacob, M. Messerle, T. Delaunay, C. Houles, F. Fieschi, F. Arenzana-Seisdedos, J. F. Moreau and J. Dechanet-Merville (2002) "Human cytomegalovirus binding to DC-SIGN is required for dendritic cell infection and target cell trans-infection" *Immunity* **17**(5): 653-64.
- Kato, M., T. K. Neil, D. B. Fearnley, A. D. McLellan, S. Vuckovic and D. N. Hart (2000) "Expression of multilectin receptors and comparative FITC-dextran uptake by human dendritic cells" *Int Immunol* **12**(11): 1511-9.

- Sallusto, F., M. Cella, C. Danieli and A. Lanzavecchia (1995) "Dendritic cells use macropinocytosis and the mannose receptor to concentrate macromolecules in the major histocompatibility complex class II compartment: downregulation by cytokines and bacterial products" *J Exp Med* **182**(2): 389-400.
- Simmons, G., J. D. Reeves, C. C. Grogan, L. H. Vandenberghe, F. Baribaud, J. C. Whitbeck, E. Burke, M. J. Buchmeier, E. J. Soilleux, J. L. Riley, R. W. Doms, P. Bates and S. Pohlmann (2003) "DC-SIGN and DC-SIGNR bind ebola glycoproteins and enhance infection of macrophages and endothelial cells" *Virology* **305**(1): 115-23.
- Valladeau, J., V. Duvert-Frances, J. J. Pin, M. J. Kleijmeer, S. Ait-Yahia, O. Ravel, C. Vincent, F. Vega, Jr., A. Helms, D. Gorman, S. M. Zurawski, G. Zurawski, J. Ford and S. Saeland (2001) "Immature human dendritic cells express asialoglycoprotein receptor isoforms for efficient receptor-mediated endocytosis" *J Immunol* **167**(10): 5767-74.
- van Kooyk, Y., B. Appelmelk and T. B. Geijtenbeek (2003) "A fatal attraction: Mycobacterium tuberculosis and HIV-1 target DC-SIGN to escape immune surveillance" *Trends Mol Med* **9**(4): 153-9.
- van Kooyk, Y. and T. B. Geijtenbeek (2003) "DC-SIGN: escape mechanism for pathogens" *Nat Rev Immunol* **3**(9): 697-709.

1D Saturation Transfer Difference NMR Experiments on Living Cells: The DC-SIGN/Oligomannose Interaction**

Silvia Mari, Diego Serrano-Gómez, F. Javier Cañada, Angel L. Corbí,* and Jesús Jiménez-Barbero*

Dedicated to Professor Luis Castedo on the occasion of his 65th birthday

Molecular recognition by specific targets is at the heart of the drug discovery process. Recently, it has been demonstrated that NMR screening is ideal for finding ligands that bind to a receptor.^[1] Although modern NMR experiments and technologies permits the exploration of the interaction process using nanomole quantities of the receptor, in some cases the availability of the target protein might pose a problem in the detection of the molecular recognition event, especially in the case of membrane-bound proteins. On this basis, we hypothesized that receptor-rich living cells could be used as a first step to screen the binding of different molecules to some types of receptors. The study of the role of the carbohydrate-binding C-type lectin and lectin-like receptors in the immune system is a topic of current interest. In particular, the dendritic-cell-specific ICAM-3-grabbing non-integrin (DC-SIGN) receptor^[2] mediates the binding and internalization of a large array of pathogens including HIV,^[3] Ebola virus,^[4] SARS-CoV,^[5] *Leishmania amastigotes*,^[6] *Mycobacterium tuberculosis*,^[7] *Schistosoma mansoni*, *Helicobacter pylori*,^[8] and *Candida*^[9] in a Man- or Fuc-dependent manner.^[10] Moreover, DC-SIGN plays an important role in the establishment of the initial contact between the dendritic cell and naive T lymphocytes through recognition of ICAM-3,^[2] and also mediates dendritic cell trafficking through interactions with endothelial ICAM-2.^[11] In this context, *Saccharomyces cerevisiae* mannan has been shown to compete with most of these DC-SIGN-dependent pathogen recognitions.^[12] Herein, we demonstrate that the direct interaction between the DC-SIGN receptor

and *S. cerevisiae* mannan may be easily detected by simple 1D saturation transfer difference (STD) NMR experiments^[13] that employ living cells directly, without the necessity for isolating the protein receptor. This evidence may open new avenues in the drug discovery process, since NMR spectroscopy may also provide key structural information.

The cell line K562 transfected with the prototype DC-SIGN, namely K562-CD209 (DC-SIGN+),^[12] and mock-transfected K562 (DC-SIGN-) were employed. These two cell types differ only in the expression of the receptor DC-SIGN on the cell surface. Thus, the untransfected cells were used as a control. The K562-CD209 DC-SIGN expression level was determined by flow cytometric analysis (see the Supporting Information). The *S. cerevisiae* mannan was obtained from Sigma and its average molecular weight was estimated to be 100 kDa by DOSY NMR experiments^[14] (see the Supporting Information).

To perform the STD NMR experiments, about 5×10^6 cells were counted, washed, and dissolved in deuterated phosphate-buffered saline (PBS), and mannan (5 mg) was dissolved in the same deuterated PBS. Typical NMR tube volumes were 600 μ L. Additional experiments with a commercial beta-glucan preparation, which does not bind to DC-SIGN, were also performed as blank experiments. All the experiments were repeated three times, and carried out using different batches. The estimated concentration of DC-SIGN in the NMR tube was approximately 0.1 μ M.^[15,16]

NMR control experiments were recorded in all cases. As illustrated in Figure 1, the STD control spectrum (Figure 1d) of mannan confirmed that the on-resonance irradiation ($\delta = 6.8$ ppm, aromatic region) did not affect the mannan signals.

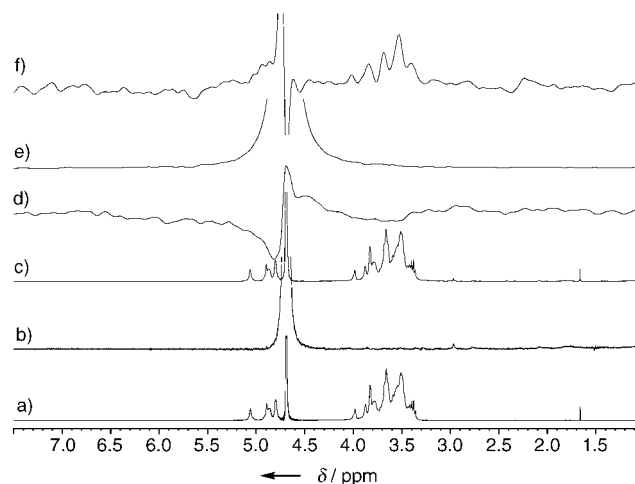


Figure 1. a) ^1H NMR spectrum of mannan in PBS; b) ^1H NMR spectrum of K562-CD209 cells in PBS, enhanced 20 \times ; c) ^1H NMR spectrum of mannan with K562-CD209 cells in PBS; d) STD reference spectrum of mannan: number of scans (NS) = 256, on-resonance frequency = 6.8 ppm, off-resonance frequency = 100 ppm, total saturation time = 2 s, enhanced 100 \times ; e) STD reference spectrum of K562-CD209 cells under the same experimental conditions as (d), enhanced 600 \times , signals enhanced 16 \times ; f) STD spectrum of mannan with K562-CD209 cells: NS = 64, on-resonance frequency = 6.8 ppm, off-resonance frequency = 100 ppm, signals enhanced 800 \times . All the samples (cells and mannan) were dissolved in deuterated PBS at pH 7.3 containing CaCl_2 (1 mM). Total sample volumes were 600 μ L.

[*] S. Mari,^[1] D. Serrano-Gómez,^[1] Dr. F. J. Cañada, Prof. Dr. A. L. Corbí, Prof. Dr. J. Jiménez-Barbero
Centro de Investigaciones Biológicas
Departamentos de Inmunología y de Estructura de Proteínas
C.S.I.C., c/Ramiro de Maeztu 9, 28040 Madrid (Spain)
Fax: (+34) 91-5360-432
E-mail: acorbi@cib.csic.es
jjbarbero@cib.csic.es

[[†]] S. Mari and D. Serrano-Gómez have contributed equally to this work and the order of authors is arbitrary.

[**] We thank the Ministerio de Educación y Ciencia of Spain (grants BQU2003-03550-CO3-01 and SAF2002-04615-CO2-01) and the European Union (HPRN-CT2002-0173) for financial support. We thank Dr. Antonio Leal for the gift of the mannan sample and Dr. Leal and Dr. P. Groves for helpful discussions. DC-SIGN = dendritic-cell-specific ICAM-3-grabbing non-integrin.

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

Identical results were obtained with saturation at $\delta = 1.3$ ppm (aliphatic side-chain region). Control NMR data of the living cells were acquired in an analogous manner (Figure 1 b and e). The cells used to obtain these spectra were then centrifuged and the pellet was dissolved in the mannan solution used to generate the spectra in Figure 1 a and d.^[17] The resulting reference spectrum is shown in Figure 1 c, while its corresponding STD NMR spectrum is displayed in Figure 1 f. The mannan signals are observed unambiguously. Thus, irradiation at the aromatic or aliphatic regions protons of the DC-SIGN receptor protein expressed in living cells results in transfer of magnetization to the polysaccharide protons. The control spectra support the notion that the observed process is specific for the interaction of living cells of K562-CD209 (DC-SIGN+) with mannan. Indeed, when the same set of experiments was repeated with mock-transfected K562 cells (DC-SIGN-) instead of K562-CD209 (DC-SIGN+) cells, no mannan signals were detected in the STD NMR experiment (Figure 2 f). Additional reference experi-

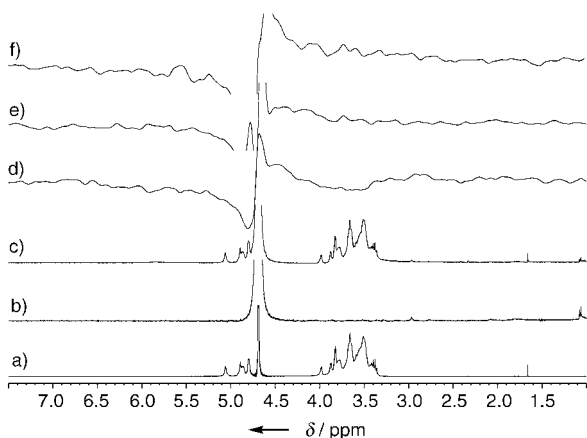


Figure 2. a) ^1H NMR spectrum of mannan in PBS; b) ^1H NMR spectrum of K562 (DC-SIGN-) cells in PBS enhanced 4 \times ; c) ^1H NMR spectrum of mannan with K562 (DC-SIGN-) cells in PBS enhanced 10 \times ; d) STD reference spectrum of mannan: NS = 256, on-resonance frequency = 6.8 ppm, off-resonance frequency = 100 ppm, total saturation time = 2 s, enhanced 600 \times ; e) STD reference spectrum of K562 (DC-SIGN-) cells under the same experimental conditions as (d), signals enhanced 400 \times ; f) STD spectrum of mannan with K562 (DC-SIGN-) cells: NS = 64, on-resonance frequency = 6.8 ppm, off-resonance frequency = 100 ppm, signals enhanced 800 \times .

ments were performed using beta-glucan as the ligand. In this case, and as expected for a molecule that does not bind to DC-SIGN, no STD signals were observed with either cell line (data not shown), which supports the NMR-based detection of the specific interaction between the K562-CD209 cells and mannan.

Direct detection of interactions between a ligand and a membrane protein by employing living cells could open new possibilities in the screening of compound libraries in a fast and efficient way. By using our current experimental setup for this K562-CD209/mannan system robust data have been obtained with 32 scans, which correspond to a total exper-

imental time of 3 minutes, and includes a recycling of the mannan solution from a previous experiment. The cells were checked by optical microscopy before and after the NMR experiments to determine their biological stability, and cell viability was also evaluated by the Trypan Blue exclusion method. Both assays demonstrated that cell viability was not significantly affected during the NMR experiments (see the Supporting Information).

In conclusion, receptor–ligand interactions with integral membrane proteins may be investigated by 1D STD NMR spectroscopy as the initial step in screening processes, which use living cells directly in a manner similar to that previously reported for detecting virus–ligand and platelet–ligand interactions.^[16] This protocol represents a significant advance in molecular recognition studies, since it eliminates time-consuming purification processes, especially in the case of receptors that are difficult to isolate.^[18]

Experimental Section

Cells: K562 cells were cultured in RPMI media supplemented with 10% fetal calf serum (FCS), 2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid (HEPES, 25 mM), and glutamine (2 mM; complete medium). DC-SIGN-expressing K562 transfectants (K562-CD209) have been previously described,^[12] and were cultured in complete medium supplemented with G418 (300 $\mu\text{g mL}^{-1}$, Gibco).

NMR experiments: All the experiments were recorded on a Bruker 500-MHz instrument at 283 K. A basic STD sequence was used, with the on-resonance frequency variable between $\delta = 6.8$ and 1.3 ppm.^[13] The off-resonance frequency was maintained at 100 ppm. A train of 40 Gaussian-shaped pulses of 50 ms each was employed, with a total saturation time of the protein envelope of 2 s. All the samples (cells and mannan) were dissolved in deuterated PBS at pH 7.3 containing CaCl_2 (1 mM) previously exchanged with D_2O . Total sample volumes were 600 μL .

Flow cytometry and antibodies: Cellular phenotypic analysis was carried out by indirect immunofluorescence. Monoclonal antibodies used for cell-surface staining included 9E10 (anti-c-Myc) and MR-1 (anti-DC-SIGN, CD209).^[12] All incubations were performed in the presence of human immunoglobulin G (50 $\mu\text{g mL}^{-1}$) to prevent binding through the Fc portion of the antibodies. Flow cytometric analysis was performed with an EPICS-CS instrument (Coulter Ceníffica, Madrid, Spain) using log amplifiers.

Received: August 7, 2004

Keywords: drug design · molecular recognition · NMR spectroscopy · proteins · receptors

- [1] B. Meyer, T. Peters, *Angew. Chem.* **2003**, *115*, 890; *Angew. Chem. Int. Ed.* **2003**, *42*, 864.
- [2] T. B. Geijtenbeek, R. Torensma, S. J. van Vliet, G. C. van Duijn-hoven, G. J. Adema, Y. van Kooyk, C. G. Figdor, *Cell* **2000**, *100*, 575.
- [3] T. B. Geijtenbeek, D. S. Kwon, R. Torensma, S. J. van Vliet, G. C. van Duijn-hoven, J. Middel, I. L. Cornelissen, H. S. Nottet, V. N. KewalRamani, D. R. Littman, C. G. Figdor, Y. van Kooyk, *Cell* **2000**, *100*, 587.
- [4] C. P. Alvarez, F. Lasala, J. Carrillo, O. Muniz, A. L. Corbi, R. Delgado, *J. Virol.* **2002**, *76*, 6841.
- [5] Z. Y. Yang, Y. Huang, L. Ganesh, K. Leung, W. P. Kong, O. Schwartz, K. Subbarao, G. J. Nabel, *J. Virol.* **2004**, *78*, 5642.

- [6] M. Colmenares, A. Puig-Kroger, O. M. Pello, A. L. Corbi, L. Rivas, *J. Biol. Chem.* **2002**, *277*, 36766.
- [7] L. Tailleux, O. Schwartz, J. L. Herrmann, E. Pivert, M. Jackson, A. Amara, L. Legres, D. Dreher, L. P. Nicod, J. C. Gluckman, P. H. Lagrange, B. Gicquel, O. Neyrolles, *J. Exp. Med.* **2003**, *197*, 121.
- [8] B. J. Appelmelk, I. van Die, S. J. van Vliet, C. M. Vandembroucke-Grauls, T. B. Geijtenbeek, Y. van Kooyk, *J. Immunol.* **2003**, *170*, 1635.
- [9] A. Cambi, K. Gijzen, J. M. de Vries, R. Torensma, B. Joosten, G. J. Adema, M. G. Netea, B. J. Kullberg, L. Romani, C. G. Figdor, *Eur. J. Immunol.* **2003**, *34*, 532.
- [10] A. Cambi, C. G. Figdor, *Curr. Opin. Cell Biol.* **2003**, *15*, 539.
- [11] T. B. Geijtenbeek, D. J. Krooshoop, D. A. Bleijs, S. J. van Vliet, G. C. van Duijnhoven, V. Grabovsky, R. Alon, C. G. Figdor, Y. van Kooyk, *Nat. Immunol.* **2000**, *1*, 353.
- [12] M. Relloso, A. Puig-Kroger, O. M. Pello, J. L. Rodriguez-Fernandez, G. de la Rosa, N. Longo, J. Navarro, M. A. Munoz-Fernandez, P. Sanchez-Mateos, A. L. Corbi, *J. Immunol.* **2002**, *168*, 2634.
- [13] M. Mayer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784. The success of the STD NMR experiments depends on the kinetics of the dissociation process and the molar ratio of ligand versus receptor. For a seminal discussion, see ref. [1].
- [14] P. Groves, M. O. Rasmussen, M. D. Molero, E. Samain, F. J. Canada, H. Driguez, J. Jimenez-Barbero, *Glycobiology* **2004**, *14*, 451.
- [15] F. Baribaud, S. Pöhlmann, G. Leslie, F. Mortari, R. W. Doms, *J. Virol.* **2002**, *76*, 9135.
- [16] a) A. Benie, R. Moser, E. Bäuml, D. Blaas, T. Peters, *J. Am. Chem. Soc.* **2003**, *125*, 14; b) B. Meyer, personal communication.
- [17] The NMR solution stays relatively clear for about 10 minutes. After this time, the cells start to sediment. However, the high sensitivity of the STD NMR experiment^[1] allows one to overcome this point without major problems.
- [18] The use of the STD technique together with high-resolution magic angle spinning may even open the window of applicability of the protocol described herein; see J. Klein, R. Meinecke, M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **1999**, *121*, 5336.

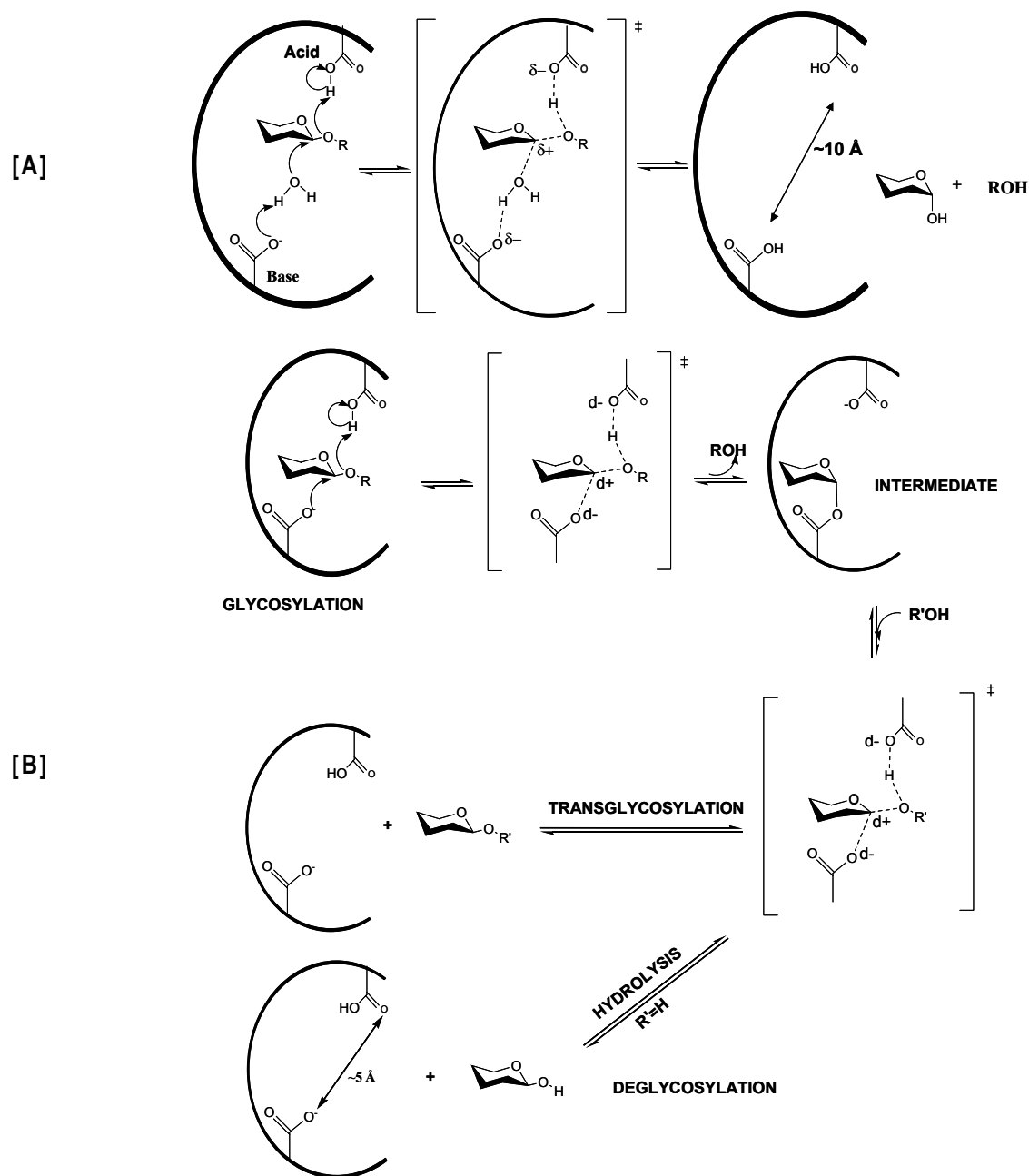
Glycosidases and Glycomimetics

Glycosidases are a large group of enzymes responsible of the degradation of poli- and oligosaccharides by hydrolysis of the glycosidic linkages. Dysfunction or deficiency of human glycosidases could lead to several diseases, like cancer (Goss et al. 1995; Yarema et al. 1998), intolerance to lactose (Sebastio et al. 1989), or pathologies relative to lysosomal storage (Kolter et al. 1999; Neufeld 1991). In that sense, knowledge of the structure and the mechanism of catalysis would help to understand the molecular basis of these processes, and the possibility to develop specific drugs.

Glycosidases: mechanism of reaction

The hydrolysis of the glycosidic linkage occurs following a general acidic catalysis mechanism, where the presence of two residues is fundamental: an acidic proton donor, and a nucleophile or basic residue (Koshland et al. 1953; McCarter et al. 1994; White et al. 1997; Zechel et al. 2000).

Based on the stereochemistry of the reaction they catalyze, the glycosidase families can be divided into two main groups. From one hand, there are those enzymes that act with retention of configuration at the anomeric centre, and hence, the anomeric configuration in the product is the same as in the substrate. On the other hand, there are other glycosidases that act with inversion of configuration, and the anomeric configuration in the product is the opposite of that in the substrate. It is generally accepted that, in both cases, the transition state is characterized by a distortion of the pyranose ring at the non-reducing end. In the case of the mechanism of inversion of configuration, two residues on the enzyme are fundamental: one residue acts as acid, and the other as basic catalyst. The two residues are approximately at 10 Å distance, which allows the entrance of the substrate and of one water molecule. As depicted in figure 36A, the reaction occurs in a single step, which implies the formation of an oxocarbenium intermediate. On the other hand, for those enzymes that act with retention of configuration, the available space between the protein residues only is around 5 Å. In this case, hydrolysis occurs in a two steps mechanism, in which the reaction intermediate is constituted by a covalently linked enzyme-substrate system (Fig. 36B). In the more complex case, [B], as proposed initially by Koshland in 1953, (Koshland et al. 1953) a residue, which is normally a carboxylic acid, acts as the general acid catalyst, and protonates the interglycosidic oxygen of the substrate. This leads also to an improvement of the properties of the remaining aglyconic part, that might act as a better leaving group. As consequence of this protonation, the pyranose ring distortion is also facilitated (White et al. 1997), and hence a weakening of the interglycosidic linkage does occur. This leads to the nucleophilic attack at the anomeric carbon and the formation of a covalent glycosyl-enzyme intermediate, which has the inverted anomeric stereochemistry respect the substrate (Vocadlo et al. 2001). In the second step of the hydrolysis, the same residue, which acted previously as proton donor, now acts as base, and deprotonates a water molecule that finally attacks the anomeric carbon and breaks the covalent-bound glycosyl-enzyme intermediate.



Glycosidase inhibitors

Due to the major roles of glycosidases in biological processes, rationalization of the recognition and of the reaction mechanism has become fundamental in order to develop possible therapeutic agents against particularly related diseases. Various reviews can be found in literature, concerning different classes of inhibitors, which can be divided in two main groups: irreversible and reversible (el Ashry et al. 2000; el Ashry et al. 2000; el Ashry et al. 2000).

The first class of compounds, the irreversible inhibitors, are characterized by permanent blocking of the enzymatic activity. Ligand's structure should be very close to the natural one, but it should also include some modifications, which allows the formation of a covalent linkage with those residues at the active site of the protein. Into this group there are highly reactive molecules, like the N-bromoacetyl glycosylamine, which can react directly with aminoacid residues like methionine, histidine or cysteine (Naider et al. 1972); the family of the glycosyl-thiocyanates that bound covalently to the amine group of the aminoacid lateral chains (Shulman et al. 1976), or the glycosyl-methyl triazines, which are very reactive against nucleophiles in acidic media (Marshall et al. 1981). A second class of irreversible inhibitors includes those compounds that have low intrinsic reactivity and that can be later activated. Belonging to this class, it is possible to consider the epoxides (Macarron et al. 1993; Wacker et al. 1992) and the deoxyfluoroglycosides (Gebler et al. 1992; McCarter et al. 1997; Vocadlo et al. 2001). All these structures have in common a glycosidic moiety which, in the first step of the reaction, allows the recognition by the enzyme.

The mayor class of used inhibitors belong to the reversible one; in particular, we refer here to those molecules that compete with the substrate for the binding to the enzyme (competitive reversible inhibitors). In this context, two mayor classes can be defined: those molecules with structures mimicking the substrate transition state, as postulated by Pauling (Pauling 1946) and those substrate analogues based on the complementarities of the enzyme and substrate, as first described by Fischer in his "lock-and-key" [Fischer, 1894] postulate.

Molecules belonging to the first class should have a 3D-structure similar of that of the alkoxy-carbenium-like transition state and, in principle, the closest these structures are to the real transition state, the strongest the binding would expect (Lillelund et al. 2002). There are two major features of the transition state, which are distortion of the 4C_1 chair conformation and the presence of a positive charge along C1-O5 bond. Based on this principles several molecules have been described in literature like the aldono-1,5 lactone (Levy et al. 1972), the nojirimycin, deoxynojirimycin, swansonine and castanospermine (Asano et al. 2000), polyhydroxylated azepanes (Li et al. 2004; Li et al. 2004) or the tetrazol-like glycosidic (Heightam et al. 1999).

Inhibitors of the second class (analogues of the substrate) are in general the simplest molecules to design, due to their structural similarity with the substrate. The major differences respect their natural analogs are the introduction of modifications that prevent their hydrolysis by the target enzyme. Examples of this structures are the carboglycosides (el Ashry et al. 2000), the C-glycosides (Espinosa et al. 1998; Wang et al. 2000), the thioglycosides (Driguez 2001; Fort et al. 2001) or the N-linked analogs (Peri et al. 2004) which are also been studied in their interaction with glycosidase (Aguilera et al. 1998; Compain et al. 2001).

In the first following paper, novel glycosidase inhibitors, with azepane-based structures, are presented; while in the second paper, a mimic of the natural 1,2-mannobioside is illustrated. NMR and computational studies of free and bound structures are also discussed.

References

- Aguilera, B., J. Jimenez-Barbero and A. Fernandez-Mayoralas (1998) "Conformational differences between Fuc(α 1-3) GlcNAc and its thioglycoside analogue" *Carbohydr Res* **308**(1-2): 19-27.
- Asano, N. N., R.J., R. J. Molynesus and G. W. J. Fleet (2000) "Sugar-mimic glycosidase inhibitors: natural occurrence, biological activity and prospects for therapeutic application" *Tetrahedron Asymmetry* **11**: 1645-1680.
- Compain, P. and O. R. Martin (2001) "Carbohydrate mimetics-based glycosyltransferase inhibitors" *Bioorg Med Chem* **9**(12): 3077-92.
- Driguez, H. (2001) "Thiooligosaccharides as tools for structural biology" *ChemBiochem* **2**(5): 311-8.
- el Ashry, E. S., N. Rashed and A. H. Shobier (2000) "Glycosidase inhibitors and their chemotherapeutic value, Part 1" *Pharmazie* **55**(4): 251-62.
- el Ashry, E. S., N. Rashed and A. H. Shobier (2000) "Glycosidase inhibitors and their chemotherapeutic value, Part 2" *Pharmazie* **55**(5): 331-48.
- el Ashry, E. S., N. Rashed and A. H. Shobier (2000) "Glycosidase inhibitors and their chemotherapeutic value, Part 3" *Pharmazie* **55**(6): 403-15.
- Espinosa, J. F., E. Montero, A. Vian, J. L. Garcia, H. Dietrich, R. R. Schmidt, M. Martin-Lomas, A. Imberty, F. J. Canada and J. Jimenez-Barbero (1998) "Escherichia coli beta-galactosidase recognizes a high-energy conformation of C-lactose, a nonhydrolyzable substrate analogue. NMR and modeling studies of solecular complex" *J Am Chem Soc* **120**: 1309-1318.
- Fort, S., A. Varrot, M. Schulein, S. Cottaz, H. Driguez and G. J. Davies (2001) "Mixed-linkage cellooligosaccharides: a new class of glycoside hydrolase inhibitors" *ChemBiochem* **2**(5): 319-25.
- Gebler, J., N. R. Gilkes, M. Claeysens, D. B. Wilson, P. Beguin, W. W. Wakarchuk, D. G. Kilburn, R. C. Miller, Jr., R. A. Warren and S. G. Withers (1992) "Stereoselective hydrolysis catalyzed by related beta-1,4-glucanases and beta-1,4-xylanases" *J Biol Chem* **267**(18): 12559-61.
- Goss, P. E., M. A. Baker, J. P. Carver and J. W. Dennis (1995) "Inhibitors of carbohydrate processing: A new class of anticancer agents" *Clin Cancer Res* **1**(9): 935-44.
- Heightam, T. D. and A. T. Vasella (1999) "Recent insight into inhibition structure and mechanism of configuration-retaining glycosidases" *Angew Chem Int Ed Engl* **38**: 750-770.
- Kolter, T., T. Doering, G. Wilkening, N. Werth and K. Sandhoff (1999) "Recent advances in the biochemistry of glycosphingolipid metabolism" *Biochem Soc Trans* **27**(4): 409-15.
- Koshland, D. E., Jr. and E. Clarke (1953) "Mechanism of hydrolysis of adenosinetriphosphate catalyzed by lobster muscle" *J Biol Chem* **205**(2): 917-24.
- Levy, G. A. and S. M. Snaith (1972) "The inhibition of glycosidases by aldololactones" *Adv Enzymol Relat Areas Mol Biol* **36**: 151-81.
- Li, H., Y. Bleriot, C. Chantereau, J. M. Mallet, M. Sollogoub, Y. Zhang, E. Rodriguez-Garcia, P. Vogel, J. Jimenez-Barbero and P. Sinay (2004) "The first synthesis of substituted azepanes mimicking monosaccharides: a new class of potent glycosidase inhibitors" *Org Biomol Chem* **2**(10): 1492-9.
- Li, H., Y. Bleriot, J. M. Mallet, Y. Zhang, E. Rodriguez-Garcia, P. Vogel, S. Mari, J. Jimenez-Barbero and P. Sinay (2004) "Polyhydroxyazepanes Mimicking Monosaccharides :Synthesis Of An a-D-Galacto-Like Iminoheptitol" *Heterocycles* **64**: 65-74.
- Lillelund, V. H., H. H. Jensen, X. Liang and M. Bols (2002) "Recent developments of transition-state analogue glycosidase inhibitors of non-natural product origin" *Chem Rev* **102**(2): 515-53.
- Macarron, R., C. Acebal, M. P. Castillon, J. M. Dominguez, I. de la Mata, G. Pettersson, P. Tomme and M. Claeysens (1993) "Mode of action of endoglucanase III from *Trichoderma reesei*" *Biochem J* **289** (Pt 3): 867-73.

- Marshall, P. J., M. Sinnott, P. J. Smith and D. Widdows (1981) "Active-site-directed irreversible inhibition of glycosidases by the corresponding glycosylmethyl-(p-nitrophenyl)triazenes" *J.C.S. Perkin I*(2): 366-376.
- McCarter, J. D., D. L. Burgoyne, S. Miao, S. Zhang, J. W. Callahan and S. G. Withers (1997) "Identification of Glu-268 as the catalytic nucleophile of human lysosomal beta-galactosidase precursor by mass spectrometry" *J Biol Chem* **272**(1): 396-400.
- McCarter, J. D. and S. G. Withers (1994) "Mechanisms of enzymatic glycoside hydrolysis" *Curr Opin Struct Biol* **4**(6): 885-92.
- Naider, F., Z. Bohak and J. Yariv (1972) "Reversible alkylation of a methionyl residue near the active site of -galactosidase" *Biochemistry* **11**(17): 3202-8.
- Neufeld, E. F. (1991) "Lysosomal storage diseases" *Annu Rev Biochem* **60**: 257-80.
- Pauling, L. (1946) *Chem. Eng. News* **24**: 1345.
- Peri, F., J. Jimenez-Barbero, V. Garcia-Aparicio, I. Tvaroska and F. Nicotra (2004) "Synthesis and conformational analysis of novel N(OCH₃)-linked disaccharide analogues" *Chemistry* **10**(6): 1433-44.
- Sebastio, G., M. Villa, R. Sartorio, V. Guzzetta, V. Poggi, S. Auricchio, W. Boll, N. Mantei and G. Semenza (1989) "Control of lactase in human adult-type hypolactasia and in weaning rabbits and rats" *Am J Hum Genet* **45**(4): 489-97.
- Shulman, M. L., S. D. Shiyan and A. Y. Khorlin (1976) "Specific irreversible inhibition of sweet-almond beta-glucosidase by some beta-glycopyranosylepoxyalkanes and beta-D-glucopyranosyl isothiocyanate" *Biochim Biophys Acta* **445**: 169-181.
- Vocadlo, D. J., G. J. Davies, R. Laine and S. G. Withers (2001) "Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate" *Nature* **412**(6849): 835-8.
- Wacker, H., P. Keller, R. Falchetto, G. Legler and G. Semenza (1992) "Location of the two catalytic sites in intestinal lactase-phlorizin hydrolase. Comparison with sucrase-isomaltase and with other glycosidases, the membrane anchor of lactase-phlorizin hydrolase" *J Biol Chem* **267**(26): 18744-52.
- Wang, Q., M. Wolff, T. Polat, Y. Du and R. J. Linhardt (2000) "Inhibition of neuraminidase with neuraminic acid C-glycosides" *Bioorg Med Chem Lett* **10**: 941-944.
- White, A. and D. R. Rose (1997) "Mechanism of catalysis by retaining beta-glycosyl hydrolases" *Curr Opin Struct Biol* **7**(5): 645-51.
- Yarema, K. J. and C. R. Bertozzi (1998) "Chemical approaches to glycobiology and emerging carbohydrate-based therapeutic agents" *Curr Opin Chem Biol* **2**(1): 49-61.
- Zechel, D. L. and S. G. Withers (2000) "Glycosidase mechanisms: anatomy of a finely tuned catalyst" *Acc Chem Res* **33**(1): 11-8.

The Conformational Behavior of Novel Glycosidase Inhibitors with Substituted Azepan Structures: An NMR and Modeling Study

Karina Martínez-Mayorga,^[a] José L. Medina-Franco,^[a] Silvia Mari,^[a] F. Javier Cañada,^[a] Eliazar Rodríguez-García,^[c] Pierre Vogel,^[c] Hongqing Li,^[b] Yves Blériot,^{*[b]} Pierre Sinaÿ,^{*[b]} and Jesús Jiménez-Barbero^{*[a]}

Dedicated to Prof. Dr. Pedro Molina on the occasion of his 60th birthday

Keywords: Conformational analysis / Nuclear magnetic resonance / Polyhydroxyazepanes / Molecular modeling / Glycosidase inhibitors / Glycomimetics

The conformational analysis of a series of configurational isomers of 2-(hydroxymethyl)azepan-3,4,5,6-tetrols **1–4** has been carried out. ¹H NMR spectroscopic data, especially vicinal *J* couplings and nuclear Overhauser enhancements (NOE), assisted by molecular mechanics, molecular dynamics and Monte Carlo calculations, have been used. A fairly good agreement between experimental and calculated data has been found. The different isomers exist in a conformational equilibrium between two chair-like structures.

TR-NOE experiments have also allowed us to demonstrate that the bound conformation of compound **2** to the β -glucosidase from almonds is the major one of this compound present in solution. Finally, molecular docking of the different conformations of these compounds in the binding site of three different enzymes has been performed in order to try to rationalize the observed inhibition of these molecules.

(© Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, Germany, 2004)

Introduction

The quest for glycosidase inhibitors has been the subject of extensive research in the past few years due to their potential use as therapeutic agents.^[1] Indeed, several molecules with this capacity have been shown to interact with receptors related to diabetes,^[2] Gaucher's disease,^[3] HIV infection,^[4] viral infections,^[5] and even cancer.^[6] Moreover, they have also been used as chemical probes, in combination with crystallography, modeling and other biochemical and biophysical methods, to provide new insights into the glycosidase mechanism^[7] and they are now expected to find an increasing number of applications as beneficial drugs.^[8]

We have recently reported the synthesis and biological evaluation of a variety of azepan derivatives with four hydroxy groups attached to the ring,^[9] as well as a hydroxymethyl moiety (Figure 1), which mimic monosaccharides

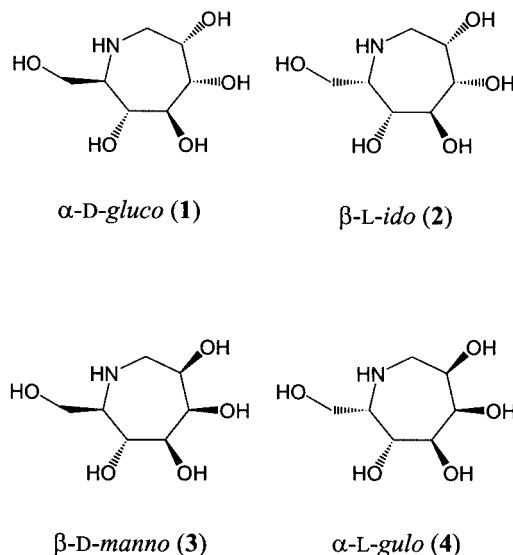


Figure 1. Schematic view of compounds **1–4**

with different stereochemistries, that is, α -D-*gluco* (**1**), β -L-*ido* (**2**), β -D-*manno* (**3**), and α -L-*gulo* (**4**). These compounds were shown to behave as moderate to good glycosidase inhibitors with **1** and **3** inhibiting bovine liver β -galactosidase

^[a] Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain
E-mail: jjbarbero@cib.csic.es

^[b] Ecole Normale Supérieure, Département de Chimie, UMR 8642,
24 rue Lhomond, 75231, Paris Cedex 05, France

^[c] Institute of Glycochemistry and Asymmetric Synthesis, Swiss Federal Institute of Technology (EPFL), BCH,
1015 Lausanne, Switzerland

Supporting information for this article is available on the WWW under <http://www.eurjoc.org> or from the author.

and coffee bean α -galactosidase in the low micromolar range, respectively, despite showing *gluco*- and *manno*-like configurations.^[9]

On this basis, and in order to try to clarify the structural basis for this behavior, we now report on the conformational study of these seven-membered ring compounds by using a NMR/modeling approach. These seven-membered ring systems, such as **1–4**, are inherently flexible and may assume several conformations that can interconvert with relative low energy barriers.

Proton magnetic resonance (¹H NMR) spectra have been recorded at neutral and acidic pH, the corresponding coupling constants have been deduced, and nuclear Overhauser effect (NOE) experiments have been evaluated for **1–3**. These data along with the results from molecular mechanics, molecular dynamics, and Monte Carlo calculations^[10] have permitted us to evaluate the conformational behavior of these iminoalditols. In a further step, STD^[11] (Saturation Transfer Difference) and TRNOE^[12] experiments have been carried out to deduce the conformation of **3** bound to β -glucosidase from almonds. Finally, molecular docking studies^[13] of these molecules have been carried out in order to characterize their binding mode to three different glycosidase enzymes^[14] and to try to rationalize the ability of **1–4** to behave as *gluco*- and *galactosidase* inhibitors.

Methods

Materials: The synthesis of these molecules has been published elsewhere.^[9]

NMR: Proton assignments were performed using standard 1D, 2D-COSY, NOESY, and HSQC experiments.^[15] The coupling constants for **1–3** were obtained from ¹H NMR spectra recorded in acidic media; additional experiments were carried out under neutral conditions for β -L-*ido* (**2**) and α -D-*gluco* (**1**). Proton–proton interatomic distances were estimated from the enhancements measured by selective 1D NOE experiments, with the DPFGE NOE sequence proposed by Shaka and co-workers.^[16]

Molecular Modeling: Molecular mechanics (MM), molecular dynamics (MD) and Monte Carlo (MC) studies^[17] were conducted with the MACROMODEL program, version 5.5.^[18] Both the AMBER*^[19] and MM3*^[20] force fields were used. The energies were minimized by using the PR conjugate gradient method. A bulk dielectric constant of 80 was used when calculations were performed "in vacuo". The GB/SA (Generalized Born Surface Area) solvation model^[21] was also used. The starting coordinates for dynamics calculations were those obtained after energy minimizations. Simulations were carried out over 2 ns at 300 K. Monte Carlo studies were conducted by using default parameters implemented in MACROMODEL; 300 trial structures were generated for each molecule. Coupling constants were calculated by using the empirical Karplus equation proposed by Haasnoot et al.^[22] Interatomic H–H distances and estimated NOEs were calculated by using the NOEPROM program, which is available from the authors upon request.^[23]

Docking simulations were conducted with AutoDock 3.0.^[24] X-ray coordinates for the enzymes were taken from the Protein Data Bank or other work.^[25–27] This program performs automated docking of the whole ligand with user-specified dihedral flexibility within a rigid protein binding site. Before docking, ligands as well as all water molecules were removed except for the putative catalytic water molecule of the glucoamylase. Docking studies with β -glucosidase were conducted with the B chain. Polar hydrogen atoms were added by using the auxiliary program Auto Dock Tools. Energies were evaluated from precalculated grids with molecular affinity potentials. Affinity grid files were generated using the auxiliary program Autogrid. The centers of the coordinates of the crystal ligands were taken as the centers of the grids, and the dimensions of the grids were $23 \times 23 \times 23$ Å with points separated by 0.375 Å.

The possible conformations of the analogues (see below) were protonated at the nitrogen atom and minimized with the AMBER* force field and the PR conjugate gradient method. The dielectric constant was set to 80. Gasteiger charges were computed with the auxiliary program Auto Dock Tools and nonpolar hydrogens merged. AutoDock randomized the initial position.

With α -galactosidase and glucoamylase, the configurational and translational exploration was conducted with a Monte Carlo simulated annealing technique. With β -glucosidase, the Lamarckian Genetic Algorithm implemented in the program was used. For each starting structure a total of 100 independent Monte Carlo or genetic algorithm simulations were made. For each Monte Carlo simulation there were 50 constant temperature cycles with a maximum of 3000 steps accepted or rejected. The initial temperature was $RT = 1000 \text{ cal}\cdot\text{mol}^{-1}$ and was reduced by a factor of 0.95 each cycle. All the other values for the Monte Carlo simulations as well as Lamarckian genetic algorithm search parameters were taken from the default values of AutoDock.

After docking, the 100 solutions were clustered into groups with RMS deviations less than 1.0 Å. The clusters were ranked on the basis of the lowest energy representative of each cluster. Note that the AutoDock energies as reported here may not represent the true energies, rather they are just a measure of the scoring function used.

Results and Discussion

The conformational analysis of the different molecules was performed by using a combination of experimental NMR spectroscopic data, assisted by modeling methods. The analysis of the shapes of seven-membered rings is challenging and a variety of forms may occur. The substitution of different carbon atoms by hydroxy groups, with different stereochemistries, and, moreover, by a bulky hydroxymethyl moiety must obviously modify the set of conformations that are accessible. The conformation of the seven-membered ring as well as of that of the hydroxymethyl group were explored.^[28] In all cases, the NMR spectrum obtained under acidic conditions was of higher quality than the one

recorded under neutral conditions and allowed a more precise measure of the key J and NOE parameters.

α -D-*gluco*-Like Compound 1

The experimental vicinal proton–proton coupling constants and NOE enhancements (from which interatomic H–H distances may be estimated) are shown in Table 1 and 2, respectively. Molecular mechanics, dynamics and Monte Carlo calculations predicted the existence of two major conformations, dubbed **A1** and **B1** (Figure 2), although with some flexibility around the conformers depicted in the figures. Conformer **B1** is about $5.6 \text{ kJ}\cdot\text{mol}^{-1}$ more stable than **A1** according to the AMBER* force field and the GB/SA solvent model. Similar results in terms of energies and geometries were obtained when either a bulk dielectric constant ($\epsilon = 80$) or when the MM3* force field was employed. The bulk hydroxymethyl group (C7) is attached to C6 in the more stable pseudoequatorial orientation in both conformers. For simplicity, and as sugar mimics, the chosen numbering is related to the sugar nomenclature, with the deoxy position of 1–4 being C-1. The hydroxy groups at positions 3, 4, and 5 of conformer **B1** are also equatorially oriented, while OH-2 has a pseudoaxial disposition. In contrast, the OH-3 of conformer **A1** is the only hydroxy group that assumes a relative pseudoaxial orientation. Table 1 also gives the AMBER*-computed proton–proton dihedral angles for **A1** and **B1**, along with the estimated couplings. Table 2 shows the AMBER*-MM/MD predicted average interatomic distances, along with the expected NOEs. It may

be observed that conformers **A1** and **B1** may be differentiated on the basis of the expected couplings for $J_{1,2}$ and $J_{3,4}$. Indeed, the experimental data seem to indicate that **B1** is predominant (Table 1) although with some contribution from **A1**-type conformers. The experimental couplings for the vicinal ring proton pairs can be explained using the same rationale. According to the data, similar conformational equilibria are seen at neutral and acidic pH.

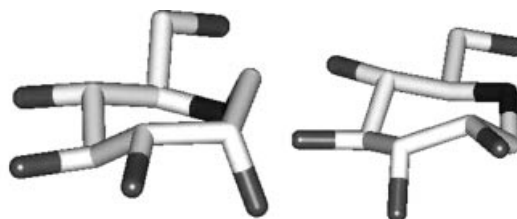


Figure 2. Conformations **A1** (right) and **B1** (left) of α -D-*gluco*-like analogue **1** (AMBER*; $\epsilon = 80$)

Similar conclusions are reached from the interatomic distances. In this case, key differences between the two calculated conformations are found for the H1'–H3 and H2–H5 proton pairs. Again, the experimental NOEs suggest a major contribution from **B1**-type conformers, since the H1'–H3 cross-peak displays an appreciable intensity and the H2–H5 cross-peak intensity is just above the noise level, in contrast to what is expected for the **A1**-type geometry. The average distances and NOE values obtained from the MD calculations are in all cases in between those values predicted for the single **A1** and **B1** conformers and are also in agreement with the observed data. Thus, the MD trajectory seems to provide a reasonable estimate of the conformational distribution (Table 2). Indeed, several reversible transitions between the **A1** and **B1** forms (Supporting Information, see also the footnote on the first page of this article) take place during the simulation, which also indicates that the energy barrier for interconversion should be relatively low.

Table 1. Experimental and expected ${}^3J_{\text{H,H}}$ and estimated torsion angles (τ) for the α -D-*gluco*-like derivative **1**

| ${}^3J_{\text{H,H}}$ | Experimental | | Expected (MM/MD/MC) | | | |
|----------------------|---------------------------|-----------|---------------------|----------|---------------------|----------|
| | ${}^3J_{\text{H,H}}$ (Hz) | | Conformer B1 | | Conformer A1 | |
| | Neutral pH | Acidic pH | τ ($^\circ$) | J (Hz) | τ ($^\circ$) | J (Hz) |
| $J_{1,2}$ | < 1.5 | < 1.5 | 69.5 | 1.0 | 86.3 | 1.6 |
| $J_{1,2}$ | 6.0 | 6.0 | 48.8 | 5.0 | 154.3 | 9.7 |
| $J_{2,3}$ | 4.4 | 4.4 | 52.6 | 3.1 | 41.8 | 4.5 |
| $J_{3,4}$ | 8.0 ± 0.5 | 9.2 | 175.4 | 10.0 | 153.3 | 6.4 |
| $J_{4,5}$ | 9.1 | 9.1 | 160.4 | 9.7 | 160.9 | 9.7 |
| $J_{5,6}$ | 8.7 | 8.7 | 160.1 | 8.6 | 171.1 | 10.3 |

Table 2. Experimental and calculated NOEs and H–H distances for α -D-*gluco*-like derivative **1**

| Proton pair | Experimental NOE | | Molecular modeling | | | |
|-------------|--------------------------|-------------------------------------|---------------------|---------------------|---------------------------------|------------------------------|
| | Intensity ^[a] | Upper limit distance ^[b] | MM (r) | | MD | |
| | | | Conformer A1 | Conformer B1 | $\langle r^{-6} \rangle^{-1/6}$ | NOE intensity ^[c] |
| 1'–2 | ms | 2.8 | 2.6 | 2.5 | 2.6 | 6.6 |
| 1–2 | m | 2.9 | 3.0 | 2.4 | 2.8 | 3.7 |
| 1'–3 | mw | 3.1 | 4.2 | 2.6 | 3.2 | 1.3 |
| 1'–5 | w | 3.5 | 4.2 | 2.4 | 3.2 | 1.5 |
| 2–3 | s | 2.6 | 2.3 | 2.4 | 2.4 | 9.7 |
| 2–5 | vw | 4.0 | 2.4 | 3.9 | 2.7 | 4.2 |
| 3–4 | m | 2.9 | 3.0 | 3.1 | 3.1 | 2.1 |
| 3–5 | ms | 2.8 | 2.8 | 2.3 | 2.7 | 4.8 |
| 4–5 | m | 2.9 | 3.0 | 3.0 | 3.1 | 2.0 |
| 4–6 | s | 2.6 | 2.3 | 2.7 | 2.5 | 8.3 |
| 5–6 | mw | 3.1 | 3.0 | 3.0 | 3.1 | 1.9 |

^[a] Average of those obtained at different mixing times. s: strong; ms: medium strong; m: medium; mw: medium weak; w: weak; vw: very weak. ^[b] Upper limits of distances (Å) are provided, according to the observed intensities, as 2.6, 2.8, 2.9, 3.1, 3.3, 3.5 Å. ^[c] NOE intensities calculated with NOEPROM (%).

Regarding the rotamer distribution around the C6–C7 linkage of the hydroxymethyl group, Table 3 gives the expected coupling values for the three possible staggered rotamers and also the experimental data for comparison. The conformers were dubbed *gg*, *gt*, and *tg*, following the currently used nomenclature for sugars. The GB/SA AMBER* calculations predicted small energy differences between the three rotamers (less than $3 \text{ kJ}\cdot\text{mol}^{-1}$), *gt* being the most populated one. However, the observed results, with one small and one medium coupling constant value, indicate that a rotational equilibrium between the *gg* and the *gt* rotamers occurs, with a similar population of both forms. Similar behavior has been described for Glc/Man pyranose rings, since in these cases, the equatorially oriented O4 in the six-membered chair of Glc/Man precludes the existence of the *tg* rotamer displaying destabilizing O1–O3 interactions. The results obtained at neutral and acidic pH are very similar, which indicates the same type of equilibrium exist under both experimental conditions. Thus, the state of protonation of the nitrogen atom does not seem to influence the conformational distribution at the C6–C7 torsion, as also observed for the seven-membered ring itself (see also Table 1).

Table 3. Experimental and expected coupling constants (Hz) [from the proton–proton torsion angles, in brackets ($^{\circ}$)] for the three staggered rotamers of the hydroxymethyl group of α -D-*gluco*-like compound **1**; the geometry of the **B1** conformer was employed

| | Exp. NMR | | | |
|------------------------------------|----------|------------|--------------|------------|
| | | <i>gt</i> | <i>tg</i> | <i>gg</i> |
| ΔE (kJ/mol) ^[a] | | 0.0 | 1.8 | 2.8 |
| $^3J_{\text{H-H}}$ (τ) | 6-7 | 3.7 (60.5) | 10.9 (177.4) | 2.5 (59.1) |
| | 6-7 | 6.5 | 4.1 (62.8) | 1.6 (59.6) |

^[a] Calculated with AMBER*; GB/SA.

β -L-*ido*-Like Compound **2**

Compound **2** differs from **1** by the stereochemistry at position C6, to which the C7 hydroxymethyl group is attached. The experimental and calculated coupling constants are given in Table 4, while the corresponding NOE-based information is gathered in Table 5. Again, only the results obtained by using AMBER* and GB/SA are shown, since those with MM3* and a bulk dielectric constant of 80 are very similar.

Table 4. Experimental and expected $^3J_{\text{H,H}}$ and estimated torsion angles (τ) for the β -L-*ido*-like derivative **2**

| $^3J_{\text{H,H}}$ | Experimental | | Expected (MM/MD/MC) | | | |
|--------------------|-------------------------|-----------|------------------------|----------|------------------------|----------|
| | $^3J_{\text{H,H}}$ (Hz) | | Conformation A2 | | Conformation C2 | |
| | Neutral pH | Acidic pH | τ ($^{\circ}$) | J (Hz) | τ ($^{\circ}$) | J (Hz) |
| $J_{1,2}$ | < 1.5 | 2.5 | 83.6 | 1.8 | 65.0 | 1.2 |
| $J_{1,2}$ | 6.0 | 6.5 | 160.3 | 10.3 | 53.1 | 4.4 |
| $J_{2,3}$ | < 1.5 | 1.5 | 42.4 | 4.3 | 78.0 | 1.0 |
| $J_{3,4}$ | 7.0 | 6.3 | 147.3 | 5.4 | 161.6 | 7.7 |
| $J_{4,5}$ | 3.0 | 2.8 | 165.4 | 10.0 | 91.0 | 0.7 |
| $J_{5,6}$ | < 1.5 | < 1.0 | 63.5 | 3.2 | 80.0 | 0.4 |

The Monte Carlo search followed by minimization produced two stable conformations of the β -L-*ido*-like compound **2**. One of them, named **A2** had a very similar geometry to that described above as **A1** for the α -D-*gluco*-like derivative **1**. The second one was not exactly like **B1**, but somehow distorted, probably due to the different stereochemistry at C6, and was dubbed **C2** (Figure 3). This was the global minimum, stabilized with respect to **A2** by $12.5 \text{ kJ}\cdot\text{mol}^{-1}$. The conformational equilibrium was evaluated as above. From the coupling values, key differences between the conformers are expected for $J_{1,2}$, $J_{2,3}$ and $J_{4,5}$ (Table 4). It is evident that the observed values are between those expected for **A2** and **C2**. From the NOE data, the major proton–proton interatomic distance differences occur for the H1'–H3, H1'–H6 and H3–H6 proton pairs (Table 5). Therefore, a conformational equilibrium between **A2** and **C2** is evident, **C2** being predominant. By comparing the major conformers of α -D-*gluco* (**1**) and β -L-*ido* (**2**), **B1** and **C2**, respectively, we can assume that the difference is due to the necessity of the hydroxymethyl group to adopt a pseudoequatorial orientation, which cannot be accommodated by the **B1**-type geometry. Distortion of the ring **B** seen in **C** can be attributed to the different conformation at C6. Despite the relatively high predicted energy difference, which would preclude the existence of **A2**-type conformers, their presence is acknowledged due to the experimental observations. Moreover, MD simulations starting from either **A2** or **C2** gave rather stable trajectories with no interconversions, somehow proving the conformational stability of these geometries.

The conformation of the lateral C6–C7 chain was also analyzed. The highest relative energy of the three possible rotamers is ca. $4 \text{ kJ}\cdot\text{mol}^{-1}$ (from *gg* to *gt*, see Table 6). The coupling constants for the hydroxymethyl group of the three main rotamers are shown in Table 6. Following the reasoning used above for compound **1**, the two intermediate values obtained at neutral pH are in agreement with an almost 50:50 conformational distribution of the *gt* and *tg* rotamers, as expected for *L-ido*-type sugars, with essentially no contribution from the *gg* form. However, in this case, lowering the pH towards acidic conditions (pH = 4.5) influences the conformational distribution of the lateral chain. Indeed, the increase in one of the $J_{6,7}$ couplings (from 7 to 9 Hz) is in agreement with a change in the conformational distribution to a 75:25 ratio in favor of the *gt* rotamer.

β -D-*manno*-Like Compound **3**

The experimental (at acidic pH) and calculated coupling constants of β -D-*manno*-like compound **3** are compared in Table 7. The difference between compounds **1** and **3** lies in the stereochemical differences at positions 2 and 3 (sugar-related nomenclature). In this case, again two conformers **A3** and **B3** were found by the MC protocol with very similar geometries to those found for **1**. The main difference lies in the relative disposition of the substituents at C2 and C3 and in this case these groups are orientated differently. Thus, the OH-3 and OH-2 groups of conformer **B3** are axially and equatorially oriented, respectively, while in con-

Table 5. Experimental and calculated NOEs and H–H interatomic distances for the β -L-ido-like derivative **2**

| Proton pair | Experimental NOE | | Molecular modeling | | MD <- -6>-1/6 |
|-------------|--------------------------|----------------------|--------------------|------------------------|------------------|
| | Intensity ^[a] | Upper limit distance | Conformer A2 | MM (r) Conformer C2 | |
| 1'-2 | ms | 2.8 | 2.6 | 2.5 | 2.5 |
| 1-2 | ms | 2.8 | 3.1 | 2.4 | 2.6 |
| 1'-3 | m | 2.9 | 4.2 | 2.4 | 3.1 |
| 1'-4 | vw | 4.0 | 4.1 | 4.5 | 4.4 |
| 1'-6 | mw | 3.1 | 4.1 | 2.4 | 3.0 |
| 2-3 | s | 2.6 | 2.3 | 2.6 | 2.5 |
| 2-4 | vw | 4.0 | 3.7 | 3.8 | 3.8 |
| 2-6 | vw | 4.0 | 3.8 | 4.0 | 3.9 |
| 3-4 | mw | 3.1 | 3.0 | 3.0 | 3.0 |
| 3-6 | ms | 2.8 | 4.9 | 2.3 | 3.1 |
| 4-5 | ms | 2.8 | 3.1 | 2.6 | 2.8 |
| 5-6 | ms | 2.8 | 2.5 | 2.6 | 2.6 |

^[a] s: strong; ms: medium strong; m: medium; mw: medium weak; w: weak; vw: very weak.



Figure 3. Conformations A2 (left) and C2 (right) of β -L-ido-like analogue **2** (AMBER*; GB/SA)

Table 6. Experimental and expected coupling constants (Hz) [from the proton–proton torsion angles ($^\circ$), in brackets] for the three staggered rotamers of the hydroxymethyl group of the β -L-ido-like compound **2**; the geometry of the major C2 conformer was employed

| ΔE (kJ/mol) ^[a] | $^3J_{\text{H,H}}$ (τ) | Experimental (neutral / acidic pH) | | |
|------------------------------------|-------------------------------|------------------------------------|------------|--------------|
| | | gg | gt' | tg |
| | 6-7 | 7.0 / 5.5 | 2.4 (59.2) | 3.7 (65.9) |
| | 6-7' | 7.0 / 9.0 | 1.6 (59.8) | 10.7 (174.5) |
| | | | 4.2 | 2.7 |

^[a] Calculated with AMBER*; GB/SA.

Table 7. Experimental and expected $^3J_{\text{H,H}}$ and estimated torsion angles (τ) for the β -D-manno-like derivative **3**

| $^3J_{\text{H,H}}$ | Experimental $^3J_{\text{H,H}}$ (Hz) | Expected (MM/MD/MC) | | | |
|--------------------|---|---------------------|----------|---------------------|----------|
| | | Conformer A3 | | Conformer B3 | |
| | | τ ($^\circ$) | J (Hz) | τ ($^\circ$) | J (Hz) |
| $J_{1',2}$ | 7.0 | 38.5 | 6.4 | 154.5 | 8.9 |
| $J_{1,2}$ | 5.5 | 79.9 | 0.9 | 35.8 | 7.9 |
| $J_{2,3}$ | < 1.5 | 32.2 | 5.5 | 81.7 | 0.8 |
| $J_{3,4}$ | < 1.5 | 41.7 | 4.6 | 66.5 | 1.6 |
| $J_{4,5}$ | 8.5 | 158.2 | 9.8 | 172.2 | 8.9 |
| $J_{5,6}$ | 9.0 | 173.2 | 10.3 | 166.4 | 10.1 |
| $J_{6,7a}$ | 2.8 | 56.7 | 4.2 | 54.3 | 2.1 |
| $J_{6,7b}$ | 7.0 | 11.1 | 11.1 | 65.1 | 1.9 |

former A3 the OH-3 and OH-2 groups are equatorially and axially oriented, respectively, with the hydroxymethyl group always adopting a pseudoequatorial orientation. The exper-

imental data indicate the existence of a conformational equilibrium between A3 and B3 with the B form prevalent, as observed for **1** (Figure 4). Actually, the application of the MM/MD/MC protocols to β -D-manno (**3**) produced results remarkably similar to those for α -D-gluco (**1**).

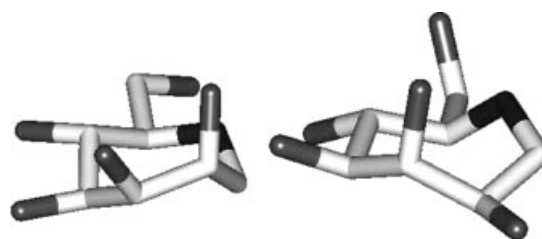


Figure 4. Conformations A3 (left) and B3 (right) for the β -D-manno-like analogue **3** (AMBER*; GB/SA)

The couplings for the hydroxymethyl group of **3** are shown in Table 7, while the reasoning can be followed with the Newman projections and the expected values for gg , gt , and tg rotamers given in Table 7. The observed values (2.8, 7.0 Hz) are in agreement with a gg – gt equilibrium similar to those predicted for Glc/Man pyranose analogues (see above for **1**), but with a higher predominance of gt rotamers than was the case with **1**.

α -L-gulo-Like Compound **4**

The peaks in the ^1H NMR spectrum of **4** showed extensive overlapping at any pH, and it was not possible to unambiguously deduce the relevant J couplings and/or NOEs to account for the conformational distribution. However, the

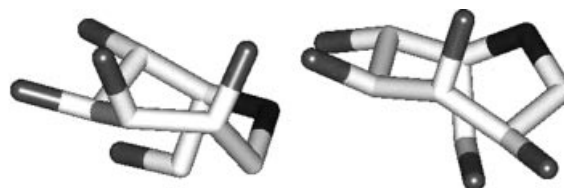


Figure 5. Conformations A4 (right) and B4 (left) for the α -L-gulo-like analogue **4** (AMBER*; GB/SA)

application of the modeling protocol also led to a participation of two chair-like forms, dubbed **A4** and **B4** (Figure 5), similar to those reported for **2**, with the **B4** conformer being the preferred one from the molecular mechanics calculations. Indeed two axial substituents may be observed for **A4**, namely the hydroxymethyl group and OH-2, while only OH-3 adopts an axial disposition in conformer **B4**.

Therefore, the conformational behavior of the four compounds can be described as a conformational equilibrium between two calculated chair-like conformations, with one being predominant. The hydroxymethyl substituent group can also adopt two conformations, with one predominating in some cases.

The Bound State

In a further step, we investigated the mode of binding of these molecules to glycosidase enzymes. NMR experiments

have been shown to be useful for deducing the conformation of carbohydrate analogues bound to lectins, and in some cases, to enzymes. Thus, we performed TRNOE and STD experiments in order to study the complexes of **3** with bovine liver β -galactosidase and with the β -glucosidase from almonds. Unfortunately, it was not possible to obtain any NMR spectroscopic data for **3** bound to the former enzyme. Very probably, given the good inhibition ability of **3**, the kinetic features of the molecular recognition phenomenon are not suitable for the strict requirements of STD and TRNOE experiments. Fortunately, both STD and TRNOE experiments (Figure 6) permitted us to deduce some information as regards the binding of **3** to the β -glucosidase from almonds. The STD experiment shows that the major transfer of magnetization from the enzymic protons to the ligand involves H2, H3, H4, and to some extent H5 atoms, that is, one of the sides of the molecule.

Figure 6 allows the NOESY spectrum (mixing time 300 ms) of free **3** to be compared with the TRNOESY spectrum

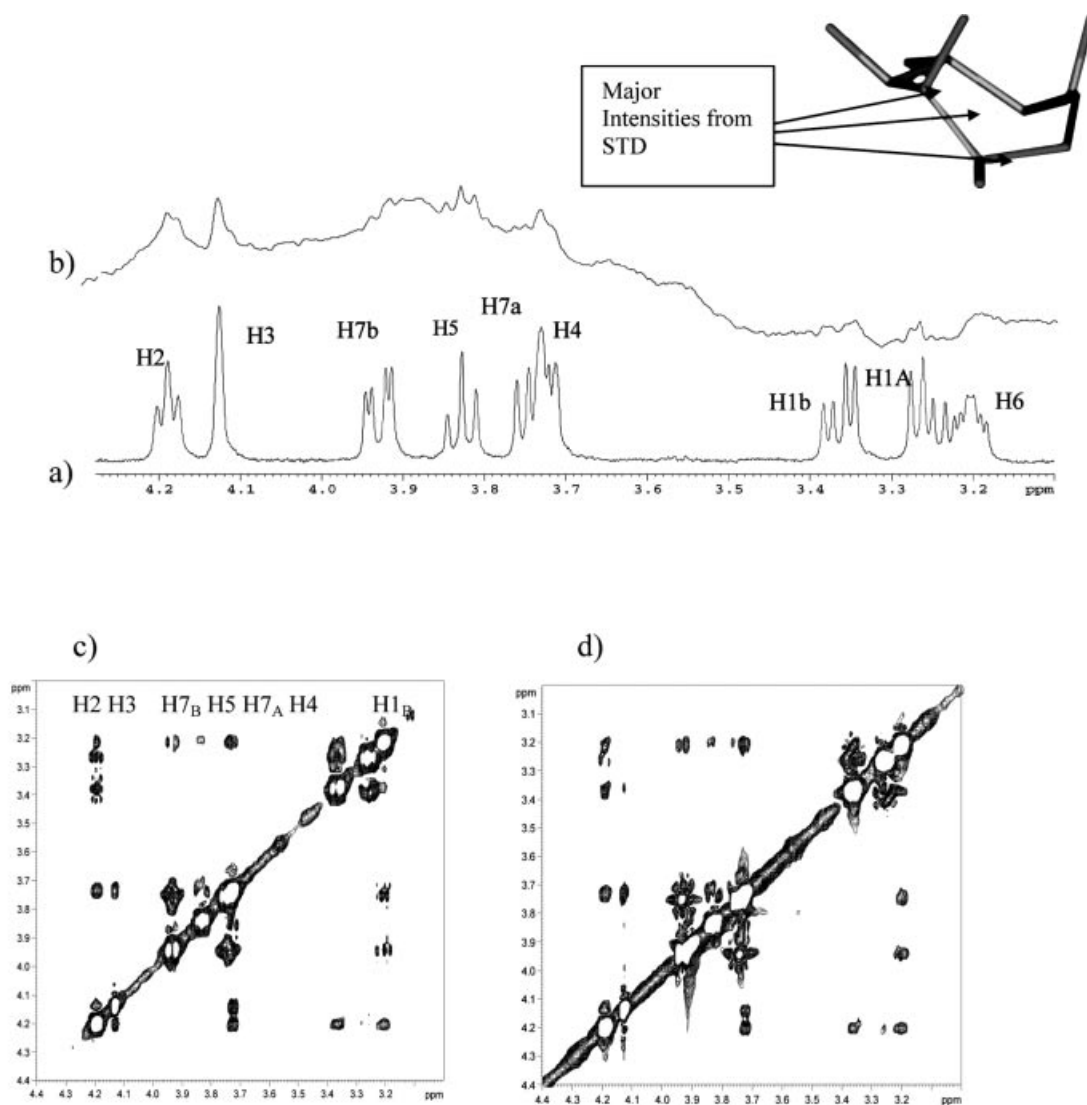


Figure 6. (a) ^1H NMR spectrum of the free ligand (1.5 mM) (50 mM in phosphate buffer, pH 5.8, and 25 °C); (b) ^1H NMR STD spectrum of the ligand (1.5 mM) bound to almond glucosidase (50 μM) (50 mM in phosphate buffer, pH 5.9, and 25 °C); NOE data for **3** in the free and enzyme-bound states: (c) NOESY (300 ms) of free **3**; (d) TRNOESY (150 ms) of **3** (1.5 mM) with almond glucosidase (50 μM)

of **3** (mixing time 150 ms) in the presence of the glucosidase (30:1 molar ratio). It is observed that despite some anti-phase character in the cross peaks of the spectrum of the free molecule, as expected for a small molecule and a relatively short NOE mixing time, the set and pattern of the observed cross peaks are basically identical. This seems to indicate that very probably, the β -glucosidase from almonds recognizes **3** in its major conformation present in solution, that is **3B**.

Inhibition Experiments

The inhibition ability of these molecules towards several glycosidases has been demonstrated previously.^[9] As shown below, docking experiments enabled us to predict the relative inhibitory ability of these compounds. In order to test whether this protocol (NMR, molecular mechanics + docking) could be extended to other glycosidases, the inhibition abilities of **1–4** towards two additional glycosidases were evaluated and the percentages of inhibition at a concentration of 1 mM of the inhibitor were determined at optimal pH, and at 35 °C. For G1 glucoamylase,^[27] the percentages of inhibition observed for compounds **1**, **3**, and **4** were 23%, 39%, and 52%, respectively, while compound **2** showed no inhibition. Curiously, the fagomine homologue, with only three substituents,^[9] displayed a 61% inhibition at a 1 mM concentration. For the A385T mutant of β -glucosidase from *Paenibacillus polymyxa*,^[26] only the β -D-manno-like analogue (**3**) was able to inhibit this enzyme (72% inhibition at a 1 mM concentration, which corresponds to an IC_{50} of 350 μ M). As will be shown in the following section, the predictions of the docking experiments perfectly match the observed inhibitions. We wish to emphasize the fact that the docking experiments were carried out before performing the inhibition assays.

Molecular Docking of Compounds 1–4 Towards Glycosidases

With all this information available, we decided to explore the atomic features of the interaction process. Therefore, molecular docking experiments of the two major conformations of each of the four compounds **1–4** in the binding site of three different glycosidases were performed to find the most favored conformation of these analogues in their binding state and to try to rationalize the observed inhibitory capacity of these molecules. They are indeed able to interact with different enzymes, even when they show different relative stereochemistries to those of the compounds targeted by the enzymes.

Thus, the two main low energy conformers of the four monosaccharide analogues **1–4** in solution were docked in the binding site of α -galactosidase from rice^[25] (family GH27, with a β/a_8 fold, and two Asp catalytic residues, Asp185 and Asp130) as an analogue of *A. niger* α -galactosidase, glucoamylase from *A. awamori* var. X-100^[26] (family GH15, with a a/a_6 fold, and two Glu catalytic residues, Glu179 and Glu400), and β -glucosidase A from *B. polymyxa*^[27] (family GH1, with a β/a_8 fold, and two Glu cata-

lytic residues Glu166, Glu352) using AutoDock 3.0. These enzymes were chosen since their crystallographic structures (see Table 8) with some inhibitors have been published. Furthermore, they belong to the same families and therefore may resemble the features of some of the enzymes whose inhibition features by **1–4** have been studied.

Table 8. Details of the enzyme X-ray crystal structures

| Enzyme | PDB Code | EC Number | Resolution (Å) | Reference |
|--|----------|-----------|----------------|-----------|
| α -Galactosidase from rice | 1UAS | 3.2.1.22 | 1.50 | [25] |
| Glucoamylase from <i>A. awamori</i> var. X100 | 1DOG | 3.2.1.3 | 2.40 | [26] |
| β -Glucosidase A from <i>B. polymyxa</i> | 1BGG | 3.2.1.21 | 2.30 | [27] |

Validation of the Docking Method

To ensure that the ligand orientations and positions obtained from the docking studies were likely to represent valid and reasonable potential binding modes of the inhibitors, the docking parameters were first validated for each crystal structure used.^[25–27] Each ligand found in the crystal structure of the enzyme–inhibitor complexes was docked into its corresponding enzyme and the AutoDock results were compared with the X-ray ones. The results for the validation method are given in Table 9. The low number of structures found for each cluster suggests that several independent docking runs can be conducted with ease, as performed here, to allow the minimization protocol to find a true energy minimum. Thus, the root mean square deviations (RMSD) for the lowest energy member indicate that the docking parameters can successfully predict the positions of the natural ligands, especially of D-galactose.

Table 9. Docking details of the natural ligands

| Protein | Ligand | Cluster (structures) | RMSD |
|--|-----------------------------------|----------------------|--------------|
| α -Galactosidase, rice | D-Galactose ^[25] | 1 (1) | 0.75 |
| Glucoamylase, <i>A. awamori</i> var. X100 | 1-Deoxyojirimycin ^[26] | 1 (2) | 1.45 |
| β -Glucosidase A, <i>B. polymyxa</i> | Gluconic acid ^[27] | 1 (7) 2 (5) | 1.12 1.08 |

Docking of Compounds 1–4 in α -Galactosidase from Rice, as a Model of Coffee Bean α -Galactosidase

After energy minimizations with the AMBER* force field, all protonated structures maintained the initial geometry, referred to as **A**, **B** or **C**. The results of docked monosaccharide analogues in the binding site of α -galactosidase from rice are listed in Table 10. The type of conformation (**A**, **B** or **C**) is given next to the compound code. The results of the inhibition studies of α -galactosidase from coffee beans are also included.^[9] All the compounds **1–4** were found to occupy the same binding site as D-galactose within α -galactosidase. The different conformers can bind in approximately the same region of the enzyme, with conformer **1A** having the most favorable interaction energy with the enzyme (see Table 10), being more favorable than

Table 10. Docking of monosaccharide analogues in the binding site of α -galactosidase from rice^[25]

| Comp. conformer | Cluster (structures) | Docked energy ^[a] | $\Delta G_{\text{binding}}$ | Intermolec. energy | Internal energy of ligand | % Inhibition ^[b] (coffee beans) |
|-----------------|----------------------|------------------------------|-----------------------------|--------------------|---------------------------|--|
| 1A | 1 (1) | -8.10 | -7.77 | -8.08 | -0.02 | 98 |
| 1B | 1 (1) | -7.98 | -7.64 | -7.95 | -0.03 | |
| 2A | 1 (4) | -7.77 | -7.53 | -7.84 | +0.07 | |
| 2C | 1 (3) | -6.47 | -6.19 | -6.51 | +0.03 | |
| 4A | 1 (2) | -7.28 | -7.38 | -7.69 | +0.41 | 74 |
| 4B | 1 (2) | -7.29 | -6.81 | -7.12 | -0.17 | |
| 3A | 1 (1) | -7.42 | -7.04 | -7.35 | -0.06 | 52 |
| 3B | 1 (1) | -7.29 | -7.12 | -7.43 | +0.15 | |

^[a] All energies are in kcal·mol⁻¹. ^[b] At a concentration of 1 mM of inhibitor, optimal pH, see ref.^[9]

conformer **1B** by about 0.5 kJ·mol⁻¹, which also binds to the enzyme better than any other conformer of the other compounds. As a key example, the conformation **1A** of the D-*gluco* analogue in the protein's binding site is shown in Figure 7.

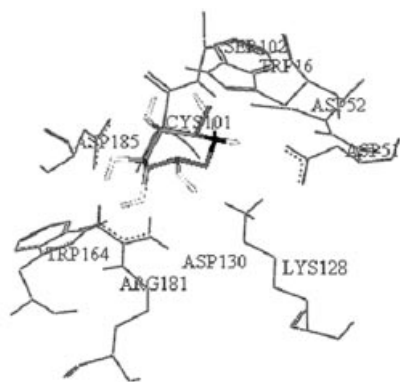


Figure 7. Docked structure of conformer **1A** of the D-*gluco* analogue in the binding site of α -galactosidase from rice^[25]

The same amino acid residues involved in the binding of galactose are also involved in the proposed complexes of the different monosaccharide mimics. In the crystal,^[25] Arg181, Trp164, as well as Asp185 formed hydrogen bonds with the O2 oxygen atom of galactose. Lys128 is hydrogen bonded to the two hydroxy oxygen atoms O3 and O4. Asp51 and Asp52 are hydrogen bonded to O4 and O6 atoms, respectively. In addition, a hydrophobic contact is observed for residue Trp16. The activity of **1A** as a galactose analogue can be explained by the replacement of the axial O4 of galactose by an axially oriented O3 in the **1A** conformer (Figure 7). Asp185 provides a bifurcated hydrogen bond to both O4 and O5, which adopt a pseudoequatorial orientation. Lys128, Arg181, Asp 51, and Asp52 are also involved in hydrogen bonding. Additionally, in this binding mode for **3**, the seven-membered ring of **1A** is surrounded by Trp16 and Trp164, a type of sugar–aromatic interaction seen in a variety of complexes between proteins and carbohydrates,^[29] which is also present in the other two docking models (see below). Also, the additional stabilization found in this conformer is probably due to hydrogen bonding between one of the protons attached to the nitrogen and Asp51, a feature that only occurs for the **1A** conformer.

Most strikingly, a fairly good quantitative relationship is observed between the experimental percentage of enzyme inhibition and the estimated free energy of binding for all compounds (Figure 8). This good agreement indicates that the binding modes proposed are reasonable. Although these results should be taken with caution due to the different nature of the enzymes tested for inhibition and used for docking, it is tempting to try to rationalize the interaction on the basis of the 3D structure of the complex.

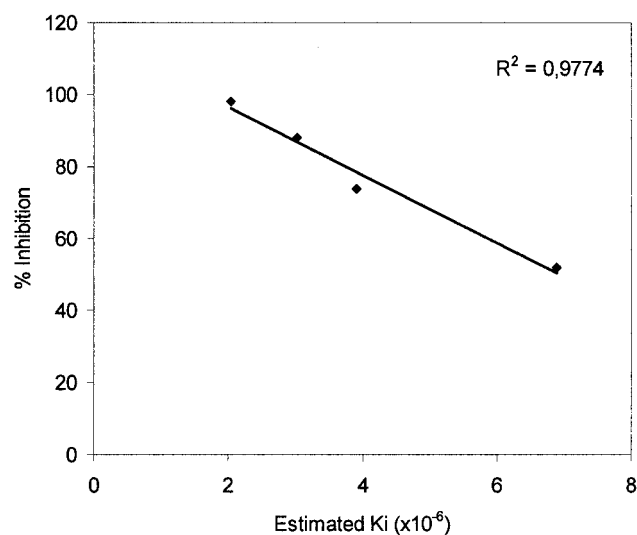


Figure 8. Relationship between the inhibition measured^[9] for *A. niger* α -galactosidase and the estimated K_i (inhibition constant) (from the interaction binding energies) of **1–4** for α -galactosidase of rice; the measured inhibition versus the free energy of binding also showed a fairly good correlation ($r^2 = 0.985$)

Glucoamylase from *A. awamori*

The results of the docking protocol for the binding of **1–4** to the glucoamylase from *A. Awamori* are summarized in Table 11. The X-ray structure of this enzyme presents its complex with 1-deoxynojirimycin.^[26] Residues that provide hydrogen bonding to 1-deoxynojirimycin are Asp55, Arg305, and Arg54. Besides, Glu179 is the catalytic acid and Glu400 is the catalytic base. The available biological data concerning inhibition^[9] with *A. Niger* and G1 glucoamylase (this work) are also included in the same table. It can be observed that the inhibition profile for both enzymes is similar.

Table 11. Docking of monosaccharide analogues in the binding site^[26] of glucoamylase from *A. awamori*

| Conformer | Cluster (structures) | Docked energy[a] | $\Delta G_{\text{binding}}$ | % Inhibition ^[b] (G1 glucoamylase) | % Inhibition ^[b] (<i>A. niger</i>) |
|-----------|----------------------|------------------|-----------------------------|---|---|
| 1A | 1 (8) | -9.33 | -9.12 | 24 | NI |
| 1B | 1 (3) | -9.73 | -9.32 | | |
| 2A | 1 (1) | -9.35 | -8.98 | | NI |
| | 3 (10) | -9.14 | -8.88 | | |
| 2C | 1 (7) | -9.62 | -9.28 | NI | |
| | 2 (14) | -9.43 | -8.95 | | |
| 4A | 1 (2) | -9.21 | -9.02 | | 45 |
| 4B | 1 (9) | -9.83 | -9.61 | 52 | |
| 3A | 1 (6) | -9.64 | -9.19 | | 34 |
| 3B | 1 (7) | -9.49 | -9.36 | 39 | |

In all cases, the docking protocol again leads to a binding mode fairly similar to that present in the crystal structure, with a perfect superimposition between 1-deoxynojirimycin and the glycomimetics. The residues mentioned above are interact similarly with the different glycomimetics. However, in this case, the best binding energy is provided by the α -L-gulo-like compound **4** in the **4B** conformation. The structure of **4B** docked in the enzyme's binding site is depicted in Figure 9 (a). Apart from hydrogen bonding involving Asp55, Arg305, Arg54, Glu179, and Glu400, there are also interactions with several aromatic residues (Tyr48, Trp52, Trp178, Trp317, and Trp417) that further stabilize **4B**. There is a fairly good superimposition with 1-deoxynojirimycin in the X-ray complex,^[26] and since only O3 of **4B** adopts a pseudoaxial orientation, the rest of the hydroxy groups may resemble the D-gluco-type configuration. Again, there is a very good correlation between docked energies, binding energies, and the inhibition data (Table 11).^[9] The compounds that provided no inhibition or weak inhibition to both enzymes are those that have the worst binding energies. The data suggest that compound **4** will be bound in a B-type conformation. The relative docked energies of α -L-gulo, **4B**, and β -D-manno, **3A**, analogues agree with the experimental relative inhibition of the two glucoamylases.^[9]

β -Glucosidase A from *B. polymyxa*

The results of these docking experiments and the inhibition data (for β -glucosidases from almonds^[9] and *B. Polymyxa*, this work) are summarized in Table 12. The best docking and binding affinity is provided by conformation **3B** of the β -D-manno-like analogue, as depicted in Figure 9 (b).

Again, the docking results predict that the monosaccharide analogues occupy the same binding site as the natural acyclic gluconic acid, as observed for the perfect superimposition for **3B** and the ligand in the crystallographic dataset^[27] [Figure 9 (b)]. In the crystal structure, hydrogen bonding to the ligand is provided by Gln20, His121, Glu405, and the two catalytic residues, Glu166 and Glu352. For all compounds, the conformation with the lowest docking energy is the one that predominates in free solution, with the exception of compound **2**. Strikingly, only the compound with the best docking energy (compound **3**) is able to inhibit *B. polymyxa*, while the best docking energies (*manno* **3** and *gulo* **4**) correlate well with the inhibition of

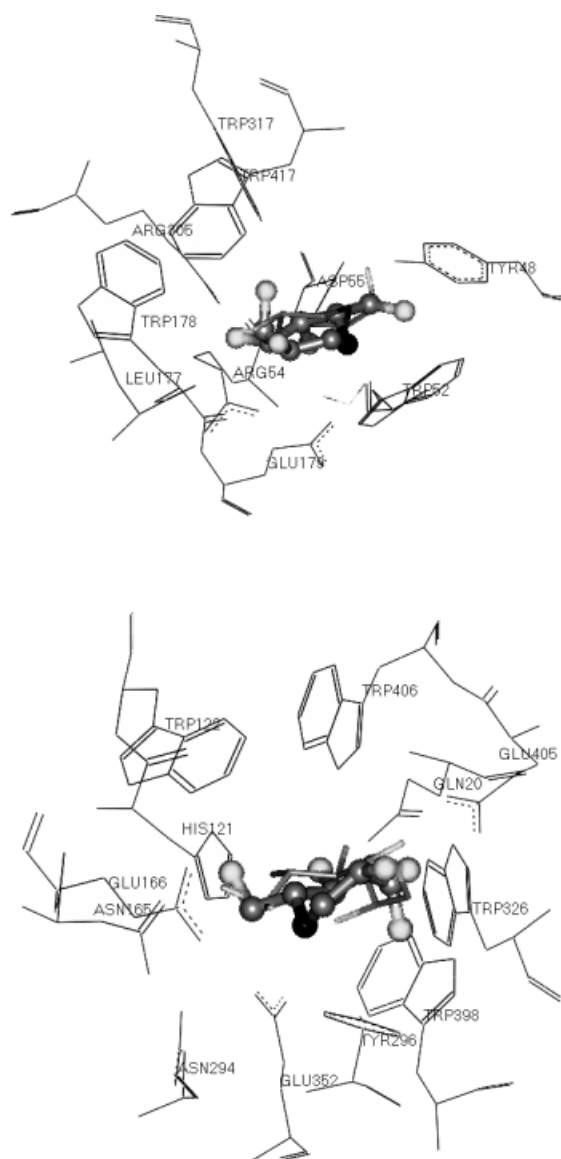


Figure 9. (a) The **4B** conformation of the α -L-gulo-like compound in the binding pocket of the glucoamylase from *A. awamori*;^[26] only O3 adopts a pseudoaxial orientation; (b) the **3B** conformation of the β -D-manno-like compound in the binding pocket of the β -glucosidase A from *B. polymyxa*;^[27] only O3 adopts a pseudoaxial orientation; monosaccharide analogues are represented as ball and stick models, and the ligands in the X-ray structure [deoxynojirimycin^[26] in (a) and gluconic acid^[27] in (b)] as stick models; hydrogen atoms have been removed for clarity

Table 12. Docking of monosaccharide glycomimetics in the binding site of β -glucosidase A from *B. polymyxa* [27]

| Conformer | Cluster (structures) | Docked energy ^[a] | $\Delta G_{\text{binding}}^{\text{[a]}}$ | % Inhibition ^[b] (<i>B. polymyxa</i>) | % Inhibition ^[b] (almonds) |
|-----------|----------------------|------------------------------|--|---|--|
| 1A | 1 (15) | -10.09 | -10.05 | NI | NI |
| 1B | 1 (57) | -10.87 | -10.55 | | |
| | 1 (13) | -10.32 | -9.94 | NI | |
| 2A | 2 (45) | -10.24 | | | NI |
| | 1 (7) | -9.83 | -9.46 | | |
| 2C | 2(35) | -9.68 | -9.42 | | |
| 4A | 1 (22) | -10.22 | -9.95 | NI | 39 |
| 4B | 1 (58) | -10.74 | -10.28 | | |
| 3A | 1 (16) | -10.23 | -9.99 | 72 | 62 |
| 3B | 1 (44) | -11.39 | -11.11 | | |

^[a] All energies are in kcal·mol⁻¹. ^[b] At a concentration of 1 mM of inhibitor, optimal pH, see ref.^[9] NI: unmeasurable inhibition.

β -glucosidase from almonds. Therefore, the binding mode shown in Figure 9 (b) may be taken as an indication of the binding of these monosaccharide analogues. In the docking complex, all the residues involved in hydrogen bonding to gluconic acid in the crystal, Asn165 and Asn294 interact with the ligand, which is in turn surrounded by Tyr296, and four Trp residues, Trp122, Trp326, Trp398, and Trp406. The almost equatorial orientation of the substituents of **3B** (except for O3) provides a good match for the required glucose shape that this glucosidase is able to hydrolyse.

Conclusions

These polyhydroxyazepan glycomimetics **1–4** may adopt two conformations in solution which display some selectivity towards different glycosidases. On this basis, they have been docked in the binding sites of three selected enzymes whose X-ray coordinates are available. These enzymes, which belong to the same family as those used in the inhibition assays, served as models to predict the actual conformation of the analogues in the binding state. In all cases a strikingly good correlation was observed between the docked energies of the models and the percentages of inhibition of the tested enzymes. Although with the α -galactosidase from rice, the preferred conformation of the best analogue in the bound state is the less favored one in solution (although there is still selectivity with respect to the other analogues), for the other two chosen enzymes, glucoamylase from *A. awamori* and β -glucosidase A from *B. polymyxa*, the preferred conformer for binding is the one which is more populated in solution. According to our results, this protocol may be used to understand the inhibition ability of glycomimetics. Further studies in our laboratories are aimed at the design of novel analogues with enhanced inhibitory potency.

Acknowledgments

We would like to thank the European Community's Human Potential Programme (under contract HPRN-CT-2002-00173), the DGICYT of Spain (Grant BQU2003-03550-C03-01) and the Swiss National Science Foundation for financial support. We also

thank the Ecole Normale Supérieure for the short stay of J.J.-B. in Paris. K.M.-M. and J.L.M.-F. thank CONACYT and the Universidad Nacional Autónoma (México) for a travel grant. We are indebted to Dr. Z. Fujimoto (Tsukuba, Japan) for providing the X-ray coordinates of rice α -galactosidase prior to availability and to Dr. A. Olson (Scripps Research Institute, USA) for providing AutoDock and auxiliary programs. We also thank Dr. J. Sanz-Aparicio (Madrid, Spain) and Dr. B. Svensson (Carlsberg Laboratory, Denmark) for samples of *B. polymyxa* glucosidase and G1 glucoamylase, respectively.

- [1] ^[1a] V. H. Lillelund, H. H. Jensen, X. Liang, M. Bols, *Chem. Rev.* **2002**, *102*, 515–553. ^[1b] O. R. Martin, P. Compain, *Curr. Top. Med. Chem.* **2003**, *3*, i–iv.
- [2] A. Mitrakou, N. Tountas, A. E. Raptis, R. J. Bauer, H. Schulz, S. A. Raptis, *Diab. Med.* **1998**, *15*, 657.
- [3] T. D. Butters, R. A. Dwek, F. M. Platt, *Curr. Top. Med. Chem.* **2003**, *3*, 561–574.
- [4] ^[4a] J. E. Groomman, *Rev. Infect. Dis.* **1990**, *12*, 908–911. ^[4b] G. S. Jacob, *Curr. Opin. Struct. Biol.* **1995**, *5*, 605–611 and references cited therein. ^[4c] M.-J. Papandreou, R. Barbouche, R. Guieu, M. P. Kieny, E. Fenouillet, *Mol. Pharmacol.* **2002**, *61*, 186–193.
- [5] ^[5a] S.-F. Wu, C.-J. Li, C.-L. Liao, R. A. Dwek, N. Zitzmann, Y.-L. Lin, *J. Virology* **2002**, *76*, 3596–3604 and references cited therein. ^[5b] A. Mehta, S. Ouzounov, R. Jordan, E. Simsek, X. Y. Lu, R. M. Moriarty, G. Jacob, R. A. Dwek, T. M. Block, *Antiviral Res.* **2003**, *57*, 56. ^[5c] P. Greimel, J. Spreitz, A. E. Stütz, T. M. Wrodnigg, *Curr. Top. Med. Chem.* **2003**, *3*, 513–523.
- [6] ^[6a] S. L. White, T. Nagai, S. K. Akiyama, E. J. Reeves, K. Grzegorzewski, K. Olden, *Cancer Commun.* **1991**, *3*, 83–91. ^[6b] P. E. Goss, J. Baptiste, B. Fernandes, M. Baker, J. W. Dennis, *Cancer Res.* **1994**, *54*, 1450–1457. ^[6c] P. D. Rye, N. V. Bovin, E. Vaslova, R. A. Walker, *Glycobiology* **1995**, *5*, 385–389. ^[6d] A. D. Elbein, R. D. Molyneux, in *Imino Sugars as Glycosidase Inhibitors: Nojirimycin and Beyond* (Ed.: A. E. Stütz), Wiley-VCH, Weinheim, **1999**, chapter 11, pp. 216–251. ^[6e] N. Zitzmann, A. S. Mehta, S. Carrouée, T. D. Butters, F. M. Platt, J. McCauley, B. S. Blumberg, R. A. Dwek, T. M. Block, *PNAS* **1999**, *96*, 11878–11882. ^[6f] N. Asano, *J. Enzyme Inhibition* **2000**, *15*, 215–234.
- [7] ^[7a] T. D. Heightman, A. T. Vasella, *Angew. Chem. Int. Ed.* **1999**, *38*, 750–770. ^[7b] A. Vasella, G. J. Davies, M. Böhm, *Curr. Opin. Chem. Biol.* **2002**, *6*, 619–629.
- [8] J. Alper, *Science* **2001**, *291*, 2338–2343.
- [9] H. Li, Y. Blériot, C. Chantereau, J.-M. Mallet, M. Sollogoub, Y. Zhang, E. Rodríguez-García, P. Vogel, J. Jiménez-Barbero, P. Sinaÿ, *Org. Biomol. Chem.* **2004**, *2*, 1492–1499.

- [10] T. Weimar, R. Woods, in *NMR spectroscopy of glycoconjugates* (Eds.: J. Jiménez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**.
- [11] M. Meyer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902–1906; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
- [12] For the first applications of TRNOE in sugar–protein interactions, see: [12a] V. L. Bevilacqua, D. S. Thomson, J. H. Prestegard, *Biochemistry* **1990**, *29*, 5529–5537. [12b] V. L. Bevilacqua, Y. Kim, J. H. Prestegard, *Biochemistry* **1992**, *31*, 9339–9349. For a detailed rigorous analysis of TRNOE cross peaks, see: [12c] P. L. Jackson, H. N. Moseley, N. R. Krishna, *J. Magn. Reson., Ser. B* **1995**, *107*, 289–292. For recent applications, see also ref.[15].
- [13] For applications of molecular docking protocols to glycosidase enzymes, see, for example: [13a] A. Laederach, M. K. Dowd, P. M. Coutinho, P. J. Reilly, *Proteins* **1999**, *37*, 166–175. [13b] J. K. Choi, B. H. Lee, C. H. Chae, W. Shin, *Proteins* **2004**, *55*, 22–33.
- [14] For information on families of glycosidases, see: P. M. Coutinho, B. Henrissat, *Carbohydrate-Active Enzymes*, **1999**; on the internet see <http://afmb.cnrs-mrs.fr/CAZY/index.html>
- [15] For applications of NMR to glycostructures, see: *NMR spectroscopy of glycoconjugates* (Eds.: J. Jiménez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**.
- [16] K. Stott, J. Stonehouse, J. Keeler, T.-L. Hwang, A. J. Shaka, *J. Am. Chem. Soc.* **1995**, *117*, 4199–4200.
- [17] For a discussion on the application of molecular mechanics force fields to sugar molecules, see: S. Pérez, A. Imberty, S. B. Engelsens, J. Gruza, K. Mazeau, J. Jiménez-Barbero, A. Poveda, J. F. Espinosa, B. P. van Eijck, G. Johnson, A. D. French, M. L. C. E. Kouwijzer, P. D. J. Grootenhuys, A. Bernardi, L. Raimondi, H. Senderowitz, V. Durier, G. Vergoten, K. Rasmussen, *Carbohydr. Res.* **1998**, *314*, 141–155.
- [18] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, *J. Comput. Chem.* **1990**, *11*, 440–467.
- [19] S. W. Homans, *Biochemistry* **1990**, *29*, 9110–9118.
- [20] N. L. Allinger, Y. H. Yuh, J. H. Liü, *J. Am. Chem. Soc.* **1989**, *111*, 8551–8559.
- [21] W. C. Still, A. Tempzyk, R. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.
- [22] C. A. G. Haasnoot, F. A. A. M. de Leeuw, C. Altona, *Tetrahedron* **1980**, *36*, 2783–2793.
- [23] A. Poveda, J. L. Asensio, M. Martín-Pastor, J. Jimenez-Barbero, *J. Biomol. NMR* **1997**, *10*, 29–43.
- [24] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639.
- [25] Z. Fujimoto, S. Kaneko, M. Momma, H. Kobayashi, H. Mizuno, *J. Biol. Chem.* **2003**, *278*, 313.
- [26] E. M. Harris, A. E. Aleshin, L. M. Firsov, R. B. Honzatko, *Biochemistry* **1993**, *32*, 1618.
- [27] J. Sanz-Aparicio, J. A. Hermoso, M. Martinez-Ripoll, J. L. Lequerica, J. Polaina, *J. Mol. Biol.* **1998**, *275*, 491.
- [28] K. Bock, J. O. Duus, *J. Carbohydr. Chem.* **1994**, *13*, 513–543.
- [29] See, for example: [29a] M. Muraki, *Protein Pept. Lett.* **2002**, *9*, 195–209. [29b] H. Kogelberg, D. Solís, J. Jiménez-Barbero, *Curr. Opin. Struct. Biol.* **2003**, *13*, 646–653, and references cited therein.

Received May 7, 2004

Synthesis, Conformational Studies and Mannosidase Stability of a Mimic of 1,2-Mannobioside

Silvia Mari,^[bl] Helena Posteri,^[al] Gilles Marcou,^[al] Donatella Potenza,^[al] Fabrizio Micheli,^[cl]
F. Javier Cañada,^[bl] Jesus Jimenez-Barbero,^{*[bl]} and Anna Bernardi^{*[al]}

Keywords: Carbohydrate mimics / Conformational analysis / Mannobioside

Dimethyl (1*S*,2*S*,4*S*,5*S*)-4-allyloxy-5-(α -mannosyloxy)cyclohexane-1,2-dicarboxylate (**3**) was designed as a structural mimic of α (1,2)mannobioside (**1**). Its synthesis and structural analysis by NMR spectroscopy and molecular modelling are described. The results show that **3**, like **1**, populates two low-energy conformations — stacked (*S*) and extended (*E*) — that are in fast dynamic equilibrium around the glycosidic linkage. Thus, the data confirm the expectation that the

pseudo-disaccharide can be used as a structural mimic of mannobioside. The mannosidase stability of **3** was found to be significantly higher (sixfold) than that of natural mannobioside. This is a very useful feature of this mimic and is encouraging for its future development.

(© Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, Germany, 2004)

Introduction

During the past few years a number of publications and reviews have illustrated the potential of sugar mimics in different fields of drug discovery.^[1–3] The main advantage of this approach is rooted in the improved drug-like character of glycomimetics compared to natural carbohydrates. Sugar mimics are generally more soluble and membrane penetrant, less hydrophilic and less metabolically labile than the sugars themselves.

Our groups are actively involved in the rational design and synthesis of glycomimetics based on the concept of scaffold replacement.^[4] In brief, we have proposed the use of conformationally stable cyclic diols to replace the non-pharmacophoric parts of bioactive oligosaccharides, while preserving the correct orientation of the pharmacophoric portion of the sugar. Computational tools have been used to predict the three-dimensional structures of virtual mimic candidates and to compare them with the structure of the natural counterpart. This approach has been validated for a 3,4-disubstituted galactose scaffold (a *cis* diol) using the GM1:CT recognition pair as a model system.^[5,6] Supported

by appropriate experimental work,^[7–10] molecular modelling has also allowed us to obtain qualitative predictions of the binding mode of the new substrates and to design further simplifications of glycomimetic structures. We now present our results relative to glycomimetics containing an analog of the α Man1,2-disubstituted framework (a *trans*-dial diol). A mimic of the 1,2-mannobioside **1** was prepared by replacing the reducing-end mannose unit with the carbocyclic diol **2** [(1*S*,2*S*,4*S*,5*S*)-dicarboxycyclohexane diol, DCCHD; Figure 1],^[4] a conformationally stable 1,2 *trans*-dial diol.^[4,11] Computer-aided molecular design (see below) suggested that **2** can be used as a mimic of the 2-substituted α -mannose unit in **1** without altering the overall shape of the molecule and therefore the presentation of the terminal (non-reducing end) residue. NMR studies confirmed that the resulting pseudo-mannobioside **3** has indeed the same conformational behavior as the natural disaccharide, but exhibits improved stability towards the activity of jack-bean mannosidase.

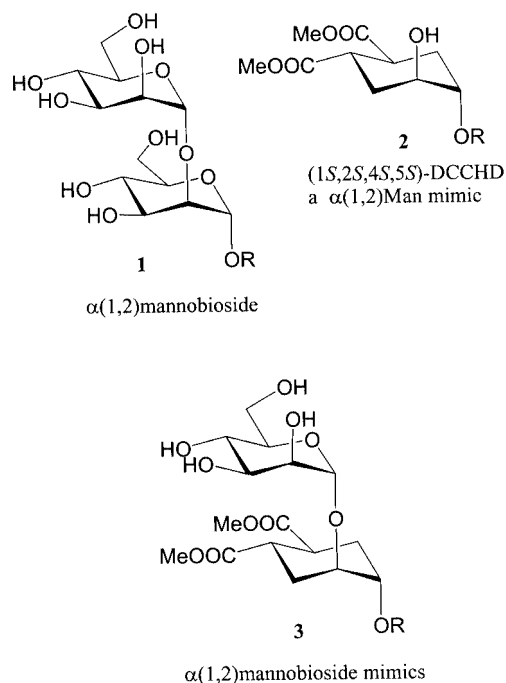
Synthesis of the Mannobioside Mimic **3**

The enantiopure *trans*-dial diol DCCHD scaffold **5** (Scheme 1) was synthesized in 81% overall yield starting from the known diacid **4**^[12] by *m*-chloroperbenzoic acid (MCPBA) double-bond oxidation, followed by Cu(OTf)₂-promoted opening of the epoxide with allyl alcohol.^[4] The pseudo-disaccharide skeleton was then assembled by glycosylation of the acceptor **5** (Scheme 1) with tetraacetylmannose trichloroacetimidate (**6**).^[13] The reaction was promoted by 0.25 molar equivalents of trimethylsilyl triflate (TMSOTf) in CH₂Cl₂ at –20 °C. At very short contact times (10 min) the orthoester **7** was initially formed by nu-

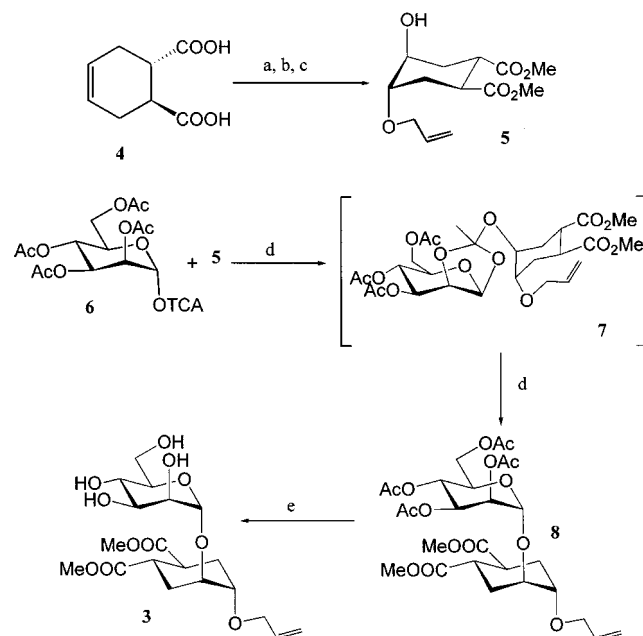
^[a] Dipartimento di Chimica Organica e Industriale e Centro di Eccellenza CISI, Università di Milano
Via Venezian 21, 20133 Milano, Italy
Fax: (internat.) +39-02-503-14092
E-mail: anna.bernardi@unimi.it

^[b] Departamento de Estructura y Funcion de Proteinas, Centro de Investigaciones Biologicas CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain
Fax: (internat.) +34-91-536-0432
E-mail: jjbarbero@cib.csic.es

^[c] GSK Medicine Research Centre
Via Fleming 4, 37135 Verona, Italy

Figure 1. Mannobioside **1** and its DCCHD-based mimic **3**

cleophilic attack of **5** on the 2-acetate protecting group of **6**.^[14] This product could be isolated by flash chromatography, and converted into **8** by treatment with TMSOTf in CH_2Cl_2 at -20°C .^[14] When allowing the reaction between **5** and **6** to stand for a somewhat longer time (20 min), the desired product **8** was obtained directly in 65% yield, after chromatography. Deacetylation (NaOMe in MeOH) quan-



Scheme 1. Synthesis of the pseudo-mannobioside **3**: (a) MeOH, cat. H_2SO_4 (95%); (b) MCPBA, CH_2Cl_2 (95%); (c) allyl alcohol, 10% $\text{Cu}(\text{OTf})_2$ (90%); (d) 25% TMSOTf, CH_2Cl_2 , -20°C (65%); (e) NaOMe, MeOH (88%)

tatively afforded **3**, whose structure was unequivocally confirmed by ^1H and ^{13}C NMR spectroscopy and by mass spectrometry.

Conformational Analysis of the Mannobioside Mimic **3**

The conformation and dynamics of the $\text{Man}\alpha 1\text{-2-Man}$ linkage were studied extensively in the oligomannose oligosaccharide $\text{Man}_9\text{GlcNAc}_2$.^[15] Computational^[15,16] and experimental^[17] results suggested that the linkage exists in two distinct, flexible conformations with similar φ values and different, slowly interconverting ψ values. The two conformations, which have been termed *S*, for “stacked”, and *E*, for “extended”, (Figure 2) are both represented in available X-ray structures of mannose oligosaccharides and account for the set of four NOE contacts typically observed across the glycosidic linkage.^[17] The H1-H1' contact (Figure 2) can arise from conformer *S*, and the H5-H1' and H1-H3' contacts (Figure 2) from conformer *E*. The fourth observed NOE contact belongs to the interglycosidic H1-H2' pair, which is at NOE distance in both conformations. (The mutually exclusive NOE contacts are marked by arrows in Figure 2).

We recently reported^[16] that modeling of **1** using the AMBER* force field and GB/SA solvation model gives results that are fully compatible with previous simulations and available experimental data. Two conformers, *S* (global minimum) and *E* (+6 kJ/mol), were located by an MC/EM conformational search, and were found to be freely interconverting and almost equally populated by MC/SD dynamic simulations. The same approach was then used to predict the three-dimensional structure of the 1,2- α -mannobioside mimic **3**; the molecule was calculated to populate the same region of conformational space as **1**. These results are presented in Figure 3. The unconventional numbering of the DCCHD ring shown in Figure 3 is used to facilitate comparison with **1**. An MC/EM conformational search of **3** located the *S* ($\varphi = 94^\circ$, $\psi = -72^\circ$; see c in Figure 3) and *E* ($\varphi = 70^\circ$, $\psi = -117^\circ$; see d in Figure 3) conformers at +2.6 kJ/mol and 0.0 kJ/mol relative energy, respectively. The dynamic simulation (10 ns, see b in Figure 3) showed that the two conformations are in fast dynamic equilibrium around the glycosidic linkage. Compared to **1**, the conformational space available to **3** is more biased towards the region populated by extended (*E*) conformations. Furthermore, the *E* region of the conformational space appears broad and is not centered around its local minimum at $\varphi = 70^\circ$, $\psi = -117^\circ$, but rather around larger values of ψ . This can clearly be seen in the map of Figure 3, and has important consequences on the calculated interproton distances, some of which display a strong dependence on the value of the ψ torsion (cf the distances calculated by the MC/SD dynamics with those of the isolated minima in Table 1). In summary, computer-aided molecular design clearly supports the expectation that the pseudo-disaccharide **3** can be used as a structural mimic of mannobioside **1**.

NOESY data were collected at 400 MHz for **3** in D_2O solution at 25°C (Table 1, column 1) and compared with the MC/SD simulation predictions. A qualitative analysis

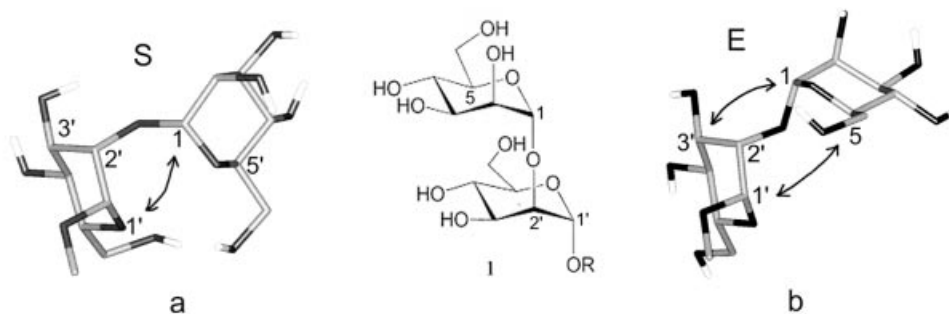


Figure 2. Low-energy conformations of the 1,2-mannobioside **1**: (a) the stacked conformation *S* ($\Phi = 88^\circ$, $\Psi = -58^\circ$); (b) the extended conformation *E* ($\Phi = 72^\circ$, $\Psi = -138^\circ$); characteristic NOE contacts are indicated by arrows

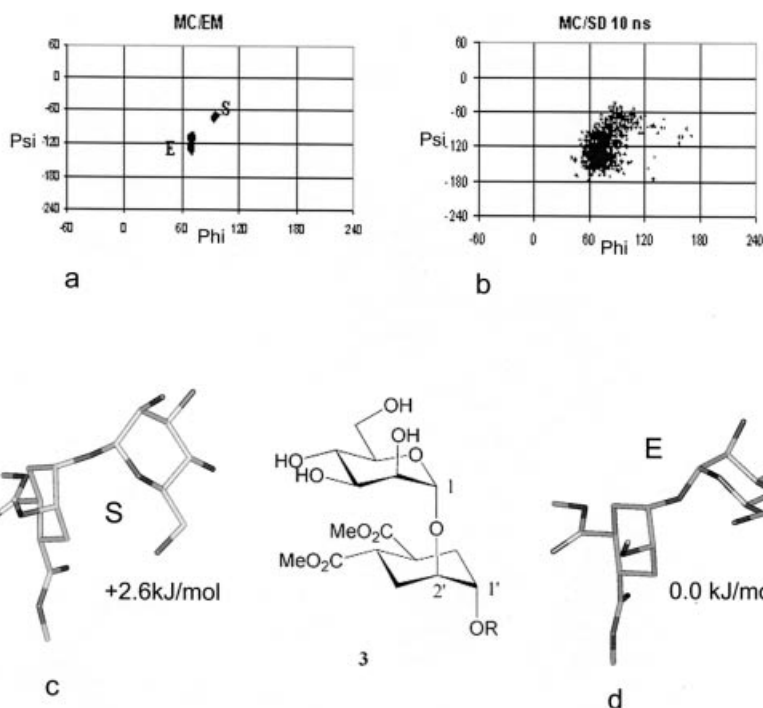


Figure 3. Conformational studies on pseudo 1,2- α -mannobioside **3** (MMOD 5.5, AMBER*, GB/SA water solvation); an unconventional numbering of the DCCHD moiety is used to help comparison with 1,2-mannobioside. Φ : O5–C1–O1–C2'; Ψ : C1–O1–C2'–C1'; (a) 10000 steps of MC/EM conformational search; (b) 10 ns of MC/SD dynamic simulation; (c) conformer *S* ($\Delta E = 2.6$ kJ/mol; $\Phi = 94^\circ$, $\Psi = -72^\circ$); (d) the lowest energy conformation *E* ($\Phi = 70^\circ$, $\Psi = -117^\circ$)

Table 1. Interresidue NOE contacts observed in the D_2O spectrum of **3**

| Proton pair ^[a] | NOE intensity ^[b] (D_2O) | Calcd. distance (Å) ^[c] | Calcd. distance <i>E</i> (Å) | Calcd. distance <i>S</i> (Å) | Exp. distance range in 1 (Å) ^[d] |
|----------------------------|--|---------------------------------------|---------------------------------|---------------------------------|---|
| H1-H1' | w | 3.4 | 3.9 | 2.6 | 2.9–3.0 |
| H1-H2' | s | 2.3 | 2.2 | 2.2 | 2.1–2.2 |
| H1-H4' | w | 3.8 | 4.2 | 4.6 | no NOE |
| H1-H3'ax | – | 4.0 | 4.1 | 4.5 | 3.2–3.6 |
| H1-H3'eq | s | 2.7 | 3.0 | 4.0 | – |
| H2-H4' | m | 3.7 | 4.0 | 4.8 | n.r. ^[e] |
| H5-H2' | m | 3.4 | 3.6 | 3.6 | n.r. ^[e] |
| H5-H1' | s | 2.6 | 2.3 | 4.4 | 2.5–2.7 |
| H5-H6'ax | m | 2.8 | 2.9 | 2.2 | – |
| H5-H4' | w | 4.2 | 4.5 | 3.2 | n.r. ^[e] |

^[a] Numbering as in Figures 2 and 3. ^[b] s: strong, m: medium, w: weak. ^[c] Evaluated from $\langle r^{-6} \rangle$ monitored during 10 ns of the MC/SD simulation. ^[d] From ref.^[15] ^[e] Not reported.

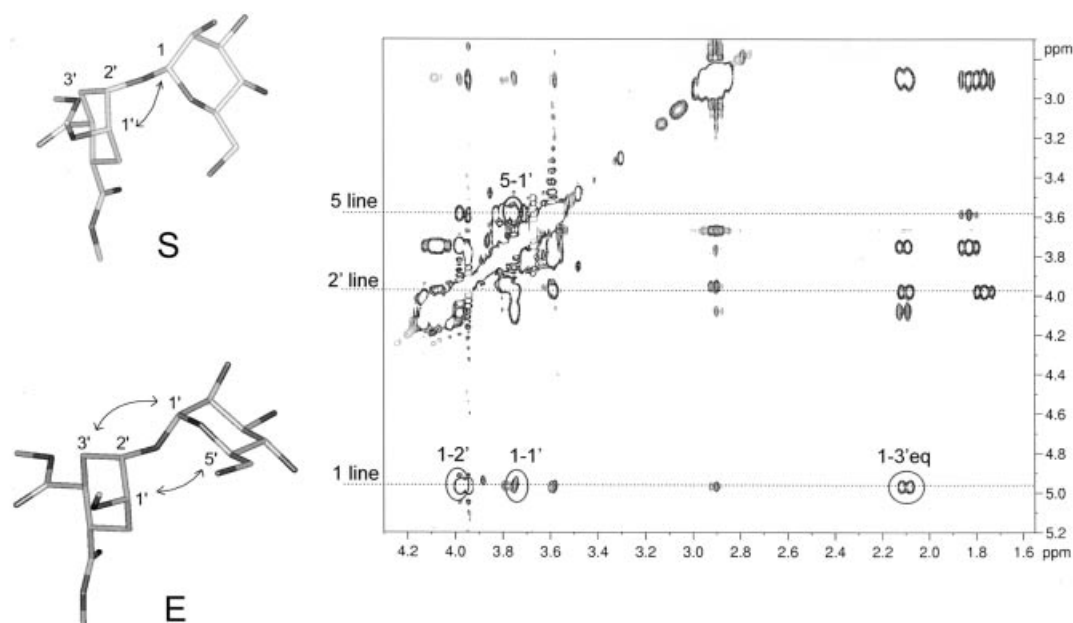


Figure 4. NOESY spectrum of mimic **3** in D₂O at 400 MHz (25 °C) with a mixing time of 800 ms; exclusive *S* and *E* conformation NOE contacts (indicated by the arrows) are found in the spectrum (circles), thus confirming that both the predicted conformations coexist

of the NOE crosspeaks using a basic strong-medium-weak intensity approach allowed us to deduce the presence of both *E* and *S* conformers in fast conformational equilibrium. Both sets of mutually exclusive NOE contacts were found in the spectra — the H1-H1' contact originating from the *S* conformers, and the H1-H3' and H5-H1' contacts originating from the *E* conformers (Figure 4).

Despite this qualitative approach, we could safely deduce that the *E* conformer is the major one in solution, given the strong differences between the intensities of the corresponding NOE cross-peaks. Indeed, the observed correlations are in excellent agreement with the simulations (Table 1).

Compared to the experimental data range reported for the equivalent distances in **1**^[15] (Table 1, last column), our data confirm that the conformational space of pseudo-mannobioside **3** is biased towards the extended conformation region. This is clearly shown by the weak intensity of the H1-H1' cross-peak (Table 1, entry 1, and Figure 4, inset of the NOESY spectrum), a reporter signal of conformer *S* (Figure 4, *S*), which is much stronger in **1** than in **3**. Furthermore, the strong interaction observed in the NOESY spectrum of **3** for the H5-H1' pair (Table 1, entry 8) is typical of extended conformations (Figure 4, *E*). This picture is also supported by the intense H1-H3'eq cross-peak (Table 1, entry 5, and Figure 4) and by the presence of a medium intensity H2-H4' NOE contact (Table 1, entry 6), which is not seen in **1**. This latter contact could not be explained by the calculated structure of the *E* local minimum (Table 1, Column 3), but is in good agreement with the results of the dynamic simulation (Table 1, Column 1) and is an indication that the dynamic behavior of the molecule is well reproduced by the AMBER* calculations.

Thus, the NMR studies experimentally support the computational hypothesis that **3** behaves as a structural mimic

of mannosidase, and preferentially populates the extended (*E*) conformation.

Mannosidase Stability of **3**

The stability of **3** to mannosidase-catalyzed hydrolysis was investigated by submitting it to hydrolytic conditions in the presence of commercial jack-bean mannosidase. The substrate was incubated with increasing concentrations of the enzyme at pH 4.5 (maximum activity pH for α -mannosidases).^[18] After 30 min of incubation at 37 °C the reaction mixture was lyophilized and the products were analyzed by GC. The extent of hydrolysis was evaluated by integration of the peaks relative to **3** (retention time, t_R , of 13.5 min) and to the aglycon **5** (t_R = 7.1 min). At the lowest enzyme concentration (0.5 $\mu\text{g/mL}$) 4% hydrolysis of mimic **3** was revealed by GC analysis, while, upon increasing the enzyme concentration up to 10 $\mu\text{g/mL}$, only 10% hydrolysis was observed. Under the same conditions 24% of the natural disaccharide **1** is already hydrolysed at an enzyme concentration of 0.5 $\mu\text{g/mL}$ and 60% is hydrolysed at 10 $\mu\text{g/mL}$ enzyme concentration. Thus, **3** indeed appears to be a substrate of jack-bean mannosidase, but it appears to be significantly more stable than **1** to the action of the enzyme. These observations prompted us to examine the ability of **3** to perform as an inhibitor for this enzyme.

Investigation of **3** as a possible inhibitor of mannosidase activity was carried out using α -mannosidase from jack-bean and *p*-nitrophenol- α -D-mannoside (PNP-Man) as the substrate (Table 2). The enzymatic reactions were run at 25 °C for 20 min and then quenched by adding a 1 M carbonate solution until pH 12, which inactivates the enzyme. The extent of the reaction was calculated from the absorbance of *p*-nitrophenoxide measured at a wavelength of 400 nm ($\epsilon_{400} = 1.77 \times 10^4 \text{ M}^{-1}\text{cm}^{-1}$). The concentration of *p*-nitro-

phenoxide versus concentration of substrate in the reaction volume was plotted and fitted to the Michaelis–Menten equation to give a K_M value of 4.7×10^{-3} M, which is similar to the reported value of 2.5×10^{-3} M.^[18] Qualitative inhibition studies were then performed using **3** as inhibitor and working with an enzyme concentration of 2.5 $\mu\text{g}/\text{mL}$ (Table 2). The inhibitory effect of **3** allows us to estimate an average value for the inhibition constant, K_i , of 1.0×10^{-2} M, assuming a competitive inhibitor model (Table 2). The K_i values obtained are similar to those reported in the literature^[18] for the known inhibitors mannono (1 \rightarrow 4) and (1 \rightarrow 5) lactone (1.0×10^{-2} M and 1.2×10^{-4} M, respectively).

Table 2. Inhibition of jack-bean-mannosidase-catalyzed hydrolysis of PNP-Man using **3**. The inhibition experiments were carried out for 20 min at 25 °C using 1.8 mM PNP-Man in citrate buffer (pH 4.5), a constant concentration of the enzyme (2.5 $\mu\text{g}/\text{mL}$) and variable concentrations of the inhibitor **3**. The reaction was quenched at pH 12 carbonate buffer and the concentration of the *p*-nitrophenoxide [PNP] was measured by UV spectroscopy. v_0 (0.025 mM/min) was obtained under the same conditions but in the absence of **3**.

| [3] (mM) | [PNP] (mM) | $v_{\text{Inhibited}}$ (mM/min) | v_i/v_0 | $K_i^{[\text{a}]}$ (mM) |
|----------------------|---------------|------------------------------------|-----------|----------------------------|
| 12.8 | 0.28 | 0.014 | 0.565 | 11.9 |
| 10.1 | 0.31 | 0.015 | 0.626 | 12.3 |
| 5.1 | 0.32 | 0.016 | 0.651 | 6.8 |
| 2.5 | 0.41 | 0.021 | 0.841 | 9.7 |

$$^{[\text{a}]} K_i = \frac{[I] \cdot K_M}{\frac{K_M + [S]}{v_i/v_0} - K_M - [S]}$$

where $K_M = 4.67$ mM, $[I] = [\mathbf{3}]$ and $[S] = [\text{PNP-Man}]$.

Discussion and Conclusions

The results presented here show that the enantiopure *trans*-diaxial diol scaffold **5** can be used as a replacement for a 1,2-disubstituted α -mannose residue to give, in a facile and high-yielding synthetic sequence, a pseudo-mannobioside **3** which shows a high structural similarity with the natural α -1,2-mannobioside. The conformational analysis of **3** performed by NMR spectroscopy and supported by computational modelling reveals that **3**, like **1**, populates two low-energy conformations *S* (stacked) and *E* (extended) that are in fast dynamic equilibrium. Relative to the natural counterpart, the conformational space available to the pseudo-disaccharide appears to be biased toward the *E* region.

The α -1,2-mannobioside is the recognized epitope in various important sugar-protein complexes.^[19] Appropriate epitope presentation is a key factor in determining ligand affinity to carbohydrate-binding proteins. For instance, it has been shown recently that, although sialic acid and galactose, per se, interact weakly with the cholera toxin (with millimolar dissociation constants), when appropriately pre-

sented on three-dimensionally well-defined structures, such as the GM1 ganglioside^[20] or appropriate structural mimics,^[5] they give rise to tight complexes with dissociation constants in the nanomolar range. The structural similarity of **3** with the natural disaccharide **1** is therefore an important element that depends critically on the nature of the dicarboxycyclohexanediol aglycon chosen to mimic the reducing-end mannose. The mimic **3** could thus be introduced into more complex oligosaccharidic structures, preserving epitope presentation, while imparting metabolic stability to the synthetic construct. Furthermore, the structure of **3** allows easy conjugation to polyvalent aglycons, using either the carboxy groups or the ether substituent. The conformational behaviour of mimic **3** that we describe here could have useful implications in recognition events where the *E* conformation of the 1,2-mannobioside is selectively recognized by the receptor.

An increased stability to enzymatic degradation is often cited as one of the advantages of glycomimetics over natural oligosaccharides in the development of drugs. The stability of **3** to mannosidase-catalyzed hydrolysis appears, indeed, to be significantly higher than the stability of the natural compound **1**. This is a very useful feature of this mimic and is very encouraging for the future developments of DCCHD-based glycomimetic drugs.

Experimental Section

NMR spectra were recorded at 25–30 °C on Bruker AVANCE 400 and 500 MHz spectrometers. The cyclohexanediol moiety is numbered as in Figure 3. Chemical shifts ¹H and ¹³C NMR spectra are expressed in ppm relative to TMS or to DSS for spectra recorded in D₂O. The NOESY spectra of **3** were recorded with mixing times of 400, 600 and 800 ms. Mass spectrometry was performed with a VG 7070 EQ-HF apparatus (FAB ionization), an Omnix apparatus (MALDI ionization) or an Apex II ICR FTMS (ESI ionization). Optical rotations $[\alpha]_D$ were measured in a 1-dm path-length cell with 1 mL capacity on a Perkin–Elmer 241 polarimeter. Thin layer chromatography (TLC) was carried out with precoated Merck F₂₅₄ silica gel plates. Flash chromatography (FC) was carried out with Macherey–Nagel silica gel 60 (230–400 mesh). Solvents were dried by standard procedures and reactions requiring anhydrous conditions were run under nitrogen. The diacid **4**^[12] and the trichloroacetimidate **6**^[13] were prepared by published procedures.

Synthesis of Dimethyl (1*S*,2*S*,4*S*,5*S*)-4-Allyloxy-5-hydroxycyclohexane-1,2-dicarboxylate (5**):** H₂SO₄ (0.5 mL, 0.009 mol) was added to a solution of the diacid **4**^[12] (7.02 g, 0.041 mol) in dry MeOH (100 mL). The solution was stirred at reflux for about 20 h and then concentrated under vacuum to about half of the original volume. EtOAc was added, the organic phase was washed with saturated NaHCO₃ and dried with Na₂SO₄ and the solvent was evaporated under reduced pressure to give dimethyl (1*S*,2*S*)-cyclohex-4-ene-1,2-dicarboxylate in 95% yield. An analytical sample was purified by flash chromatography: $[\alpha]_D = +127.3$ ($c = 1.23$, CHCl₃). ¹H NMR (200 MHz, CDCl₃, 25 °C): $\delta = 2.6$ – 2.1 (m, 4 H, CH₂), 2.85 (m, 2 H, H¹, H²), 3.7 (s, 6 H, OCH₃), 5.7 (m, 2 H, H⁴, H⁵) ppm. ¹³C NMR (53.3 MHz, CDCl₃, 25 °C): $\delta = 24.5$, 40.96, 51.45, 124.66, 174.79 ppm.

The crude reaction product was then submitted to epoxidation conditions. Thus, MCPBA (1.85 g, 7.6 mmol) was added to a solution of the diester (1.5 g, 5.5 mmol) in CH_2Cl_2 (13 mL). The solution was stirred at room temperature for 3 h and the organic phase was washed with saturated NaHCO_3 , dried and the solvents evaporated under reduced pressure. The crude was purified by flash chromatography on silica gel (hexane/EtOAc, 85:15) to yield the epoxide in 95% yield. $[\alpha]_{\text{D}} = +71.6$ ($c = 1.27$, CHCl_3). $^1\text{H NMR}$ (200 MHz, C_6D_6 , 25 °C): $\delta = 1.55$ (ddd, $J = 15$, $J = 11$, $J = 2.3$ Hz, 1 H), 1.95 (m, 2 H), 2.30 (ddd, $J = 15$, $J = 4$, $J = 2$ Hz, 1 H), 2.55 (m, 1 H), 2.65 (m, 1 H), 2.75 (m, 1 H), 3.10 (ddd, $J = 11$, $J = 11$, $J = 4.6$ Hz, 1 H), 3.28 (s, 3 H), 3.30 (s, 3 H) ppm.

The catalyst $\text{Cu}(\text{OTf})_2$ (0.07 mmol) was added to a solution of the epoxide (150 mg, 0.7 mmol) in allyl alcohol (1.5 mL, 2.2 mmol). The reaction mixture was stirred for 3 h at room temperature, whilst monitoring by TLC (hexane/EtOAc, 6:4). After completion of the reaction a saturated solution of NH_4Cl and NH_3 was added and the mixture was extracted with EtOAc. The organic layer was dried and the solvents evaporated under reduced pressure. The crude mixture was purified by flash chromatography on silica gel (hexane/EtOAc, 1:1) to give **5** in 90% yield. $^1\text{H NMR}$ (200 MHz, C_6D_6 , 25 °C): $\delta = 2.05$ (m, 5 H), 2.52 (br. s, 1 H, exch), 3.40 (m, 9 H), 3.70–4.00 (m, 4 H), 5.15 (m, 2 H), 5.80 (m, 1 H) ppm. $^{13}\text{C NMR}$ (75 MHz, CDCl_3 , 25 °C): $\delta = 27.0$, 30.6, 39.0, 39.4, 51.9, 67.3, 69.8, 75.9, 116.9, 134.8, 175.0 ppm. MS (FAB+): $m/z = 273$ $[\text{M} + \text{H}^+]$.

Mannosylation of 5. Synthesis of 8: Ground 4-Å molecular sieves were added to a solution of **5** (493 mg, 1.81 mmol) and $\text{6}^{[13\text{C}]}$ (1.33 g, 2.71 mmol) and the mixture was dried in vacuo overnight. Dry CH_2Cl_2 (50 mL) was then added under N_2 and the solution was stirred for about 15 min. The temperature was then adjusted to -20 °C and TMSOTf (63 μL , 0.724 mmol) was added. The solution was stirred at -20 °C for about 10 min, until TLC revealed the disappearance of **5** (hexane/EtOAc, 6:4). The reaction was neutralized by adding triethylamine and concentrated. The molecular sieves were then removed by filtration and the product was isolated by flash chromatography (hexane/EtOAc, 6:4), and purified a second time by chromatography (CHCl_3 /acetone, 95:5), to give the orthoester **7** in 86% yield. $[\alpha]_{\text{D}} = +47$ ($c = 1$, CHCl_3). $^1\text{H NMR}$ (400 MHz, CDCl_3 , 25 °C): $\delta = 1.75$ (s, 3 H, CH_3), 1.8–1.99 (m, 4 H, $\text{H}^{3\text{Dax,eq}}$), 2.2–2.15 (m, H^9 , COCH_3), 2.95–3 (m, 2 H, $\text{H}^{4\text{D}}$, $\text{H}^{5\text{D}}$), 3.51 (d, 1 H, $\text{H}^{1\text{D}}$), 3.69 (s, 3 H, OCH_3), 3.7 (s, 3 H, OCH_3), 3.71 (m, 1 H, $\text{H}^{5\text{D}}$), 3.89 (m, 1 H, $\text{H}^{2\text{D}}$), 3.92 (dd, $J = 12$, $J = 5$ Hz, 1 H, $\text{H}^{7\text{a}}$), 4.09 (dd, $J = 12$, $J = 6$ Hz, 1 H, $\text{H}^{7\text{b}}$), 4.11 (dd, $J = 12$, $J = 5$ Hz, 1 H, $\text{H}^{6\text{a}}$), 4.28 (dd, $J = 12$, $J = 2.5$ Hz, 1 H, $\text{H}^{6\text{b}}$), 4.6 (m, 1 H, H^2), 5.12–5.31 (m, 2 H, H^9), 5.19 (s, 1 H, H^3), 5.25 (s, 1 H, H^4), 5.45 (d, $J = 2$ Hz, 1 H, H^1), 5.8–6.0 (m, 1 H, H^8) ppm. $^{13}\text{C-HETCOR}$ (100.6 MHz, CDCl_3 , 25 °C): $\delta = 34$, 35, 40, 62.5, 66, 69, 70.5, 71, 73, 75, 77.5, 98, 117, 135 ppm. MS (FAB+): $m/z = 603$ $[\text{M} + \text{H}^+]$.

Rearrangement of the orthoester was achieved by adding TMSOTf (0.65 μL , 0.0035 mmol), to a cold (-20 °C) solution of the orthoester **7** (21.3 mg, 0.035 mmol) in CH_2Cl_2 (600 μL). The mixture was stirred under nitrogen whilst monitoring by TLC (CHCl_3 /acetone, 95:5). After 10 min the reaction was complete. The mixture was then treated with triethylamine and the solvents evaporated under reduced pressure. The residue was purified by flash chromatography on silica gel (CHCl_3 /acetone, 97:3) to give **8** in 50% yield. The same product was also obtained in 65% yield by running the mannosylation of **5** with **6** at -20 °C for 20 min, and monitoring the conversion of **7** by TLC (CHCl_3 /acetone, 95:5). $[\alpha]_{\text{D}} = +38$ ($c = 1.26$, CHCl_3). $^1\text{H NMR}$ (400 MHz, C_6D_6 , 25 °C): $\delta = 1.74$ (s, 3 H, COCH_3), 1.77 (s, 3 H, COCH_3), 1.81 (s, 3 H, COCH_3), 1.85 (s, 3

H, COCH_3), 2.04–2.17 (m, 4 H, $\text{H}^{3\text{D}}$, $\text{H}^{6\text{D}}$), 3.3–3.5 (m, 2 H, $\text{H}^{4\text{D}}$, $\text{H}^{5\text{D}}$), 3.40 (s, 3 H, OCH_3), 3.46 (s, 3 H, OCH_3), 3.57 (m, 1 H, $\text{H}^{1\text{D}}$), 3.85 (dd, $J = 13$, $J = 6$ Hz, 1 H, $\text{H}^{7\text{a}}$), 3.95 (m, 2 H, $\text{H}^{7\text{b}}$, $\text{H}^{2\text{D}}$), 4.34 (m, 1 H, H^5), 4.37 (dd, $J = 12$, $J = 8$ Hz, 1 H, $\text{H}^{6\text{a}}$), 4.49 (dd, $J = 12$, $J = 6$ Hz, 1 H, $\text{H}^{6\text{b}}$), 5.11 (dd, $J = 9.2$, $J = 1.5$ Hz, 1 H, $\text{H}^{9\text{a}}$), 5.15 (d, $J = 1.5$ Hz, 1 H, H^1), 5.27 (dd, $J = 16$, $J = 1.5$ Hz, 1 H, $\text{H}^{9\text{b}}$), 5.65 (dd, $J = 3$, $J = 2$ Hz, 1 H, H^2), 5.8 (t, $J = 9.7$ Hz, 1 H, H^4), 5.82–5.91 (m, 2 H, H^3 , H^8) ppm. $^{13}\text{C-HETCOR}$ (100.6 MHz, C_6D_6 , 25 °C): $\delta = 28.5$, 52, 63, 70, 71, 72, 73, 75, 96, 118 ppm. MALDI-TOF: $m/z = 625.45$ $[\text{M} + \text{Na}^+]$.

Synthesis of the Pseudo-1,2- α -mannobioside 3: 1 M Methanolic sodium methoxide (398 μL , 0.398 mmol) was added to a solution of **8** (120 mg, 0.199 mmol) in MeOH (20 mL) and stirring was continued at room temperature for 18 h. The reaction was monitored by TLC (CHCl_3 /MeOH/ H_2O , 60:35:5). Amberlite IR-120 was then added, the resin was removed by filtration and the solution was concentrated. The residue was purified by flash chromatography (CHCl_3 /MeOH, 9:1), to give **3** in 88% yield. $[\alpha]_{\text{D}} = +58.2$ ($c = 1$, CH_3OH). $^1\text{H NMR}$ (500 MHz, D_2O , 25 °C): $\delta = 1.75$ (m, 1 H, $\text{H}^{3\text{Dax}}$), 1.82 (m, 1 H, $\text{H}^{6\text{Dax}}$), 2.09 (m, 2 H, $\text{H}^{3\text{Deq}}$, $\text{H}^{6\text{Deq}}$), 2.88 (dt, $J = 11.5$, $J = 3.4$ Hz, 2 H, $\text{H}^{4\text{D}}$, $\text{H}^{5\text{D}}$), 3.57 (m, 1 H, H^5), 3.58 (m, 1 H, H^4), 3.65 (s, 3 H, OCH_3), 3.66 (s, 3 H, OCH_3), 3.68 (m, 1 H, $\text{H}^{6\text{b}}$), 3.74 (q, $J = 3$ Hz, 1 H, $\text{H}^{1\text{D}}$), 3.77 (q, $J = 3.4$ Hz, 1 H, H^3), 3.79 (m, 1 H, $\text{H}^{6\text{a}}$), 3.94 (dd, $J = 3.4$, $J = 1.8$ Hz, 1 H, H^2), 3.97 (q, $J = 3$ Hz, 1 H, $\text{H}^{2\text{D}}$), 4.03 (m, 1 H, CH_2Oallyl), 4.09 (m, 1 H, CH_2Oallyl), 4.95 (d, $J = 1.8$ Hz, 1 H, H^1), 5.20 (m, 1 H, $\text{CH}_2=$), 5.29 (m, 1 H, $\text{CH}_2=$), 5.91 (m, 1 H, $\text{CH}=\text{O}$) ppm. $^{13}\text{C NMR}$ (125 MHz, D_2O , 25 °C): $\delta = 27$, 39.2, 52.6, 61.05, 66.9, 70, 70.5, 70.6, 70.7, 73.5, 73.7, 98.7, 118.3, 134.2 ppm. MS (FAB+): $m/z = 457$ $[\text{M} + \text{Na}^+]$. HRMS (ESI+) for $\text{C}_{19}\text{H}_{30}\text{O}_{11}\text{Na}^+$: calcd. 457.16803; found 457.16586.

Enzymatic Studies: Mannosidase stability assays were performed in 20 μL samples of 5 mM substrate (**3** or **1**) in pH 4.5 0.1 M phosphate-citrate buffer at 37 °C. Substrates were submitted to increasing concentrations of jack-bean mannosidase. Three enzymatic reactions were run, with 0.5, 2.5 and 10 μg protein/mL solutions prepared from stock commercial enzyme suspension (Sigma-M7257, 5 mg/mL). After 30 min incubation the crude mixtures were frozen and freeze-dried. In all cases lyophilized samples were completely trimethylsilylated by treatment with trimethylsilylimidazole and pyridine at 60 °C for 30 min, and were then analyzed by gas chromatography. GC analysis was performed with a Perkin–Elmer Autosystem. A fused silica column was employed (dimension 5 m \times 0.25 mm \times 0.22 μm) with SPB-1. Carrier gas: He flow at 20 psi. Volume injected: 1 μL . Program temperature used: 80 °C for 2.5 min, 10 °C/min to 150 °C and 15 °C/min to 250 °C. Injector temperature: 280 °C; FID temperature: 280 °C. Retention times: **3** 13.5 min, **5** 7.1 min, mannose 8.1 min, mannoside **1** 13.1 min.

Determination of the K_{M} (4.67 mM) and v_{MAX} (0.13 mM) parameters was performed in 50 μL reaction volume with 0.25 $\mu\text{g}/\text{mL}$ enzyme concentration with *p*-nitrophenol- α -D-mannoside (PNP-Man) as substrate. Seven different substrate concentrations were used (from 0.3 mM to 15 mM), with an incubation time of 20 min. Quenching was performed by adding 300 μL of 1 M carbonate buffer at pH 12. The absorbance of the *p*-nitrophenoxide released during the enzymatic reactions was measured at 400 nm and 25 °C on a UV/Vis Spectrometer (Perkin–Elmer Lambda 6).

Mannosidase inhibition studies were performed using a fixed 1.8 mM substrate (PNP-Man) concentration and variable inhibitor (**3**) concentrations in a reaction volume of 20 μL (Table 2). Enzymatic reactions were run with 2.5 $\mu\text{g}/\text{mL}$ enzyme concentration in

pH 4.5 0.1 M phosphate-citrate buffer at 25 °C. After 20 min the reactions were quenched by addition of 130 μ L of 1 M carbonate solution (pH 12) and the absorbance was measured at 400 nm and 25 °C on a UV/Vis Spectrometer.

Computational Methods: All calculations were performed using the MacroModel/Batchmin^[21] package (version 5.5) and the AMBER* force field with the Senderowitz–Still all-atom pyranose parameters.^[22] Charges were taken from the force field (all-atom charge option). Water solvation was simulated using MacroModel's generalized Born GB/SA continuum solvent model.^[23] The conformational search for **3** was carried out using 10000 steps of the usage-directed MC/EM procedure. The glycosidic linkages and all the extra-annular C–C bonds were used as explicit variables during the Monte Carlo search. Extended non-bonding cut-off distances (a van der Waals cut-off of 8.0 Å and an electrostatic cut-off of 20.0 Å) were used both for the MC/EM and MC/SD calculations. The same degrees of freedom of the MC/EM searches were used in the MC/SD runs. All simulations were performed at 300 K, with a dynamic time-step of 1.5 fs and a frictional coefficient of 0.1 ps⁻¹. One Monte Carlo step was performed after each dynamic step. The average acceptance ratio was 6%. Two simulations of 5 ns each were performed starting from the lowest energy *E* and *S* conformations. Structures were sampled every 2 ps and saved for later evaluation. The interatomic distances reported in Table 1 under the MC/SD header were evaluated from $\langle r^{-6} \rangle$, which was monitored during the simulation (option MDDI of Batchmin).

Acknowledgments

The project was supported by the Improving Human Potential Program under contract HPRN-CT-2002–00173 (Glycidic Scaffolds Network). Funding from the Ministry of Science and Technology of Spain (BQU2003-3550-C03), MIUR, and CNR is also acknowledged.

- [1] P. Sears, C. H. Wong, *Angew. Chem.* **1999**, *111*, 2446–2471; *Angew. Chem. Int. Ed.* **1999**, *38*, 2300–2324, and references cited therein.
- [2] R. Thomson, M. von Izstein, in *Carbohydrate Based Drug Discovery* (Ed.: C. H. Wong), **2004**, vol. 2, pp. 831–861.
- [3] R. Horstkorte, O. T. Keppler, W. Reutter, in *Carbohydrate Based Drug Discovery* (Ed.: C. H. Wong), **2004**, vol. 2, pp. 863–873.
- [4] A. Bernardi, D. Arosio, L. Manzoni, F. Micheli, A. Pasquarollo, P. Seneci, *J. Org. Chem.* **2001**, *66*, 6209–6216.
- [5] A. Bernardi, P. Brocca, A. Checchia, S. Sonnino, F. Zuccotto, *J. Am. Chem. Soc.* **1999**, *121*, 2032–2036.
- [6] A. Bernardi, D. Arosio, S. Sonnino, *Neurochem. Res.* **2002**, *27*, 539–545, and references cited therein.
- [7] A. Bernardi, D. Potenza, A. M. Capelli, A. García-Herrero, F. J. Cañada, J. Jiménez-Barbero, *Chem. Eur. J.* **2002**, *8*, 4597–4612.
- [8] A. Bernardi, D. Arosio, L. Manzoni, D. Monti, H. Posteri, D. Potenza, S. Mari, J. Jimenez-Barbero, *Org. Biomol. Chem.* **2003**, *1*, 785–792.
- [9] [9a] E. A. Merrit, S. Sarfaty, F. v. d. Akker, C. L'Hoir, J. A. Martial, W. G. J. Hol, *Protein Sci.* **1994**, *3*, 166–175. [9b] E. A. Merrit, S. Sarfaty, M. G. Jobling, T. Chang, R. K. Holmes, T. R. Hirst, W. G. J. Hol, *Protein Sci.* **1997**, *6*, 1516–1528. [9c] E. A. Merrit, W. G. J. Hol, *Curr. Opin. Struct. Biol.* **1995**, *5*, 165–171 and references cited therein.
- [10] [10a] D. Arosio, S. Baretta, S. Cattaldo, D. Potenza, A. Bernardi, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3831–3834. [10b] A. Bernardi, D. Arosio, D. Potenza, I. Sanchez-Medina, S. Mari, F. J. Cañada, J. Jimenez-Barbero, *Chem. Eur. J.* **2004**, *10*, 4395–4406.
- [11] V. V. Samoshin, V. A. Chertkov, L. P. Vatlina, E. K. Dobretsova, N. A. Simonov, L. P. Kastorsky, D. E. Gremyachinsky, H.-J. Schneider, *Tetrahedron Lett.* **1996**, *37*, 3981–3984.
- [12] A. Bernardi, D. Arosio, D. Dellavecchia, F. Micheli, *Tetrahedron: Asymmetry* **1999**, *40*, 3403–3407.
- [13] [13a] T. Ren, D. Liu, *Tetrahedron Lett.* **1999**, *40*, 7621–7625. [13b] R. R. Schmidt, J. Michel, *Angew. Chem. Int. Ed. Engl.* **1980**, *19*, 731–732. [13c] R. R. Schmidt, *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 212–235.
- [14] W. Wang, F. Kong, *J. Org. Chem.* **1998**, *63*, 5744–5745; W. Wang, F. Kong, *Angew. Chem. Int. Ed.* **1999**, *38*, 1247–1250.
- [15] R. J. Woods, A. Pathiaseril, M. R. Wormald, C. J. Edge, R. A. Dwek, *Eur. J. Biochem.* **1998**, *259*, 372–386, and references cited therein.
- [16] A. Bernardi, A. Colombo, I. Sánchez-Medina, *Carbohydr. Res.* **2004**, *339*, 967–973, and references cited therein.
- [17] M. R. Wormald, A. J. Petrescu, Y.-L. Pao, A. Glithero, T. Elliot, R. A. Dwek, *Chem. Rev.* **2002**, *102*, 371–386 and references cited therein.
- [18] Y.-T. Li, *J. Biol. Chem.* **1967**, *242*, 5474–5480.
- [19] [19a] Y. Bourne, V. Zamboni, A. Barre, W. J. Peumans, E. J. M. Van Damme, P. Rouge, *Structure* **1999**, *7*, 1473–1482. [19b] D. N. Moothoo, B. Canan, R. A. Field, J. Naismith, *Glycobiology* **1999**, *9*, 539–545. [19c] D. A. R. Sanders, D. N. Moothoo, J. Raftery, A. J. Howard, J. R. Helliwell, J. H. Naismith, *J. Mol. Biol.* **2001**, *310*, 875–884. [19d] C. A. Bewley, *Structure* **2001**, *9*, 931–940; L. C. Chang, C. A. Bewley, *J. Mol. Biol.* **2002**, *318*, 1–8. [19e] S. R. Shenoy, L. G. Barrientos, D. M. Ratner, B. R. O'Keefe, P. H. Seeberger, A. M. Gronenborn, M. R. Boyd, *Chem. Biol.* **2002**, *9*, 1109–1118. [19f] C. A. Bewley, S. Kiyonaka, I. Hamachi, *J. Mol. Biol.* **2002**, *322*, 881–889. [19g] C. N. Scanlan, R. Pantophlet, M. R. Wormald, E. O. Saphire, R. Stanfield, I. A. Wilson, H. Katinger, R. A. Dwek, P. M. Rudd, D. R. Burton, *J. Virol.* **2002**, *76*, 7306–7321.
- [20] W. B. Turnbull, B. L. Precious, S. W. Homans, *J. Am. Chem. Soc.* **2004**, *126*, 1047–1054.
- [21] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, *J. Comp. Chem.* **1990**, *11*, 440–467.
- [22] H. Senderowitz, W. C. Still, *J. Org. Chem.* **1997**, *62*, 1427–1438.
- [23] W. C. Still, A. Tempzyk, R. Hawley, T. Hendrickson, *J. Am. Chem. Soc.* **1990**, *112*, 6127–6129.

Received July 23, 2004

Conformer selection of a phenyl-lactic GM1 glycomimetic by three different lectins

Article in preparation

In this last section, our preliminary results on the conformational selection of ligands by different proteins are discussed. In particular, we here refer to a specific molecule designed and synthesized to be a CTB5 ligand, but which could be also recognized by other galactose-specific lectins, such as VAA (Viscum Album agglutinin) and CG14 (Chicken Galectin 14). The interesting behavior of these systems is the different conformer selection of the same molecule by different proteins.

These preliminary NMR results suggested us to try to model also the binding site of VAA, by using the same protocol already described for the CTB5 case.

NMR studies of phenyl-lactic mimic with VAA and CG14

Characterization of the binding features of **1** (see Chart 1 for definition) in solution was carried with a final concentration of 1.24 mM of **1** and 25 μ M of the corresponding lectin in D₂O, thus with a molar ratio ligand:subunit of about 25:1 (considering that both VAA and CG14 have two binding sites for compound **3**).

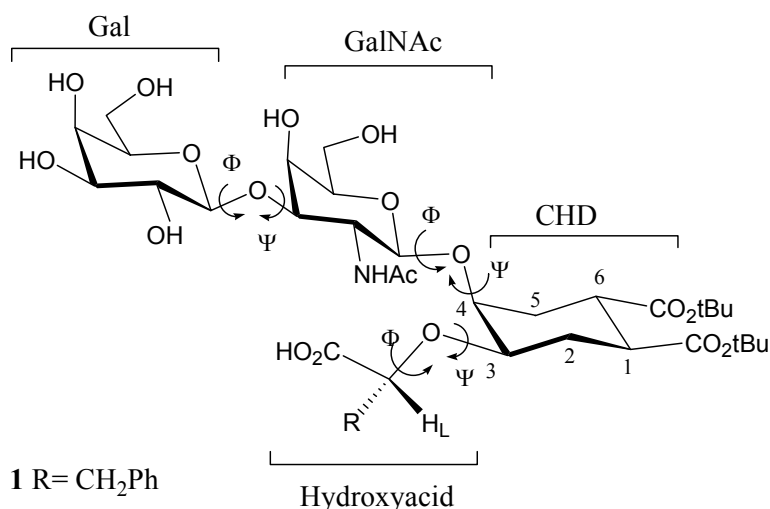


Chart 1. Phenyl-lactic mimic (compound **1**). Description and nomenclature employed.

TR-NOESY experiments were performed at different mixing times (150 and 300 ms for the VAA containing sample, and 100, 200 and 300 ms for the CG14 one) and ROESY spectra were also recorded (with 200 ms of mixing time for the VAA, while 150 and 300 ms for the CG14 one). In order to identify the binding epitope, All the experiments were recorded on a Bruker 500 MHz at 298 K (for the VAA containing samples) and 288 K (for the CG14 containing samples). No differences in chemical shifts were detected passing from the free to the bound state, while a general broadening of the signals was

obtained. Moreover, in the bound state both with VAA and with CG14, scalar coupling less than 5 Hz - that were easily detected in the free state- were impossible to be distinguished; furthermore, the typical increasing in the background signals of the aliphatic region were observed. These behaviors confirmed the binding of the ligand to the proteins (Fig. 1).

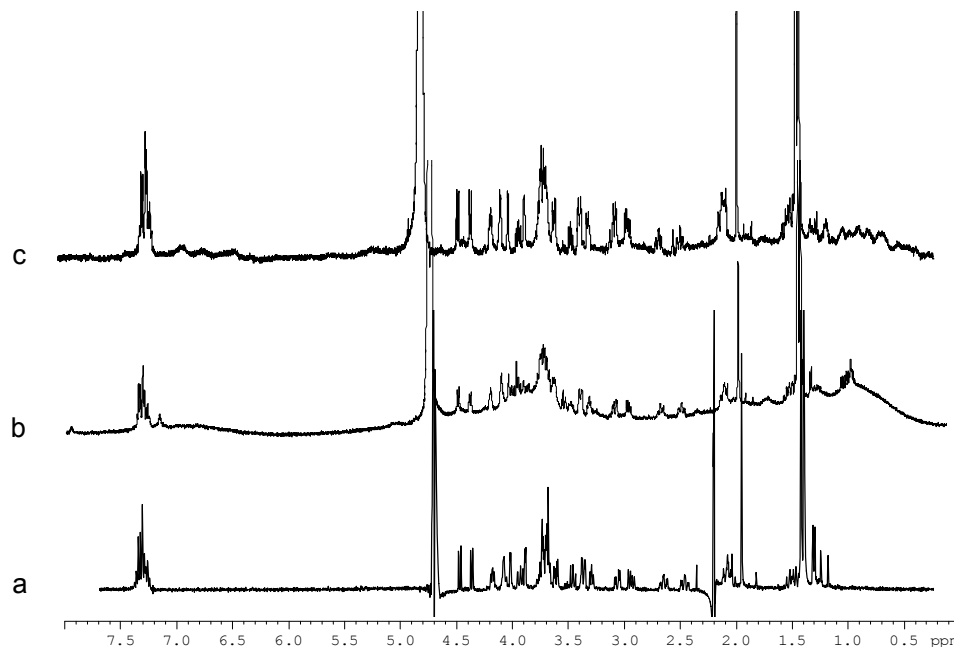


Figure 1. 1D spectra of compound 1. Reference spectrum (a) of the free state, and 1D spectra of the bound states with VAA (b) and CG14 (c).

The NOEs contacts detected in the TR-NOESY and ROESY spectra are listed in the table 1. Comparing the NOE sets for the free ligand (fig.2) with those for the bound states with VAA (fig.3A) and CG14 (fig.3B), it is possible to suggest that the pseudo-trisaccharide unit [Gal(β 1 \rightarrow 3)GalNAc(β 1 \rightarrow 4)diol] maintains its conformation.

The NOEs intra-residue contacts, GalNAc H1 with H3 and H5, and Gal H1 with H3 and H5 Gal, as well as those for the diol moiety, confirmed that all the three six-membered rings are adopting a chair conformation in free and bound states. Regarding the inter-residue NOEs contacts, the strong H-1 Gal / H-3 GalNAc contact can be correlated with a major syn conformation around the Gal(β 1 \rightarrow 3)GalNAc linkage. The same behavior is found for the GalNAc(β 1 \rightarrow 4)diol moiety, where a strong H-1 GalNAc / H-4 CHD cross peak is observed.

Table 1. Key NOE contacts for **1** in the free state, bound to VAA, and bound to CG14. The last column shows the NOE cross peak set of the bound state with Cholera Toxin B Pentamer, as described in reference a. The estimated distances (r , Å) according to a full matrix relaxation approach are also given ($\pm 10\%$).

| Proton Pair | | Intensity FREE STATE | r (Å) | Intensity bound state with VAA | r (Å) | Intensity bound state with CG14 | r (Å) | Intensity bound state with CTB5^a | r (Å) |
|-------------|---------|------------------------------------|---------|--|---------|---|---------|---|---------|
| GN1 | GN3 | strong | 2.5 | strong | 2.5 | strong | 2.5 | medium | 2.5 |
| | GN5 | strong | 2.6 | strong | 2.6 | strong | 2.6 | strong | 2.6 |
| | CHD-4 | strong | 2.5 | strong | 2.5 | strong | 2.5 | strong | 2.5 |
| | HL | N.O. | >3.5 | N.O. | >3.5 | N.O. | >3.5 | N.O. | >3.5 |
| G1 | G3 | medium | 2.5 | strong | 2.5 | strong | 2.5 | medium | 2.5 |
| | G5 | strong | 2.6 | strong | 2.6 | strong | 2.6 | strong | 2.6 |
| | GN3 | strong | 2.4 | strong | 2.4 | strong | 2.4 | strong | 2.4 |
| Ph | GN1 | medium | 2.8 | N.O. | | N.O. | | medium | 2.8 |
| | HL | medium | 2.7 | N.O. | | N.O. | | medium | 2.7 |
| | CHD-4 | medium | 2.8 | N.O. | | N.O. | | medium | 2.8 |
| | GN6 | weak | | N.O. | | N.O. | | medium | |
| | GN3* | medium | | N.O. | | N.O. | | medium | |
| | GN5 | medium | | N.O. | | N.O. | | medium | |
| HL | CHD-3* | strong | 2.3 | medium | | very weak | >3.3 | strong | 2.3 |
| | CHD-2eq | strong | 2.4 | very weak | | very weak | >3.3 | strong | 2.4 |
| | CHD-1 | weak | 3.0 | N.O. | >3.5 | very weak | >3.3 | weak | 3.0 |
| CHD-4 | CHD-5eq | strong | | strong | | strong | | strong | |
| | CHD-3 | strong | | strong | | strong | | strong | |
| | GN1 | strong | | strong | | strong | | strong | |
| | HL | N.O. | >3.5 | medium ^o | 2.5 | medium | 2.5 | weak | |

* GN3 and H3-CHD are isochronous

N.O. = not observable noe contact

^o observable only on CHD4 line

a) A. Bernardi, J. Jiménez-Barbero et al. *Chem. Eur. J.* 2004, 10, 4395 – 4406

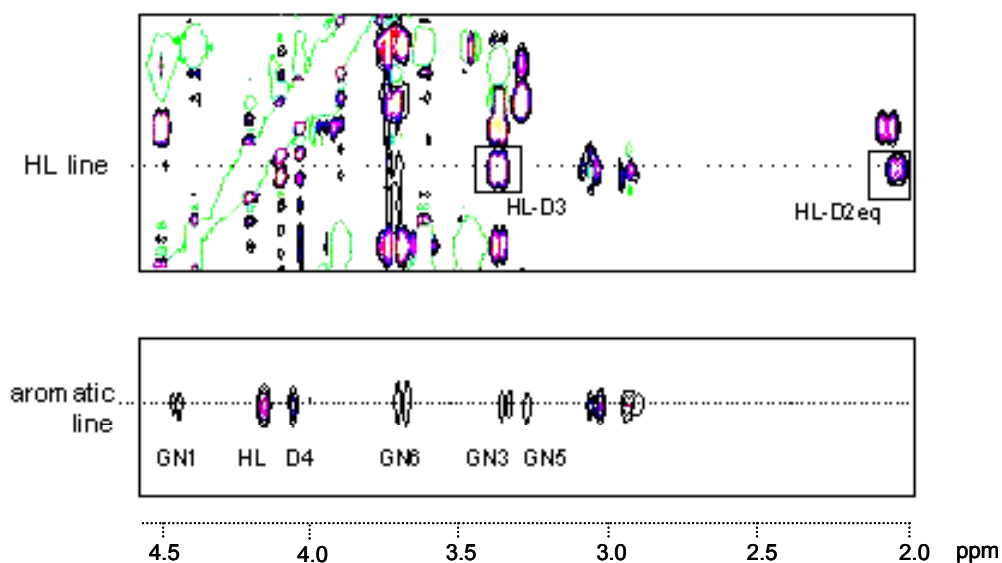


Figure 2. 2D NOESY of compound 1 in the free state

The lateral chain, represented by the phenyl and HL cross peaks sets, shows the most interesting behaviour. For **1** bound to the Cholera Toxin B Pentamer, the observed NOE contacts were the same both in free state and when bound to CTB5, thus suggesting that this mimic exists in a well defined conformation that was the same found in the toxin binding site. Therefore, ligand **1** did not suffer from any conformational penalty for toxin binding.

However, for the bound state with both VAA and CG14 lectins, a complete different situation we found. Major differences, in both cases, are found in the aromatic noe contacts (Tab. 1). As shown in the corresponding spectra (fig. 3A and B), no cross peaks are detected at the aromatic resonance line. This fact indicates that the phenyl group adopts a different orientation in the VAA and CG14 recognition sites and does not stack with the N-acetyl galactosamine ring, in contrast with the observed behavior for the bound state to cholera toxin. In these cases, the phenyl ring should move away from the GalNac moiety and to point in a different orientation with respect to both galactosamine and cyclohexandiol rings.

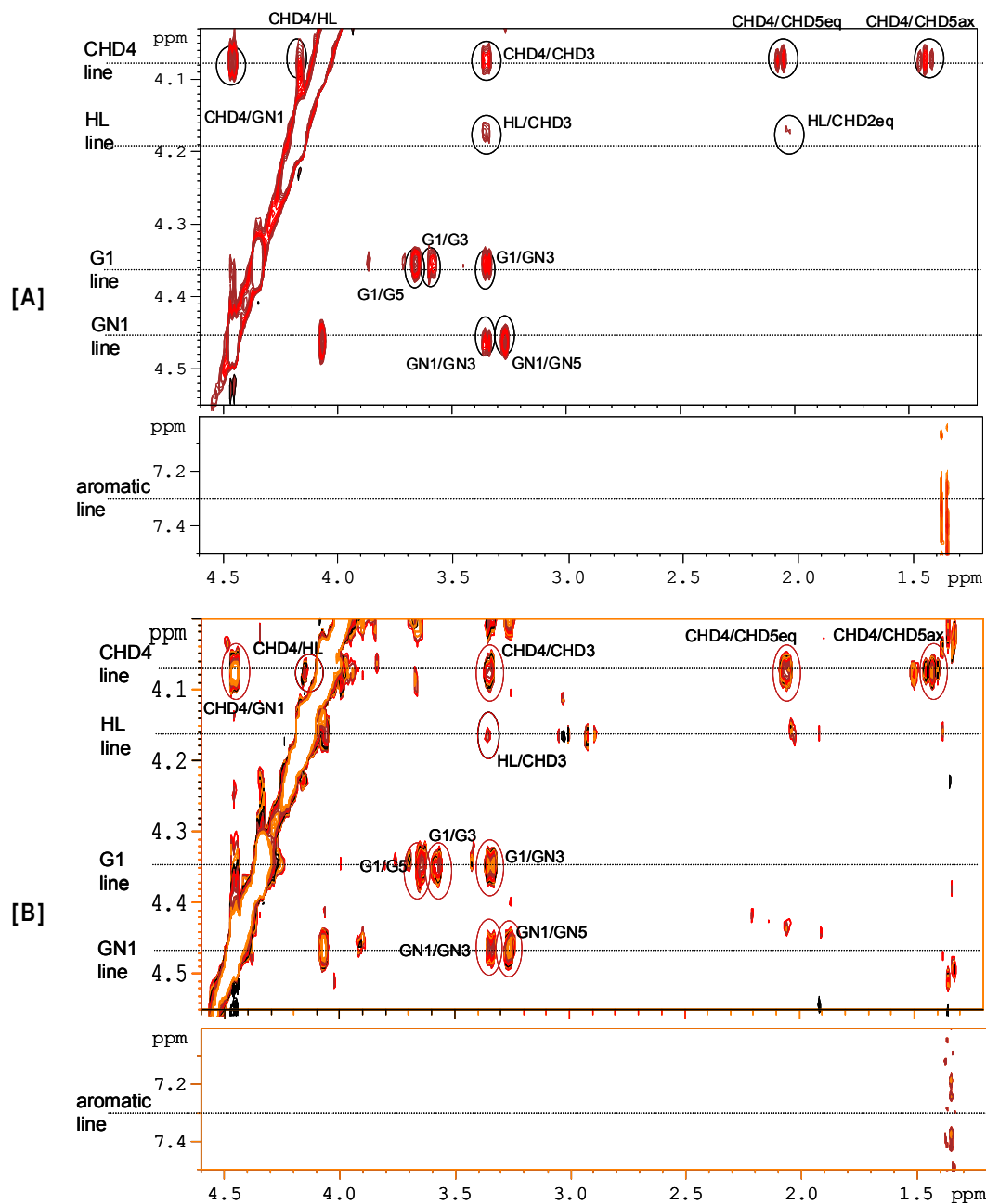


Figure 3. TR-NOESY spectra of **1** bound to VAA [A] and CG14 [B] (D_2O , mixing time 300ms)

Moreover, the distinct intensities of the exclusive H4-CHD / HL NOE cross peak suggested that different rotamers are selected by these two lectins. This cross-peak is not present either in the free state or in the bound state to CTB5 (fig. 4), while it clearly appears in the presence of VAA or CG14.

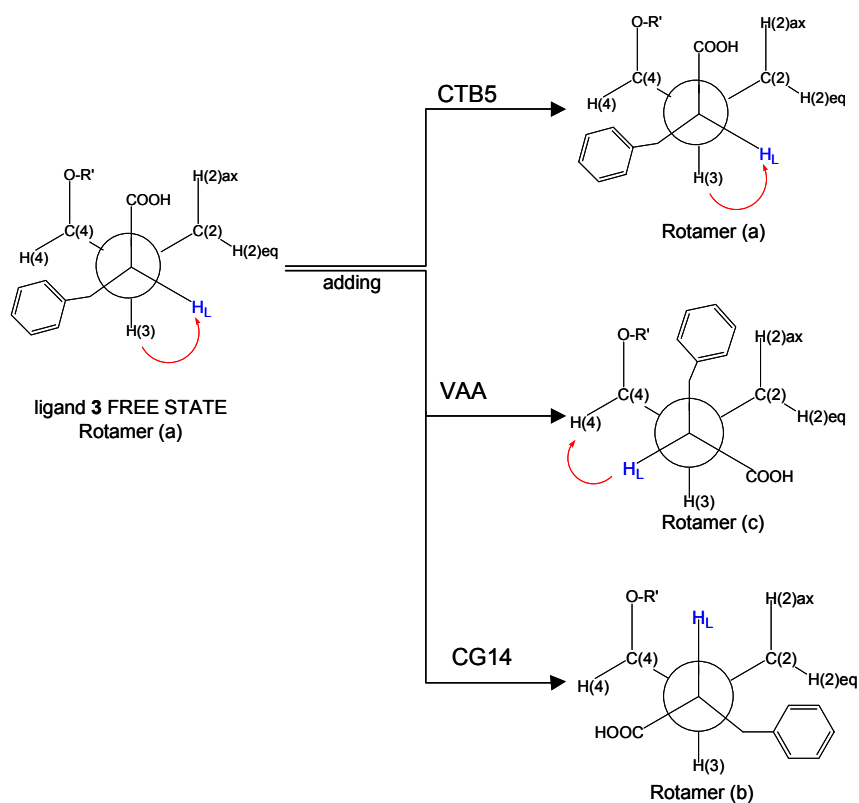


Figure 4. Preferred rotamers for the lateral chain orientation of 1 in the different free or bound states.

Nevertheless, conformational differences could also be found for the bound state of the glycomimetic with these two lectins. The TR-NOESY spectra for the VAA containing sample (fig.3A) revealed the HL / H₄-CHD and HL / H₃-CHD cross-peaks that could only be explained by the existence of rotamer (c) in fig.4. However, for the CG14 containing complex, only the HL / H₄-CHD cross-peak was evidenced, and the HL / H₃-CHD contact was negligible. Thus, rotamer (b) of 1 seems to be predominant in the CG14 binding site. Thus, rotamer (c) is preferred in the bound state with VAA (fig.4), while rotamer (b) is predominant in the bound state with CG14.

As conclusion, it can be safely stated that the three different receptors bind distinct conformations of the same analogue. Reasonably, this fact has to do with the different architecture of the corresponding binding sites. Thus, we are presently carrying out a variety of modeling protocols, including docking methods to present a definitive structural view to further justify the non ambiguous experimental NMR data.

Contributions to Symposia

One of the fundamental aspects of my Thesis period has been the possibility to participate to international congresses and courses, both in terms of poster presentation, oral communications and invited talks. Thus, I would like to list all this contributions:

12th European Carbohydrate Symposium
Grenoble, France
6-11 July 2003
(<http://eurocarb12.cermav.cnrs.fr/>)

Poster:

Mimics of Ganglioside GM1 as Cholera toxin ligands: structural study by NMR of their free and toxin-bound conformation.

XX Reunión Bienal de Química Orgánica
Zaragoza, Spain
9-12 June 2004
(<http://wzar.unizar.es/actos/xrbqo/>)

Oral communication:

Mimics of Ganglioside GM1: Structural Study by NMR of their Free and Bound Conformations with Cholera toxin, Viscum album agglutinin and Galectin CG14

Course:

Taller de Resonancia Magnética Nuclear de sistemas paramagnéticos y diamagnéticos. Interacciones Moleculares,
Barcelona, Spain
20-22 April 2004

II Reunión Bineal de RMN,
Santiago de Compostela, Spain
11-12 September 2004
(<http://desoft03.usc.es/bienal.asp>)

Poster:

1D-STD NMR Experiments on living cells

Course:

Curso Avanzado de Resonancia Magnética Nuclear,
Jaca, Spain
13-17 June 2005

Euromar 2005 (18th EENC)
Veldhoven, The Netherlands
3-8 July 2005
(<http://www.eenc2005.org/>)

Poster:

Conformational Analysis of oGM3 and its Derivative using Residual Dipolar Coupling and Aqueous Molecular Dynamic

SMASH Congress
Verona, Italy
25- 28 September 2005
(<http://www.smashnmr.org/home.htm>)

Oral communication:

Small Molecules Binding to Protein Receptors: the NMR point of view

In the next pages, the illustrations of the above mentioned posters are given.



Acknowledgements

...a great teamwork makes the difference

First of all I would like to thank my Thesis' director Prof. Jesús Jiménez-Barbero for giving me the opportunity to join the Glycidic Scaffold group and start this great adventure in Spain. My thanks go also to all the members of Network, in particular to Prof. Anna Bernardi, Prof. Francesco Nicotra and Dr. Yves Blierot with whom I had the possibility to start proliferate collaborations which possibly will last over the Network.



My great thank goes to all the member of Jesus' group, I just point out Dr. Javier Cañada but my words go to all of you, for being so helpful in my day-after-day problems and questions. A particular mention goes to Carmen and Víctor, my tireless Spanish's teachers and above all my great friends: I wish you all the success you desire and at least all the happiness you gave me.

My thanks go also to Prof. James H. Prestegard for giving me the great opportunity to cross the ocean and join his Lab at CCRC in Georgia. A particular acknowledgment goes to Beverly, Pete, Kristein, Xiaobin, Greg (and his cats), LianMei, Anita and Lora for the great help given to me and to my husband.



My thank is also for Prof. Robert J. Woods, for his great help running my MD simulations and for the free access to his giant cluster. I hope to see his wonderful Jaguar running on Italian roads.



Last but not least, my great acknowledgment to all the facilities I had access to and particularly to all the persons behind those. From protein expression and purification at IQOG to NMR services at UCM and CCRC. Lola, Angel and John: thank you so much.