UNIVERSITY OF NAPLES
"FEDERICO II"

FACULTY OF MATHS, PHYSICS AND NATURAL SCIENCES

PHD IN CHEMICAL SCIENCES
XXV CYCLE

Nuclear Magnetic Resonance spectroscopy studies of proteins-glycoconjugates interactions

ROBERTA MARCHETTI

TUTOR
Dr. Alba Silipo

SUPERVISOR
Prof. Lucia Falcigno

Contents

Papers related to PhD thesis:

1. X-ray structural studies of the entire extracellular region of the serine/threonine kinase PrkC from Staphylococcus aureus. (*Biochem J*. **2011**, 435(1): 33-41)

2. Chemical Basis of Peptidoglycan Discrimination by PrkC, a Key Kinase Involved in Bacterial Resuscitation from Dormancy. (*JACS*. **2011**, 133(51): 20676–20679)

3. Structural Study of Binding of α-Mannosides to Mannan-Binding Lectins. (*Eur. J. Org. Chem*. **2012,** 2012(27): 5275–5281)

4. *Burkholderia cenocepacia* lectin A binding to heptoses from the bacterial lipopolysaccharide. (*Glycobiology*. **2012**, 10:1387-98)

5 NMR Spectroscopic Analysis Reveals Extensive Binding Interactions of Complex Xyloglucan Oligosaccharides with the *Cellvibrio japonicus* Glycoside Hydrolase Family 31 α-Xylosidase. (*Chemistry*. **2012**;18(42):13395-404)

6. NMR binding studies of the interaction between a monoclonal antibody and the LPS O-antigen of *Burkholderia anthina* (Submitted)

7. Molecular basis of chitin recognition and activation of rice chitin elicitor receptor (Manuscript in preparation)

Papers do not related to PhD thesis:

1. The structure of the carbohydrate backbone of the lipooligosaccharide from an alkaliphilic *Halomonas* sp. (*Carbohydr Res*. **2010**, 345(13): 1971-5)

Abstract

This PhD thesis work has been focused on the analysis of the structural requisites for recognition and binding between proteins and glycoconjugates, essential for the comprehension of mechanisms of paramount importance in chemistry, biology and biomedicine.
A large variety of techniques, such as crystallographic analysis, titration microcalorimetry (ITC), surface plasmon resonance (SPR) and fluorescence spectroscopy, allows the elucidation of molecular recognition events. In the last years, NMR-based methods have become among the most powerful and versatile techniques for detection and characterization of binding processes between a ligand and its receptor [1]. In detail, Saturation transfer difference NMR (STD NMR) is a technique that allows to demonstrate the interaction between species and to detect ligand regions more involved in the interaction process; in parallel, important information on the bioactive conformation of the substrate in the bound state can be obtained by tr-NOESY experiments.

Cell wall microbial glycoconjugates are very important players in the dynamic host-guest recognition that, in case of pathogens, is followed by the immune response and, in case of symbiosis, by its suppression. This PhD project proposes to use NMR techniques to achieve a deeper comprehension of the role of bacterial glycoconjugates in eukaryotic-pathogen and microbe-microbe interactions.

For what concern eukaryotic host and pathogen interactions, we have studied the binding between a monoclonal antibody, designated as 5D8, and the O-polysaccharide chain isolated from the lipopolysaccharide of a Gram negative bacterium, *Burkolderia*

[1] Meyer, B.; Peters, T. *Angew Chem Int Ed*, **2003**. 42, 864-890.

anthina, an opportunistic pathogen of cystic fibrosis (CF) patients [2,3]. Lipopolysaccharides (LPS) are endotoxins intercalated into the outer membrane of Gram negative organisms, essential for bacterial survival and defined as one of the most important virulence factor of bacterial infection in cystic fibrosis (CF). Among the LPS components, the O-chain is responsible of the activation of the adaptative immune response, that operates in the latter stage of the infection, providing a specific response and creating an immunological memory. Once activated by the innate immunity, leukocytes and B cells expose on their surfaces bacterial antigenic targets leading the production of effector T cells and specific antibodies, which are able to recognize, as non-self, bacterial components. The above interaction system, mAb 5D8 - O-chain, was analyzed in detail to go in-depth, at molecular level, into the mechanism of infection, key step toward the possible vaccines development. Thus, tr-NOESY and STD NMR experiments have been performed on the complex mAb 5D8 - O-chain from of the LPS of *Burkolderia anthina*, isolated from a CF patient from UK. Furthermore, in order to obtain more accurate information on structural requirements fundamental in the recognition process, we have also used, as carbohydrate ligands, the *ad hoc* synthesized trisaccharide and hexasaccharide O-antigen repeating units. The combined use of molecular dynamics simulations and NMR spectroscopy allowed to determine the binding epitope and the contribution to the binding of the sugar units composing the O-chain.

As mentioned above, innate immunity represents the first line of defence against invading microorganisms in vertebrates and the only line of defence in invertebrates and plants. This ancient

[2] Marchetti R,Lanzetta R, Nilsson I., Vogel C., Reed D.E., AuCoin D.P., Silipo A., Jimenez-Barbero J., Molinaro A., **2013**. Submitted.
[3] AuCoin, D. P.; Crump, R. B.; Thorkildson, P.; Nuti, D. E.; LiPuma J.J. and Kozel, T. R. *J.of Medical Microbiology* , **2010**. 59, 41-47.

system is exemplified by C-type lectins, which have developed a sophisticated ability to differentiate between infectious agents and self components. Lectins achieve this by recognizing invariant exposed structures adorning pathogens, so called Pathogen Associated Molecular Patterns (PAMPs), such as carbohydrates and acetylated compounds, among others. The precise dynamics of binding between these human pattern recognition receptors (PRRs) and their ligands are poorly understood, and so, among the mechanism of interaction analyzed in this PhD project, we undertook a study to elucidate the physical and biochemical binding properties of the prototypic lectin, mannose-binding lectin (MBL) [4]. Previous studies have revealed that this protein modulates immune response and is able to mediate the phagocytosis, representing a potential clinical application against glycosylated enveloped viruses, such Ebola and Marburg. Unsuitable for industrial-scale production, due to the costs and the difficulties inherent in its extremely complex quaternary structure, its potential for use in successful anti-infective therapy is limited; to produce better, less expensive therapeutic agents, less complex chimeric fusion proteins, such as L-FCN/MBL76, with similar ligand recognition and enhanced effector functions, have been developed. Performing a large variety of NMR experiments, we have studied the binding activity of both, recombinant human, rhMBL, and its chimer achieving information, at a molecular level, on the mechanism of the interaction useful for the design and optimization of new therapeutic proteins.

Besides being involved in the immune response, as a result of their interactions with glycoproteins, glycolipids and oligosaccharides, lectins play a large variety of biological functions in all living organisms, from viruses to bacteria, fungi, plants and animals. In

[4] Marchetti, R.; Lanzetta, R.; Michelow, C.J.; Molinaro, A. and Silipo A. *Eur. J. Org. Chem*, **2012**. 27, 5275–5281.

pathogenic micro-organisms, lectins are mostly involved in host recognition and tissue adhesion. Indeed, soluble lectins, are often secreted as virulence factors from opportunistic pathogens, such as *B. cenocepacia* that produces, among the others, a lectin called BclA with specificity for fucosylated and mannosylated glycoconjugates. Since the biological role of that lectin is unknown, in order to ascertain a correlation between its structure and function, a combination of techniques, crystallography, micro-calorimetry, fluorescence, electron microscopy and NMR spectroscopy, have been used [5]. In detail, we have studied by NMR the affinity of BclA for different ligands with a structure that remembers the glycoconjugates exposed on the external membrane of Gram negative organisms, such as the oligosaccharide portion (core) of the lipooligosaccharide of *Burkolderia cenocepacia*. By this approach, we have verified the hypothesis that this lectin could be involved in quorum sensing mechanisms defining its biological role.

Moving to plant-pathogen interactions, a combination of NMR techniques was used even in the analysis of the binding between glycoside hydrolases and xyloglucan oligosaccharides [6]. Xyloglucans are plant cell wall polysaccharides that can be degraded by glycoside hydrolases, ubiquitous among species, providing essential energy to organisms from microbes to higher animals. Thus, the study of this system of interaction is fundamental to diverse applications in medicine, food production, but in particular to biomass resource utilization. In the present project, we focused our attention on the main α-xylosidase, member of glycoside hydrolase family 31, Xyl31A, from the Gram

[5] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*, **2012**. 10, 1387-98.
[6] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A. *Chem Eur J*, **2012**. 15;18(42), 13395-404.

negative bacterium *Cellvibrio japonicus*. NMR techniques have been used to perform the epitope mapping on both the natural substrate and the product of the enzyme. Results obtained from STD NMR and tr-NOESY experiments, taken together with molecular dynamic simulations and docking, allowed a rationalization of previous kinetic data, underlying a peculiar architecture of enzyme active site. Indeed, we demonstrated that Xyl31A recognized the entire backbone of glycosidic units with four sites and also makes significant interactions, with other sub-sites, with branched xylose residues. This is due to the presence of an appended PA14 domain that plays a key role in the interaction and explains the strong affinity of the enzyme for long oligosaccharide substrates.

Other important components of fungal, cell wall, on which we have focused our attention, are chitooligosaccharides, potent elicitors of defense responses in a wide range of plant species [7,8]. Their perception and transduction, that plays a fundamental role in the establishment of plant resistance to pathogens, thought to be mediated by a chitin elicitor binding protein CEBiP, that is a key receptor in rice cells. It contains a transmembrane domain, two extracellular domains (known as Lysin Motif domains), that bind to the chitoligosaccharides, but lacks any intracellular region, so it could require additional factors for signal transduction, such as a chitin elicitor signaling, denoted as CERK1. In order to study the mechanism of plant immune reaction to fungal oligosaccharides, we performed many STD NMR experiments on chitooligosaccharides of different length and the extracellular domain of CEBiP. Combining NMR results with molecular biology experiments, it has been possible to define the ligand regions more

[7] Liu et al. *Science*, **2012**. 336(6085), 1160-4.
[8] Marchetti, R.; Silipo, A.; Berisio, R.; Shibuya, N.; Molinaro, A. Manuscript in preparation.

involved in the interaction and, at the same time, to detect the portion of the protein essential for the binding.
The above work is a part of a large project in which we are studying molecules possessing a similar primary structure and immunity-stimulating activities, involved in nodulation process and in the establishment of plant resistance to pathogen. Between these molecules, the peptidoglycan stands out; it is a signaling molecule involved not only in the establishment of bacterial resistance in plant but even in the interaction between bacteria. Indeed, it is known that PGN fragments, named muropeptides, are produced during bacterial growth and represent a molecular signal that growing conditions are promising, playing a key role in the germination of dormant bacterial spores. They are indeed recognized by the extracellular domain of bacterial eukaryotic-like Ser/Thr membrane kinase that, in response to muropeptides, allows the germination of spores [9]. By coupling protein mutagenesis and NMR techniques we succeeded in defining the structural requirements of both protein and ligand necessary for recognition and binding of muropeptides. In particular, an arginine residue belonging to the third PASTA (Penicillin and Serine Threonine Associated) domain played a key role in the interaction with *meso*-diaminopimelic moiety of PGN. It is worth to note that this aminoacid is a peculiar component of PGN of all Gram negative bacteria and of few Gram positive bacteria, such as bacilli; thus, our results explain also the specificity of the interaction between the protein and only DAP-type PGN.

In conclusion, the results obtained have allowed the analysis, at a molecular level, of fundamental mechanisms of interaction, thus allowing to clarify the molecular determinants involved in recognition and binding events. Therefore, the obtained results

[9] Squeglia, F.*; Marchetti, R.*; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; and Silipo, A. *equal contribution. *JACS*, **2011**. 133(51): 20676–20679.

represent an important platform pre-requisite that will permit a better comprehension of recognition events, key step for the design and optimization of new molecules with therapeutic aims.

SECTION I

INTRODUCTION

Chapter 1

The Glycoconjugates

1.1 Glycans and Glycoconjugates

Glycans (oligo- and polysaccharides) are one of the four fundamental classes of macromolecules that comprise living systems, along with nucleic acids, proteins and lipids. They play fundamental roles in the development and function of all living organisms, including critical roles in the areas of cell signaling, molecular recognition, immunity and inflammation. Therefore, they represent significant components involved in a wide variety of essential biological mechanisms; for example, it is known that in humans they determinate the phenotype of blood groups (Figure 1.1).

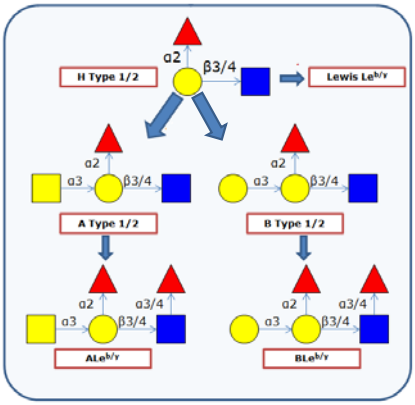


Fig.1.1: Histo-Blood groups antigens also called HBGs (A, B, H) and related antigens are complex carbohydrates present on the surface of red blood cells. In few cases, human red blood cells contain, on their surfaces, a core carbohydrate sequence known as “H antigen”; the more common ABO blood types derive from further modifications to this H carbohydrate chain.

In *vivo,* glycans are commonly found bound to specific locations on proteins or lipids modulating aspects of their biological activity.

These aggregates, known collectively as glycoconjugates, cover human, plant, fungal, bacterial and viral cell surfaces providing the first point of contact for any kind of host-guest interaction. In many type of cells, such as in humans and animals, a dense and complex array of glycans forms the glycocalix (Figure 1.2), while in other species, such as prokaryotes, plants or fungi, glycoconjugates comprise the cell wall (Figure 1.3).

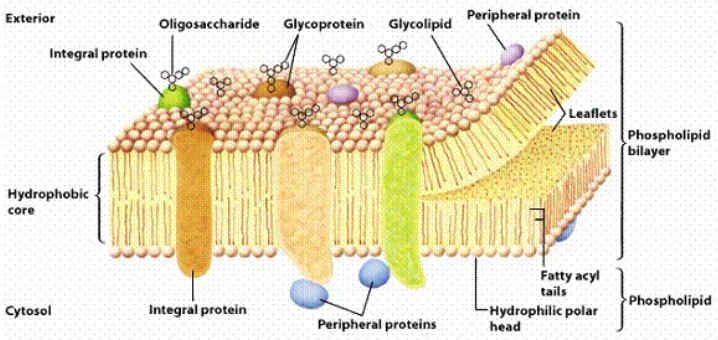


Fig. 1.2: Scheme of glycocalix (adapted from Bertozzi Research Group Homepage "Biomaterial and Interfacial Design").

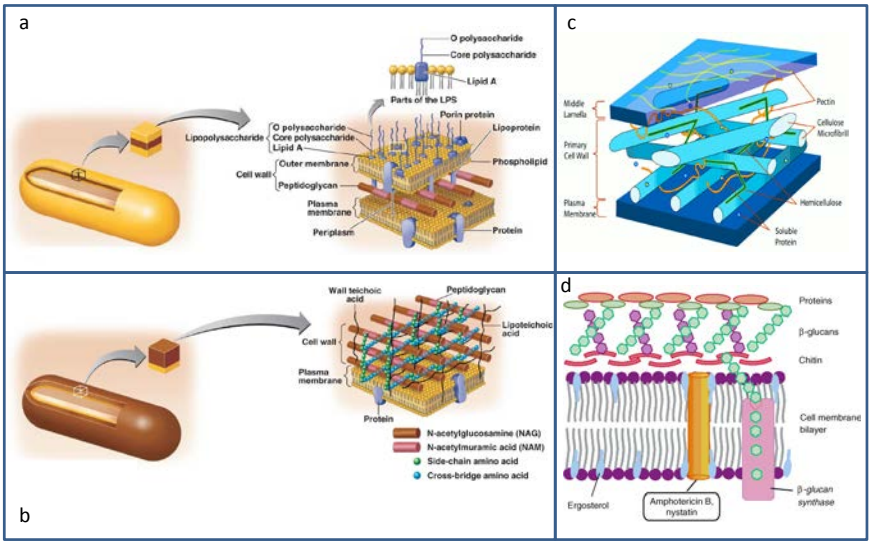


Fig. 1.3: Gram negative (a) and Gram positive (b) cell wall envelope. (adapted from Pearson Education, Inc.) Plant (c) and fungal (d) cell wall.

These complex structures hold high information content resulting from many ways glycans can be assembled from simple sugar building blocks which can be linked together at different sites, with different spatial orientations and creating both linear and branched polymers with a large range of shapes (Figure 1.4). This great diversity and flexibility are critical to many biological functions of glycoconjugates, especially in mediating cell-cell or cell-matrix interactions.

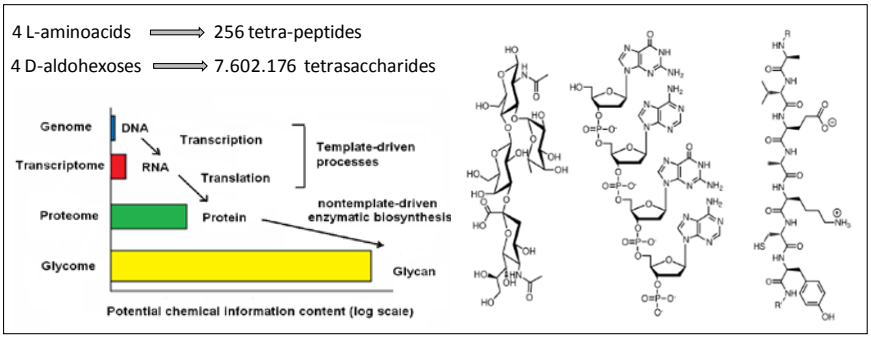


Fig. 1.4:. Schematic potential chemical information content of glycans respect to that of proteins and nucleic acids. Despite of sugars, protein and nucleic acid biopolymers are linear, and every building block is linked to the next through the same kind of connection. Between the combination of structural diversity and different possible connection sites, the complexity of glycans increases rapidly; therefore, even a small number of monosaccharide units can provide a large number of different oligosaccharides.

A brief outline of the structures of the glycans and glycoconjugates investigated in this thesis is reported in the following sub-paragraphs.

1.1.1 Chitin structure

After cellulose, chitin is the second most abundant natural biopolymer and the main structural component of invertebrate exoskeletons as well as of cell walls of fungi. It is a polysaccharide composed of β-(1,4) linked units of the amino sugar N-acetyl-

glucosamine, which can be arranged in antiparallel (α-chitin) or parallel (β-chitin) manner (Figure 1.5).

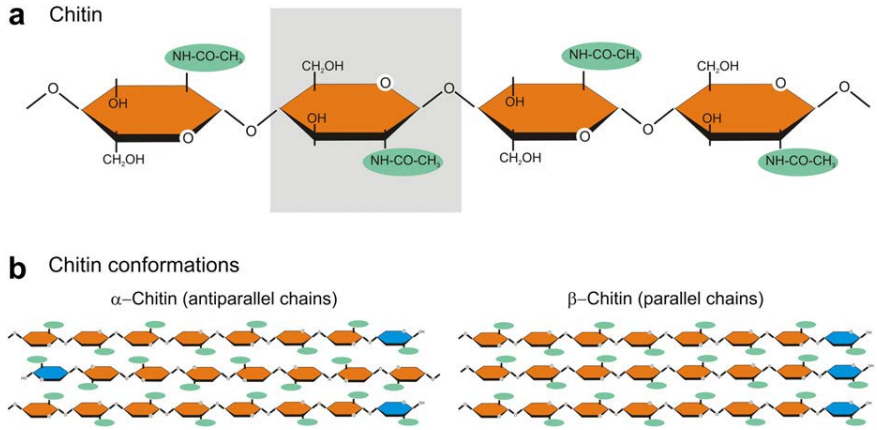


Fig. 1.5: Chitin structure (a). Arrangement of the chains in α-chitin and β-chitin (b).

This component is essential for many cellular functions but it serves, particularly, as a fibrous strengthening element responsible for cell wall rigidity.

1.1.2 Xyloglucans structure and nomenclature

Xyloglucans are members of a group of plant cell wall polysaccharides, including xylan, gluco- and arabino-xylan, mannan, gluco-and galacto-mannan, not solubilized by water, typically referred to as hemicelluloses. Due to the similarity of the structure and conformational homology with the cellulose (composed by β-(1,4) linked units of D-glucose), hemicelluloses establish strong non-covalent association with cellulose microfibrils, regulating cell wall properties.

Xyloglucans are quantitatively predominant hemicellulosic polysaccharides in the primary walls of dicots and non-graminaceous monocots. They possess a backbone composed of β-

(1,4)-D-Glcp residues highly substituted at position 6 by mono-, di-, or triglycosyl side chains [1] (Figure 1.6a).

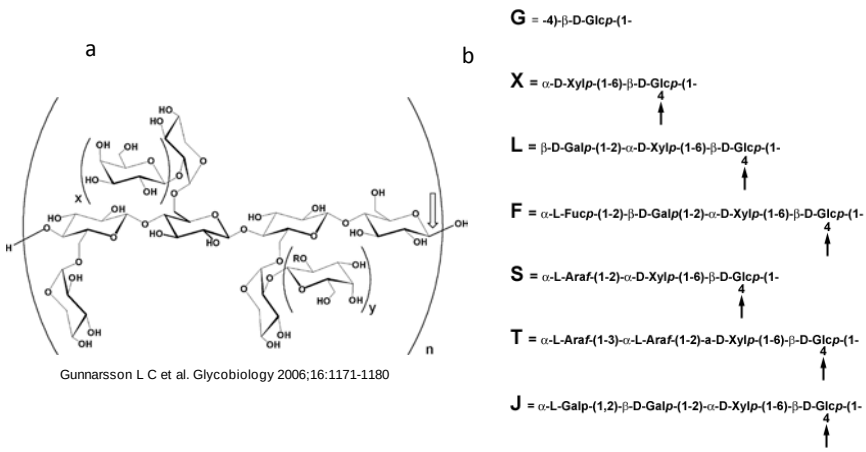


Fig. 1.6: General xyloglucan structure (a) and nomenclature (b).

In order to simplify the nomenclature of these macromolecules, single letters are used to represent different branched or unbranched domains [2] (Figure 1.6b). Depending on their structure, xyloglucans can be classified as XXXG-type, composed by three consecutive backbone residues substituted by Xylp and a fourth unbranched glucosyl residue, or XXGG-type, constituted by two branched and two unbranched domains.
XXXG-type xyloglucans can be digested by many endo-active enzymes, generating a series of well-defined oligosaccharide (GXXG), providing essential energy to organisms from microbes to higher animals.

1.1.3 Lipopolysaccharide structure

Lipopolysaccharides are glycoconjugates intercalated into the outer membrane of Gram negative bacteria. They are heat-stable amphiphilic molecules, characterized by a common structural motif

in which three different domains can be distinguished (Figure 1.7), namely lipid A, core oligosaccharide and O-polisaccharide chain.

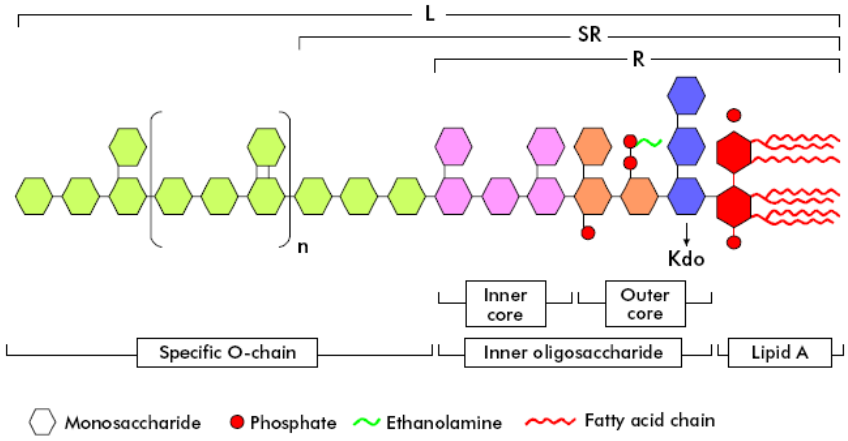


Fig. 1.7: Scheme of the general chemical structure of bacterial lipopolysaccharide: Lipid A, internal oligosaccharide and specific O-chain. Rough type LPS (R) do not contain specific O-chain, semi-rough type (SR) just contain a repetitive unit of specific O-chain, and smooth type (S) contain two or more repetitive units of specific O-chain.

The lipid A is the hydrophobic portion of the LPS that enables the anchorage of the entire macromolecule to the bacterial outer membrane and represents the endotoxic principle of LPS [3].

The inner core, proximal to lipid A, consists of characteristic monosaccharides, such as heptoses [4] (L-glycero-D-manno heptose or D-glycero-D-manno heptose) and Kdo (3-deoxy-D-*manno*-octulosonic acid), a monose marker of all Gram negative bacteria that connects the core oligosaccharide to the lipid A backbone. On the other hand, the outer core is usually constituted by a few glycosidic residues: neutral carbohydrates, uronic acids and aminosugars.

The O-polysaccharide chain is the hydrophilic component of LPS and is usually composed by a regular polysaccharide [5], homopolymeric or heteropolymeric, linear or branched, composed by up to 50 identical repeating units consisting of up to eight

different sugars. It is highly variable and heterogeneous even within the same bacterial species, contains different kinds of monosaccharides and can be decorated with non-carbohydrate substituents thus increasing the structural heterogeneity.

1.1.4 Peptidoglycan structure

The peptidoglycan is an essential and unique bacterial glycoconjugate that provides rigidity and structure to the cell wall of bacteria. It is a large and complex heteropolymer formed by a network of glycan strands of alternating β-(1,4)-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) possessing interconnected oligopeptide stems. These are composed of alternating L and D amino acids [6], attached to the lactyl group of the muramic acid. During cell-wall biosynthesis, the oligopeptide stem of one glycan strand is cross-linked by the transfer of an acyl group to an amine on the stem of an adjacent strand [7].

Although PGN composes the cell envelope of all bacteria, its location, amount and specific composition vary among species. In monoderm (single membrane) bacteria, such as *Staphylococcus aureus* (a Gram-positive-staining bacterium), the PGN is found as a thick exposing layer; whereas in Gram negative microorganisms, it constitutes a thin layer overlaid by an outer membrane surface (Figure 1.3a,b). According to the residue at position three of the peptide stem, it is possible to classify the PGN in two major categories: L-lysine type (Lys-type) and *meso*-diaminopimelic acid type (DAP-type) PGN [8]. In DAP-type PGN, characteristic of all Gram negative bacteria and Gram positive bacilli, such as *Bacillus anthracis*, the peptide stems are directly linked. On the other hand, in Lys-type PGN, present in many Gram positive organisms, peptide stems are connected by a peptaglycine bridge (Figure 1.8).

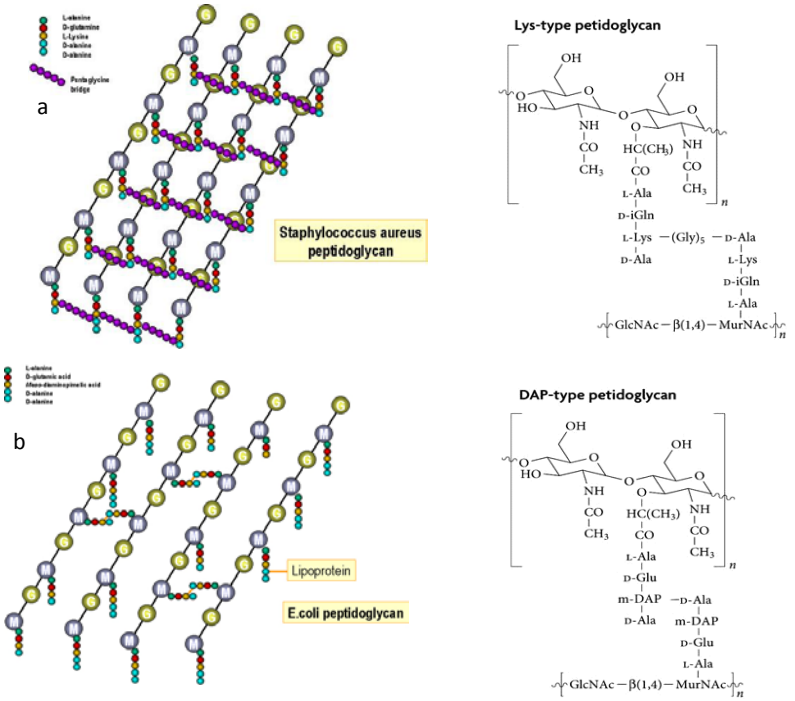


Fig. 1.8: Schematic structure of Lys-type PGN (a) and DAP-type PGN (b).

1.2 Communication: glycoconjugates and the bacterial social life

The complex glycans and glycoconjugates which decorate the surface of microbial cells are essential not only for the bacterial survival but also in cell-cell signaling and communication, representing key players of microbial social life. The following examples underline the huge importance and role of glycoconjugates in a large variety of aspects of the bacterial social life.

1.2.1 Glycoconjugates as key mediators in microbe-microbe interaction

Bacteria rarely exist as single-species planktonic forms, the majority are found thriving in complex polymicrobial biofilm communities held together by a self-produced extracellular matrix [9]. Exopolysaccharides, secreted by microbial cells, constitute a significant component of biofilm contributing directly to its properties [10]. For instance, *P. aeruginosa* is a Gram-negative pathogen that makes biofilms by producing three distinct exopolysaccharides: alginate, PEL, and PSL, depending of the environment and the strain studied. Alginate is produced by mucoid strains of *P. aeruginosa*, often isolated from lungs of cystic fibrosis patients; the pel gene cluster encodes a glucose-rich polymer (PEL), found in most of the strains, while others exhibit the PSL galactose and mannose-rich polysaccharide [11]. Moreover, the ability of exopolysaccharides to interact with other macromolecules, polysaccharides and cells, as well as with ions, provide a multitude of microenvironments within any biofilm that seems to be important for the protection of the microbial cell. There are numerous benefits that a bacterial community might obtain from the formation of biofilms; it contribute to bacterial persistence in a variety of disease, facilitating chronic infections and reducing the efficacy of antimicrobial therapy, playing essential roles in microbial social life.

Glycoconjugates intercalated into the outer membrane of Gram negative organisms, such as lipopolysaccharides, greatly contributing to the bacterial structural integrity and to the protection of the cell envelope, are directly involved in bacterial colony aggregation. Since they are exposed toward the external environment, LPSs are responsible of many biological interactions of the bacteria. In particular, they regulate not only (pathogenic, symbiotic) strategies for adhesion to host tissues, but can also be recognized by specific bacterial lectins playing a crucial role in the

dialogue between microbes (Section II, Chapter 4). Different studies revealed that soluble bacterial lectins, such as PAIIL from *Pseudomonas aeruginosa* [12], are involved in biofilm formation and cohesion between microbes since they have the ability to bind to glycoconjugates present at the surface of bacteria and parasites [13-14]. Moreover, the structural basis for the interaction of C-type lectin, surfactant D, and heptose residues present in bacterial lipopolysaccharide has been demonstrated [15].

Important mediators of intracellular communication mechanisms are fragments of bacterial cell wall (PGN), known as muropeptides, released into the extracellular milieu as metabolic byproducts of bacterial growth. Since the peptidoglycan is a well-conserved molecule, PGN derived muropeptides, released during the turn-over, can behave as key signaling molecules in the resuscitation pathway of dormant spores [16]. Gram positive bacteria, such as *Bacillus* and *Clostridium* spp., can exist in metabolically inactive states, such as dormant endospores, that allow them to survive severe physical and chemical conditions that are not conducive for growth [17-19]. However, the endospores are able to monitor the extracellular environment to identify growth-promoting conditions allowing them to reinitiate metabolism [20]. The presence of muropeptides into the extracellular milieu represents a clear interspecies signal of favorable conditions that support bacterial growth. In order to allow the exit from dormancy in response to the secreted PGN fragments, it is necessary the intervention of the members of a family of serine/threonine protein kinases (STPKs), characteristic only of Gram positive bacteria, which are able to specifically recognize DAP-type PGN fragments (Section II, Chapter 3).

1.2.2 Glycoconjugates as key mediators in host-microbe interactions

The nature of the relationship between microbes and hosts spans the spectrum from mutually beneficial (mutualistic symbiosis), to benefiting the microbe without necessarily harming the host (commensal), to benefiting the microbe at the expense of the host (pathogenic). One of the key factors determining the establishment of any particular microbe–host interaction along these potential outcomes is the pattern of glycoconjugates exposed on host-cell surfaces [21].

Symbiotic host-microbe interactions

Eukaryotes coexist from millions of years in contact with a great variety of microorganisms.
In humans, the number of microbes, comprising the endogenous bacterial flora, exceeds the total number of somatic and germ cells by at least an order of magnitude (10^{13} human cells vs 10^{14} bacterial cells). The mammalian gut is particularly populated by different bacterial species that colonize specific parts of gastroinstestinal tract. A non-pathogenic, mutually beneficial host-microbe relationship is allowed by the presence of intestinal glycoconjugates that play a key role in intercellular communication. *Bacteroides thetaiotaomicron* [22] provides an example in which host glycans are involved in the establishment of an advantageous interaction with a member of the colonic microflora. This anaerobic bacterium has evolved a mechanism to increase the production of its own nutrients (fucosylated glycans) from the host and, at the same time, it is able to express, when it is necessary, polysaccharide binding proteins in order to forage and consume host hexose sugars. This also benefits the host because resource competition provides an additional barrier to colonization by potential pathogens. In addition, products of microbial fermentation are important energy sources for the host.

Plants in their natural habitats are surrounded by a large number of microorganisms that can directly interact with them in a mutually beneficial manner. One of the most familiar example is represented by the symbiosis established, under condition of nitrogen limitation, between legominous and some species of Rizhobia. This interaction is based on the communication between bacteria and plant roots through glycoconjugates, known as NOD factors, produced in response to chemical signals released from plants (flavonoids). The binding between NOD factors and receptor on plant roots cells lead to the formation of specific organs (nodules) in which bacteria carry out the fixing of atmospheric nitrogen in ammonia, useful as plant nitrogen source [23]. Moreover, plant defense metabolism can be enhanced by certain rhizo-bacteria, through the production of antagonistic compounds against phytopathogens, inducing a high resistance against microbial or fungal invasion [24].

Pathogenic host-microbe interactions

The interactions of microbial pathogens with their hosts are also influenced, to an important degree, by the repertoire of glycoconjugates and glycan-binding receptors that each expresses (Section III). They play an essential role in all stages of infection, from early colonization of host epithelial surfaces, to the induction of immune response and inflammation.

Microbial molecules, such as lipopolysaccharide and peptidoglycan, expressed and secreted by pathogens, are known as virulence factors since they are responsible of causing disease in the host; indeed, these important glycoconjugates enable bacterial adhesion to host tissue, immunoevasion and immunosuppresion. Actually, due to their nature of highly conserved structures, not found in the host and invariant among entire classes of pathogens, they act as PAMPs (Pathogen Associated Molecular Patterns). The ability to detect these structures, distinguishable from “self”

components, is essential for the host survival and is mediated by the immune response. This ancient system is divided into two different and complex mechanisms: the innate and the adaptive immune response (Figure 1.9).

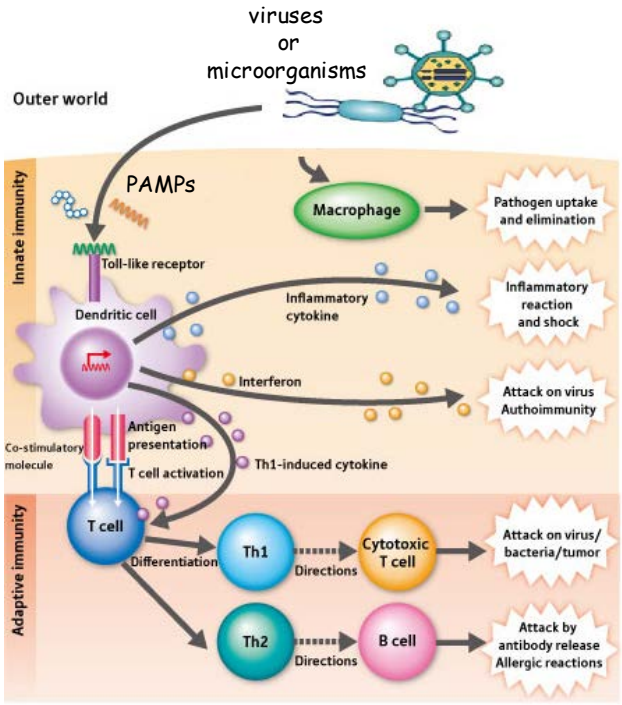


Fig. 1.9: Scheme of innate and adaptative immune response.

The innate immune system is well conserved from invertebrates through higher mammals and it represents the first line of defence against pathogens. This process is stimulated by the recognition of peculiar bacterial molecules (PAMPs) by specific host glycoconjugates, known as PRRs (Pathogen recognition receptors). Two groups of PRR, that are the C-type lectins (CTL) and C-type lectin receptors (CTLR), and the Toll-like receptors (TLR), play a crucial role in the activation of immune response with the resulting recruitment of phagocytes and leukocytic cells that exert their bactericidal action. The CTLs can recognize a wide variety of bacteria, fungi and parasites and are important in phagocytosis and

endocytosis. TLRs are expressed on the surfaces of a variety of cells, including dendritic cells and macrophages, and they play a major role in innate immunity. The interaction of TLRs with PAMPs initiates a cascade of events activating the adaptative immunity; it operates in the latter stage of the infection providing a specific response and creating an immunological memory. Once activated by the innate immunity, leukocytes and B cells expose on their surfaces bacterial antigenic targets (glycans and glycoconjugates) leading the production of effector T cells and specific antibodies.

In conclusion, complex carbohydrate structures, exposed on the surface of host cells as well as produced by pathogens, including bacteria, viruses or fungi, act as key mediators in the wide range of host-pathogen interactions.

References

[1] York et al. *Carbohydr Res*, **1990**. 200, 9-31.
[2] Fry, S. C.et al. *Physiol.Plant*, **1993**. 89, 1–3.
[3] Wollenweber, H.W.; Schlecht, S.; Luderiz, O.; Rietschel, E.T. *Eur. J.Biochem*, **1983**. 130, 167-171
[4] De Soyza, A.; Silipo, A.; Lanzetta, R.; Govan, J.R.; Molinaro, A. *Innate Immun.*, **2008**. 14, 127-144.
[5] Konig, H.; Herald, C.; Varma, A.; *Prokaryotic cell wall*, chapter 4 : 133-154
[6] Dziarski, R. *CMLS*, **2003**. 60, 1793-1804.
[7] Cloud-Hansen, A.K. et al. *Nature Reviews*, **2006**. 4, 710-716.
[8] Guan, R. and Mariuzza, A.R. *TRENDS in Microbiology*, **2007**. 15: 126-134.
[9] Peters, B.M.; Jabra-Rizk, M.A.; O'May, G.A.J.; Costerton, W. and Shirtliff, M.E. *Clin. Microbiol. Rev.*, **2012**. 25, 193. DOI: 10.1128/CMR.00013-11.
[10] Framework, I.W. *Microbiology*, **2001**. 147, 3–9 .
[11] López, D.; Vlamakis, H.; Kolter, R. Cold Spring Harb Perspect Biol., **2010** (7): a000398. doi: 10.1101/cshperspect.a000398.
[12] Tielker et al. Microbiology, **2005.** 151, 1313-23.
[13] Imberty et al. *Mol Microbiol*, **2004.** 52(3):691-700.
[14] Imberty et al. *Microbes Infect*, **2004.** 6(2), 221-8.
[15] Wang, H.; Head, J.; Kosma, P.; Brade, H.; Muller-Loennies, S; Sheikh, S.; McDonald, B.; Smith, K.; Cafarella, T.; Seaton, B.; and Crouch, E. *Biochemistry*, **2008**. 47, 710-720.
[16] Boudreau, M.A.; Fisher, J.F. and Mobashery, S. *Biochemistry*, **2012**. 51, 2974−2990.
[17] Setlow, P. *Cell*, **2008**. 135, 410-412.
[18] Setlow, P. *Current Opinion in Microbiology*, **2003**. 6: 550-556.
[19] Dworkin, J.; Shah I.M. *Nature Reviews Microbiology*, **2010**. 8, 890-896.
[20] Shah I.M. et al. *Cell*, **2008**. 135, 486-496.
[21] Varki, A.; Cummings, R.D.; Esko, J:D. et al Essentials of *Glycobiology. 2nd edition., Cold Spring Harbor Laboratory Press*; **2009**. Cap 39
[22] Hooper, L.V.; Gordon, J.I. *Glycobiology*, **2001** 11(2):1R-10R.

[23] Ogutcu, H.; Adiguzel, A.; Gulluce, M.; Karadayi, M.; Sahin, F. *Romanian Biotechnological Letters*, **2009**. 14, 4294-4300.
[24] Shoresh, M. et al. *Annu. Rev. Phytopathol*., **2010**. 48, 21–43.

Chapter 2

NMR methods for protein-carbohydrate interactions studies

2.1 NMR techniques for free-state structure determination

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful tool to investigate structural features of macromolecules in solution and, particularly, to gain insights at both atomic and molecular level in carbohydrates structure, conformation, geometry and dynamic [1-3].

The present thesis will not discuss the basis and the fundamentals of NMR, whereas it will be focused on the methodologies that yield the most complete picture of a carbohydrate, in terms of oligo- or polysaccharide structure and behaviour in solution.

In order to assign all the spin systems, determine the nature and the location also of non-carbohydratic substituents, thus characterizing the entire sugar structure, a combination of homo- and hetero-nuclear 1D and 2D NMR experiments has to be performed. As a first step, the monosaccharide composition can be achieved by the evaluation of NMR chemical shifts and coupling constants. The anomeric proton resonances are found in the shift range 5.8-4.4 ppm, the integration of corresponding signals provides an initial estimate on the number of different monosaccharide residues present; the anomeric carbon resonances achieved by ^{13}C or ^{13}C-^{1}H HSQC spectra can confirm the above evaluation.

Additionally, the values of coupling constants $^{1}J_{C1,H1}$ and $^{3}J_{H1,H2}$ are diagnostic of the anomeric orientation of pyranose rings. In sugar with *gluco* or *galacto* configuration (H-2 axial), a $^{3}J_{H1,H2}$ around 8 Hz is indicative of a β-anomer, whereas below 3 Hz of an α-anomer. On the other hand, *manno* configured sugars (H-2 equatorial) show both $^{3}J_{H1,H2}$ below 3 Hz. The values of $^{1}J_{C1,H1}$ are also indicative of the anomeric configuration, indeed, a $^{1}J_{C1,H1}$

below 165Hz indicates a β-anomer whereas above 170 Hz indicates the presence an α-anomer.
The ring proton signals resonate in the range 4.4-2.8 ppm.
The key values of proton resonances in pyranose rings are reported below:

- 8.5 – 7.5 ppm : ammide resonances;
- 5.5 – 4.2 ppm : anomeric protons;
- 4.5 – 2.8 ppm : sugar ring protons;
- 2.6 – 1.8 ppm : α-methylene protons of deoxy sugars;
- 2.2 – 0.8 ppm : β,γ,…ω protons of methyl and distal methylene groups.

While the principal values of ^{13}C resonances are:

- 160 – 180 ppm : carbonyl carbons ;
- 105 - 95 ppm : anomeric carbons;
- 80 – 60 ppm : sugar ring carbons;
- 60 – 45 ppm : nitrogen bearing carbon signals;
- ~ 30 ppm : aliphatic methylene carbons of deoxy sugars;
- 20 – 17 ppm : methyl carbons of deoxy sugars, acetyl groups.

The identification of the sugar moieties is possible by determining the number of the different spin systems corresponding to each sugar moiety, that can be achieved by analysis of COSY and TOCSY spectra, in turn used to attribute the carbon resonances in the HSQC spectrum.
NOESY and ROESY spectra (cfr. 2.3.1) are very useful both to confirm the *intra*-residual assignment and the anomeric configuration. Actually, in β-configured sugars, H-1 gives *intra*-residue NOE correlation with H-3 and H-5, while in α-configured only with H-2. In addition, *inter*-residual contacts, together with the long range correlations present in HMBC experiments, are crucial

to understand the sugar sequence in the saccharide chain (Figure 2.1)

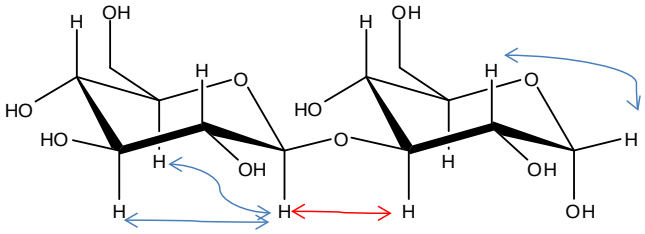


Fig. 2.1: *intra*-(blu) and *inter*-(red)residual NOE contacts.

Furthermore, in the HSQC spectrum, the down-field shift of carbon resonances (glycosidation shift) allows to locate the positions of glycosidation.

The attachment of a non-carbohydrate substituent, such as methyl, acetyl or phosphate residue, often affects proton and carbon resonances where the group is located. Normally, downfield shifts of ~0.5-0.2 ppm are observed for protons and higher $\Delta\delta$ values for ^{13}C. For instance, the acylation sites can be easily determined analyzing the downfield shift of ring proton signals (acylation shift). It is worth to note that such appended groups may also contain other NMR-active nuclei, which may give rise to additional splittings due to couplings (e.g., ^{31}P-^{1}H long-range couplings). The use of other homo- or heteronuclear correlations may help in the determination of their position.

2.2 NMR techniques in protein-ligand interactions

NMR techniques are crucial in the study of interaction between different species, such as proteins-glycoconjugates recognition events. As already mentioned in Chapter 1, major roles of glycoconjugates in health and disease are mediated by binding proteins that decode the information content of the glycome through their recognition of glycans as ligands [4-7]. Thus, understanding the basis of the complex interplay between microbial

glycoconjugates and their glycan receptors, at a molecular level, is fundamental for the possible development of novel therapeutics and diagnostics.
A large variety of techniques, such as crystallographic analysis, titration microcalorimetry (ITC), surface plasmon resonance (SPR) and fluorescence spectroscopy, allow the elucidation of molecular recognition events; however, in the last years, a huge amount has been published on NMR investigations of protein-ligand interactions, including numerous reviews [8-14]. Thus, NMR-based methods have become among the most powerful and versatile techniques for detection and characterization of binding processes between a small substrate and its large receptor. These experiments identify binding events either by looking at the resonance signals of the ligand or of the protein, delivering a picture of ligand-receptor complexes.

2.2.1 Binding equilibrium and chemical exchange

The binding of small molecule ligand (L) to large receptor (P) can be described, in its simplest form, by a bimolecular association reaction with second order kinetics:

$$[P] + [L] \underset{K_{off}}{\overset{k_{on}}{\rightleftharpoons}} [PL] \qquad \text{[Equation 2.1]}$$

where k_{on} and k_{off} represent the association and dissociation rate constants (measured in $M^{-1}s^{-1}$), respectively. The unimolecular k_{off} is inversely proportional to the mean lifetime τ_B of the protein-ligand complex (PL); the bimolecular k_{on} measures the probability of a productive encounter between free receptor and ligand. It is often assumed to be controlled by the diffusion and can vary between 10^7 and 10^9, being slow if large conformational rearrangements of either ligand or receptor occur upon binding.

The affinity of protein-ligand interaction can be described by the dissociation constant K_D, that is given by the Equation 2.2:

$$K_D = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}} \qquad \text{[Equation 2.2]}$$

Ligands with weaker affinity have larger K_D and thus require the addition on more substrate to saturate the receptor binding site. Indeed, when [L] « K_D, the fractional occupation of the receptor binding site by ligand is proportional to [L]; when [L] = K_D, the protein is half-saturated; if [L] » K_D the receptor is completely saturated and, on average, each bound ligand exchanges with another every $1/K_{off}$.
At a given concentration of total ligand [L_0] and protein [P_0], and a given value of K_D, the actual concentration of the complex PL can be calculated by Equation 2.3 and the fraction of ligand and protein bound can be derived (Equation 2.4 and 2.5):

$$[PL] = \frac{1}{2}\left(K_D + [P]_0 + [L]_0\right) - \sqrt{\frac{1}{4}(K_D + [P]_0 + [L]_0 2 - [L]_0[P]_0)}$$

[Equation 2.3]

$$f_{LB} = {}^{[PL]}\!/\!_{[L]_0} \qquad f_{PL} = {}^{[PL]}\!/\!_{[P]_0}$$

[Equation 2.4 and 2.5]

The binding process is a dynamic chemical exchange between free and bound state in which different conditions are possible depending on k_{off} values. On chemical shift time scale, the exchange can be fast (koff » ω), intermediate (koff ~ ω), or slow (koff « ω), where $\Delta\omega = 2\pi\Delta\nu$, that is the chemical shift difference between the free and bound states, measured in Hz.

It is worth to note that, in NMR, chemical exchange refers to any process in which a nucleus exchanges between two or more environments (for example free and bound state) in which its NMR parameters (chemical shift, scalar coupling, dipolar coupling, relaxation rate) differ; their observation forms the basis for all NMR screening experiments.
In the free state, both receptor and ligand retain their intrinsic chemical shifts, relaxation rates, diffusion coefficients, etc.. When the ligand binds to the protein, the mutual binding affinity drives an exchange process that modulates the NMR parameters of both molecules. Upon binding, the small ligand adopts characteristics of the much larger receptor, while the protein site microenvironments are perturbed by the interaction with the substrate.
The HMM (Hahn, Maxwell, and McConnell [15-17]) equations provide a complete description of the exchange phenomena under all exchange regimes, but it beyond the scope of this Thesis in which it has been considered only the case of fast exchange.

2.2.2 Fast exchange approximation

In NMR screening practice, the HMM equations are almost never solved and fast exchange is simply assumed.

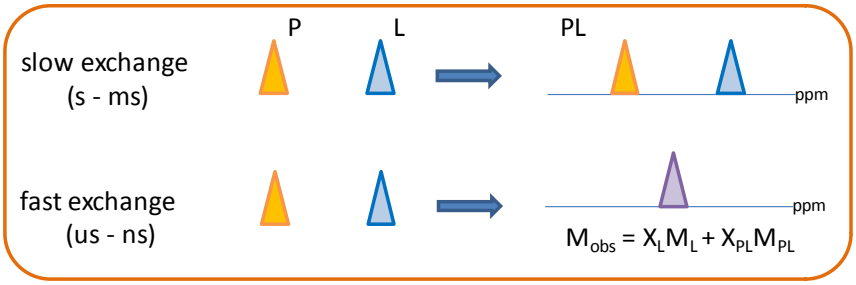


Fig. 2.2: Schematic representation of slow and fast exchange.

For a system with a long lifetime (slow exchange), the frequency of free and bound state is sampled before exchange so solved and integrated signals may be expected for each state (Figure 2.2). In

practice, a situation of this type is very rare to achieve because of the extreme difficulty to detect and integrate μM signals into a complex and crowded spectrum. More generally, systems assumed to be in fast exchange are studied. For these systems the observed NMR response of a ligand (or protein) is the mole fractions weighted average of the NMR parameters of the free and bound states:

$$M_{obs} = X_L M_L + X_{PL} M_{PL} \qquad \text{[Equation 2.6]}$$

where M_{obs} is any NMR observable (chemical shift, scalar coupling, dipolar coupling, relaxation rate) characteristic of the equilibrium system; X_L and X_{PL} are the mole fractions of the free and bound ligand; and M_L and M_{PL} are the NMR parameters of the ligand in the free and bound state, respectively. As mentioned above, this equation is applicable if the exchange is fast on the relaxation and chemical shift timescales of the system. This assumption can be made since typically the analyzed binding processes are characterized by $K_D \geq 100$ μM and the slowest exchange rate k_{ex} values lies in the range $1000 < k_{ex} < 100\,000\ s^{-1}$. Ligand-based NMR screening methods are primarily ^{1}H-based and, consequently, the exchange constant exceeds most differences in intrinsic ^{1}H relaxation rates and rotating frame precession frequencies, thus ensuring that the fast exchange assumption is valid. Under this conditions, the NMR observable M_{obs}, grown on the ligand molecule in the bound state, is transferred to the free ligand and viceversa; this transfer process is mediated by a rapid exchange between M_{PL} and M_L during the time necessary to measure M_{obs}. The individual populations X_L and X_{PL} depend on the K_D value and on the concentrations of L and P. From Equation 6 it is evident that the resonances of the ligand in a ligand-target mixture can only be observed if the ligand is in excess with respect to the target.

2.3 Ligand's perspective

Since chemical exchange upon binding modulates the NMR parameters of both receptor and ligand, the screening may proceed by ligand- or receptor-based methods. In this Thesis the focus will be pointed out particularly on NMR methodologies based on the ligand point of view. These are particularly useful in the medium–low affinity range, do not require labeled protein since only the ligand NMR signals are detected, and only a small amount of receptor is required; therefore, they have been adopted to detect a wide range of different systems of interaction.
Among NMR ligand-based methods, transferred-NOE (trNOE) and Saturation Transfer Difference (STD) NMR are the most used in order to gain insight, at molecular level, in various binding mechanisms [8-14].

2.3.1 NOE and transferred NOE

It is well established that NOE effects are extremely useful in the determination of 3D structure of molecules in solution [18]; such information can be obtained because the NOE depends on *inter*-nuclear separation. NOE effects are due to homonuclear dipolar interactions that drop of very fast with distance, so they can only occur with a distance of <5-6 Å (500-600 pm). The intensity and the sign of NOE contacts are also affected by the characteristics of molecular tumbling in solution. The variation in the maximum theoretical homonuclear steady-state NOE in a two-spin system as a function of molecular tumbling rates (defined by the dimensionless parameter $\omega_0\tau_c$), is reported in Figure 2.3.

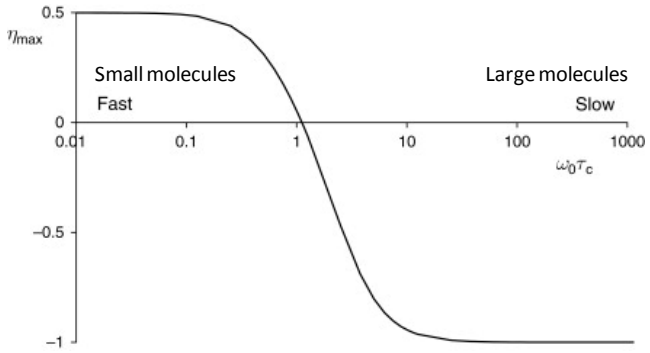


Fig 2.3: Dependence between the maximum NOE intensity, η_{max}, and $\omega_0\tau_c$, where ω is the resonance frequency and τ_c the molecular correlation time. The region of fast motion is the extreme narrowing limit and that of slow motion is the spin-diffusion limit.

Molecules with a short correlation time, that corresponds to a fast random rotation in solution, exhibit high positive NOE contacts, reaching at the most the value 0.5; while molecules with a long τ_c, that means a rather sluggish rotation, undergo negative NOE with a maximum value of -1.

It is worth to note that the correlation time is determined by different factors: mainly, the molecular weight, MW. The larger a molecule, the slower its re-orientation, therefore the longer its τ_c. As a rule of thumb, the correlation time of a molecule can be estimated from the molecular weight, in detail, τ_c [ps] » 0.5 MW [Da], for example, a MW of 1000 Da corresponds to ca. 0.5 ns. However, the correlation time is also influenced by other parameters, such as temperature (the higher the temperature, the shorter τ_c), solvent viscosity (the more viscous, the longer τ_c), aggregation in solution etc..

At a given value of the parameter $\omega_0\tau_c$ (1.115), the NOE effect disappears and no NOE signals are visible although the nuclei are dipolar coupled. A modification of NOE experiment (ROE), with a different dependence to $\omega_0\tau_c$, permits to override this inconvenience (Figure 2.4).

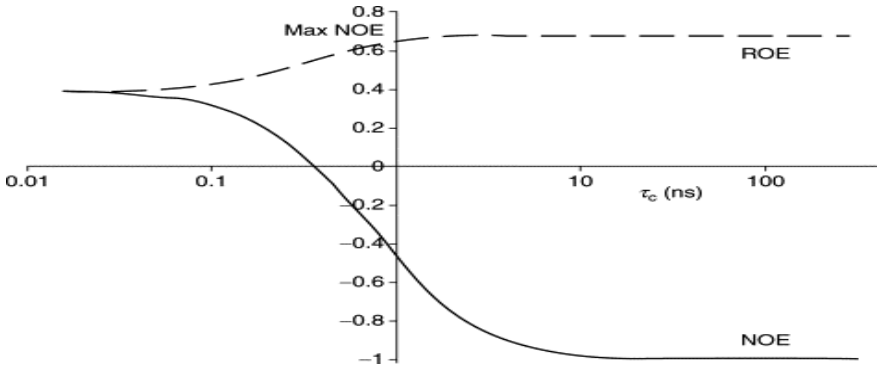


Fig. 2.4: Dependence of NOE and ROE to log $\omega_0\tau_c$.

Even if the intensity of ROE signals is lower respect to that of NOE effects, their sign is always positive and so observable.
Any quantitative interpretation of steady state NOEs requires knowledge of the arrangement of all protons relative to each other, therefore transient or kinetic NOEs have to be determined. Typically, 2D NOESY experiments are used to detect transient NOEs. The following diagram shows the pulse sequence of the NOESY experiment in which high-frequency pulses are used to generate a non-equilibrium state that during mixing time (τ_{mix}) returns to equilibrium by relaxation.

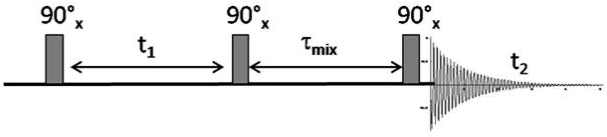


Fig. 2.5: Pulse sequence of 2D NOESY experiment.

During the mixing time, NOEs build up to a maximum value and then decrease to zero because of T_1 relaxation.
When the pulse sequence reported in Figure 2.5 is applied to a protein-ligand system in dynamic exchange in which ligand is

present in excess, we talk about transferred NOE (trNOE) experiment.

Transferred NOE (trNOE) experiments provide key information on the ligands' binding mode in the natural environment and allows to deduce information on the structural requirements of the ligand in the bound state (bioactive conformation).

Transferred NOE is a transient NOE experiment in which *intra-* and *inter*molecular trNOEs build up during the mixing time. Although both effects can be used for the detection of binding activity of ligands, we have focused our attention on the intramolecular transferred NOEs, which are well established as unique sources of structural information on bound ligand, and additionally do not require protein assignment and isotope enrichment.

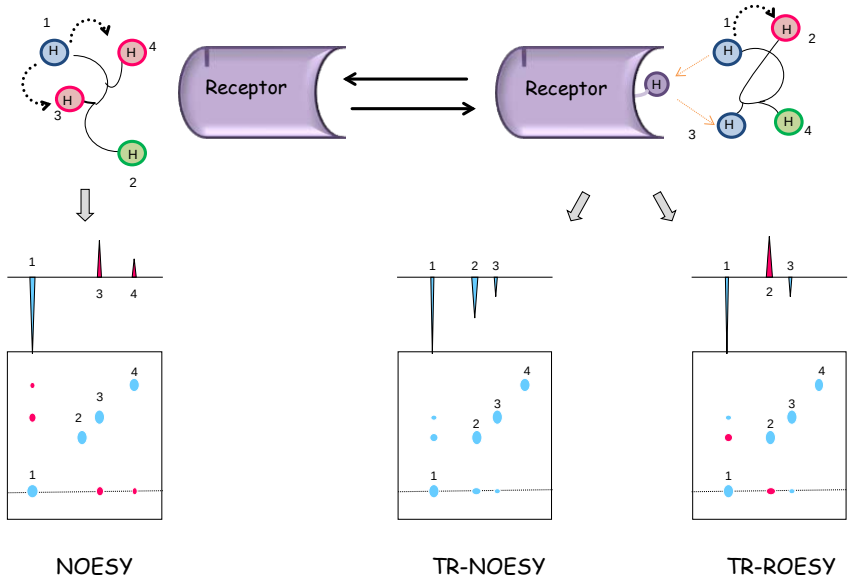


Fig. 2.6: *Left.* Schematic representation of a NOESY spectrum of a ligand in free state; cross peaks and diagonal peaks have different signs. *Right.* Schematic representation of a tr-NOESY and a tr-ROESY spectra recorded on ligand–protein mixture. In tr-NOESY spectrum, cross peaks and diagonal peaks 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 [19]. In tr-ROESY spin diffusion cross peaks (H1/H3) and diagonal peaks have the same signs, while direct cross peaks (H1/H2) show different sign to diagonal peaks.

The observation of intramolecular trNOEs relies on different tumbling times τ_c of free and bound molecules. As mentioned above, in ideal conditions, low molecular weight molecules exhibit positive NOEs, while larger molecules show negative NOEs. Thus, when a small ligand molecule binds to a high molecular weight protein, it behaves as a part of a large molecule and adopts the corresponding NOE behavior. As a consequence, the NOEs signs undergo drastic changes and lead to the appearance of transferred NOEs (Figure 2.6).
The conditions for the applicability of this approach are established considering the well known equilibrium and the molar fractions of free and protein bound sugar [Equations 2.1-2.5].
Furthermore, in order to perform tr-NOESY experiments it is important that the Equations 2.6, 2.7 and 2.8 hold

$$p_b \cdot \sigma^B > p_f \cdot \sigma^F \qquad \text{[equation 2.7]}$$

$$k_{-1} \gg \sigma^B \qquad \text{[equation 2.8]}$$

where p_b and p_f are the molar fractions of bound and free ligand, σ^B and σ^F are the cross relaxation rates of bound and free state (defined as the rates of cross-relaxation for exchange of magnetization between two spins.), while k_{-1} is the off rate constant.
Under these conditions, the dipole-dipole cross relaxation rate constant become the population-weighted averages of the free and the bound states:

$$\sigma_{avg} = p_b \cdot \sigma^B > p_f \cdot \sigma^F \qquad \text{[Equation 2.9]}$$

In the fast exchange regime, during the reversible binding, the ligand acquires the NOE characteristics of the large receptor then transferred from the bound state of the ligand to its free state. The ligand signals are still sharp owing to the rapid total correlation time of the free ligand.

It is worth to note that in NOESY spectra, spin diffusion effects, which are typical for large molecules, are often observable. They lead to the appearance of negative cross peaks between protons which are far apart in the macromolecule because spins which are not close in space (including those of the receptor) may mediate the exchange of magnetization. Thus, receptor-mediated indirect tr-NOE effects may produce interpretation errors in the analysis of the ligand bioactive conformation. However, using trNOEs in the rotating frame (tr-ROESY) experiments [20], 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 2.6).

Since in many systems of interaction, the ligands are molecules with high molecular weight and exhibit negative NOEs in free state, the discrimination between trNOEs originating from the bound state and NOEs of the ligand in solution can also be achieved by determining the build-up rate. It is the time required to achieve maximum NOE intensity which for tr-NOEs is in the range of 50 to 100 ms, whereas for non-binding molecules it is four- to ten-times longer. Therefore, the maximum enhancement for tr-NOEs is observed at significantly shorter mixing times than for isolated smaller molecules in solution (Figure 2.7).

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, which has considerable implication for a rational structure-based drug design.

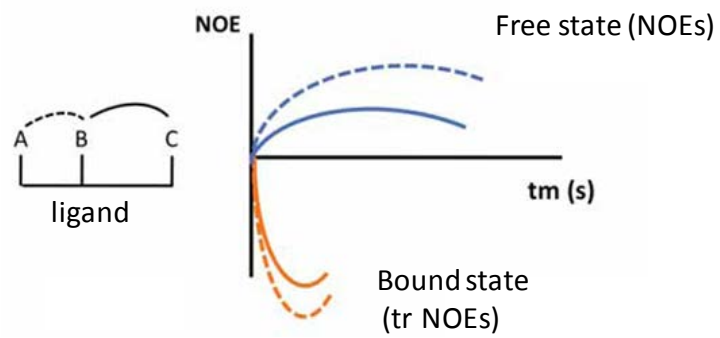


Fig. 2.7: NOE build up depending on mixing time.

2.3.2 Saturation transfer difference NMR

Saturation transfer difference NMR was a pioneering technique developed in 1979 by Wuthrich's group [21] for the resonances assignment of the heme group in oxidized cytochrome c-557 from *Crithidia oncopelli*. More recently, Mayer and Meyer proposed the same NMR method in order to elucidate ligands binding to UDP-GlcNAc 2-epimerase/ManNAc kinase [22]. Originally proposed as a technique for the rapid screening of compound libraries, nowadays STD NMR is one of the most powerful and widespread NMR method for detection and characterization of transient receptor-ligand interactions in solution. Indeed, it allows to determine the binding epitope of ligands, detecting the regions in closer contact to the receptor protein.

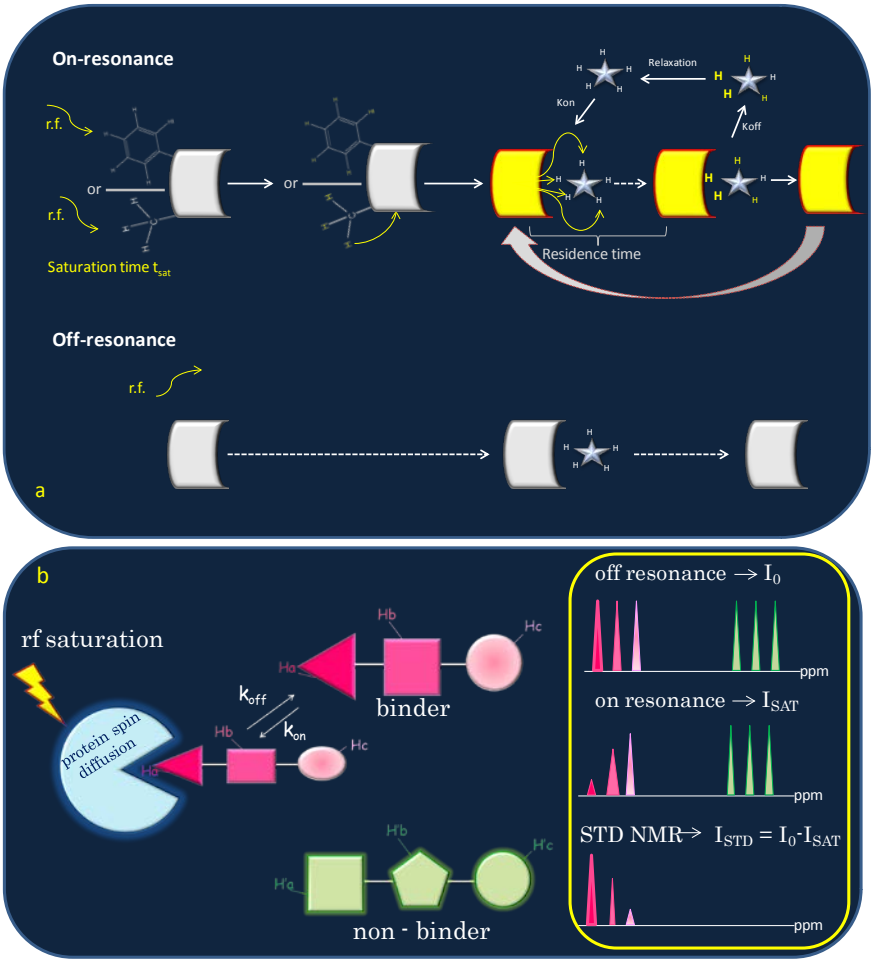


Fig. 2.8: Representation of STD NMR spectroscopy.
a) STD involves two experiments: the on- and the off- resonance. During the on-resonance, frequency selective irradiation (r.f.) causes saturation of the receptor protein. A train of shaped pulses is applied for a sustained period during which saturation propagates between target protons by spin diffusion; hereafter the saturation is transferred to binding compounds during their residence time in the protein binding site. Once experienced the saturation transfer, the binders ligands turn in solution, by chemical exchange, where the saturation is detected. In the off-resonance experiment, rf saturation is applied in an off-resonance region from both receptor and ligand producing a reference spectrum.
b) The off resonance experiment (right top panel) produces spectra with intensity I_0 since the irradiation frequency is set to a value where resonances of both protein and ligand are absent. Whereas, during the on resonance experiment

(middle pannel), the irradiation frequency is set to a value where only resonances from the protein nuclei and no resonances from ligand nuclei are located. This manifests as the decreased signal intensity I_{SAT}. The STD spectrum (bottom pannel) is produced by the difference between the on and off resonance spectra containing only the resonances of binding compounds, with intensity $I_{STD} = I_0 - I_{SAT}$. Ligand protons in closer contact to the binding site show higher STD signals respect to more external residues, whereas, non binders ligands are unaffected.

This experiment is performed on a sample containing the receptor (large molecule with MW $\geq$ 15kDa) and a large molar excess of small compounds. It is based on the transfer of saturation from the selected receptor protons to the bound ligand via the vast network of intermolecular ^1H-^1H cross-relaxation pathways. Hereafter, by chemical exchange, the ligand molecules turn into solution where the saturated state persists due to their small R_1 (enthalpic relaxation) values and it is detected (Figure 2.8a).
As the name suggests, the STD NMR spectrum is produced by the difference between two experiments: the on resonance and the off resonance (Figure 2.8a/b) [10]. In the first one, the magnetization of the protein is selectively saturated applying a train of frequency selective rf pulses to a region of the spectrum that contains receptor resonances while ligand signals are absent. The on resonance irradiation frequency values are typically set to around -1ppm since no ligand nuclei resonances are found in this spectral window, whereas the protein signals still allow the selective saturation. The rf value could even be placed in the aromatic proton spectral region (or also downfield, around 11-12 ppm) if ligand resonances are absent in that ppm range. Once saturated via spin diffusion, quite efficient due to the large molecular weight of the target, the receptor transfers the saturation to binding compounds via intermolecular ^1H-^1H cross relaxation at the ligand-protein interface. The small molecules then dissociate back in solution where the saturation is detected and, at the same time, unsaturated

ligands exchange on and off the receptor thus increasing the population of saturated free molecules in solution.
During the off resonance experiment, the identical rf train is applied far off- resonance such that no NMR resonances are perturbed; it represents the reference spectrum.
Both the experiments (on and off resonance) are recorded in an interleaved fashion and subtracted; the resulting difference is the real STD NMR spectrum which yields only resonances of the receptor and binding compounds (Figure 2.8b). It is worth to note that receptor resonances usually are not visible since the protein concentration in solution is very low and, in addition, they easily can be deleted by R_2 relaxation filtering prior to detection.
An STD pulse scheme is depicted in Figure 2.9, where, d1 is the relaxation delay; the duration of saturation time (for on and off resonance alternatively) typically is 1-3 seconds and consists of a cascade of Gaussian shaped pulses (each 50 ms long, separated by an interpulse delay (δ) of ~1ms) in order to reach the desired selectivity and to avoid side-band irradiation. Selectivity of pulses in general is achieved by reducing the amplitude intensity of rf pulse and at the same time increasing the duration; rectangular pulses are rarely used for selective excitation due to their unfavorable profile and associated sidebands responsible of unwanted excitation signals, whereas shaping functions are the most effective for this purpose. In detail, the Gaussian shaped pulse, described by equation 2.10 (where t_0 is the center of pulse envelope, S the intensity of the pulse, *a* comprises the pulse duration determining the pulse width, and *t* is the time), ensures a bandwidth of the irradiation of the order of 40Hz.

$$S(t) = \exp [-1 (t - t_0)^2] \quad \text{[Equation 2.10]}$$

Later the 90° pulse, a spin lock filter of 10-50 ms is used for the suppression of receptor signals. The subtraction between on and off

resonance is performed after every scan by phase cycling; the phases of 90° pulse, $T_{1\rho}$ filter, and receiver are respectively:

ϕ1: (x, -x, -x, x, y, -y, -y, y, -x, x, x, -x, -y, y, y, -y);
ϕ2: 2(y, -y), 2(-x, x);
ϕrec: 2(x), 2(-x), 2(y), 2(-y), 2(-x), 2(x), 2(-y), 2(y).

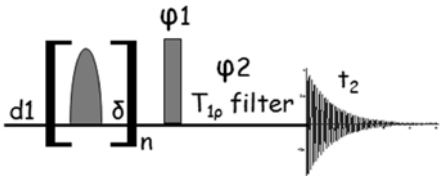


Fig. 2.9: Graphical representation of STD sequence.

The signals in the on resonance spectrum possess intensity I_{SAT}, while those in the off resonance exhibit the equilibrium value I_0. Subtraction of the two intensities yields to the STD response, which is of the same form as a traditional steady state NOE:

$$\eta = (I_0 - I_{SAT}) / I_0 = I_{STD}/I_0 \qquad \text{[Equation 2.11]}$$

It is worth to note that, for a given ligand, not all its protons will receive the same amount of saturation since the transfer of magnetization from the receptor to small molecules depends on the inverse sixth power of intermolecular ^1H-^1H distance in the bound state. It means that ligand protons in closer contact to the binding site of the protein receive an higher degree of saturation and therefore give rise to stronger STD signals. This allows the identification of the portions of the ligand more involved in the interaction process, known as “epitope mapping” in reference of the original meaning of detecting antigen regions necessary for antibody or T-cell recognition.
Mayer and Meyer introduced the refined group epitope mapping (GEM) for identifying binding surfaces on the ligand using STD

methods. This approach is based on the comparison the STD response for different protons within a ligand. This is done by normalizing all the measured STD signals against the most intense one, which is arbitrarily assumed to be 100%. The set of resulting STD percentages qualitatively delineates the chemical moieties which are key for molecular interaction, since they are intimately recognized by the protein (STD values close to 100%), and the regions of ligand situated far from the receptor binding site.

The degree of ligand saturation naturally depends on its residence time in the protein-binding pocket. In addition, the STD sensitivity depends on the number of ligands receiving the saturation from the receptor and can be described in terms of the average number of saturated ligands produced per receptor molecule. Indeed, the saturation of the protein and the bound ligand is about 100 ms and, assuming fast ligand exchange, the information about saturation is transferred very 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; thus, the proportion of saturated ligands in solution increases during the sustained rf pulse train (saturation time) and the information about the bound state resulting from the saturated protein is amplified. It means that the more ligand that is used, the longer the irradiation time, the stronger the STD signal is. In general, an irradiation time of 2 seconds and a 100-fold excess of ligand give good results. Obviously, from the high ligand : protein ratios, it is clear that only a relatively small amount of protein is required for the measurements (nM-μM).

Ideally, epitope information would not depend on the chosen saturation time; however, significantly different R_1 relaxation rate of the ligand protons can produce artifacts in the epitope definition. Indeed, protons with slower R_1 relaxation efficiently accumulate

saturation in solution giving rise to higher STD relative intensity, which means that their proximity to the binding pocket may be overestimated. Thus, in order to quantitatively well investigate the interaction of a ligand with a receptor protein, the use of STD build-up curves has been proposed [23-24]. Deriving the initial slope (STDfit) of an STD build up curve, which is acquired by collecting STD spectra at different saturation times (typically from 1 s to 5 s), represents a good way to cancel out all these artifacts, since at zero saturation time virtually no accumulation of saturated ligand occurs, and at the same time intramolecular spin diffusion within the bound state is minimized.

The slope of the curve at zero saturation time is obtained from fitting the STD NMR intensities data to the monoexponential equation 2.12, where STD is the STD signal intensity of a given proton at a saturation time t, STDmax stands for the asymptotic maximum of the curve, and ksat is the observed saturation rate constant that measures the speed of STD build up.

$$STD = STDmax(1 - \exp(-ksat\ t)) \quad \text{[Equation 2.12]}$$

The epitope map of the ligand is then obtained by normalizing all the values of different ligand protons to the largest STDfit, giving STD epitopes fit, that represents the real STD intensity, in the absence of T_1 bias and merely dependent on the proximity of the ligand protons to the protein.

Data obtained from STD NMR experiments are fundamental not only in the definition of the epitope mapping, but they can even be used in the determination of K_D values, since STD intensity reflects the concentration of ligand-receptor complex present in solution. For this frame, it is important to define the STD amplification factor (STD-AF) that has been introduced in order to well quantify the amplification of resonances in the STD experiments (Equation 2.13) [25],

$$\text{STD-AF} = \varepsilon(I_0 - I_{SAT}) / I_0 = \varepsilon\eta_{STD} \quad \text{[Equation 2.13]}$$

By multiplying the observed STD for the molar excess of ligand over the protein (ε), it is possible to convert the STD intensity, which depends on the fraction of bound ligand, into a factor that is a function of the fraction bound protein. It provides a convenient means for gauging the inherent sensitivity of the experiment; in addition, spectra from samples having different receptor concentrations may still be compared. The evolution of the STD-AF along a titration series enables construction of a saturation curve in the form of an association isotherm (Equation 2.14), analogous to the well known equation for the Henri-Michaelis-Menten enzymatic reaction rate.

$$\text{STD-AF} = \alpha_{STD}[L] / [L] + K_D \quad \text{[Equation 2.14]}$$

with α_{STD} being a dimensionless scaling factor that represents the maximum STD amplification. The value of dissociation constant results from the mathematical fit of the experimental curve; however, in many cases, apparent K_D values are obtained depending strongly on experimental conditions, such as saturation time, relative concentration of the species and monitored proton. The origin of this discrepancies is due to ligand rebinding in solution; saturated ligand molecules return to the complex before their saturation has completely relaxed back to equilibrium. If the lifetime of the free state is shorter than the T_1 relaxation time of the proton considered, the accumulation of saturated ligand in solution decreased and the observed STD signal is significantly reduced [26]. The strongest effects on K_D are observed for low ligand : protein ratios (so for larger protein concentrations), for ligand protons with the largest STD intensities, and for titration with larger affinity. Furthermore, careful studies performed by Angulo [10], showed that

STD data tend to underestimate K_D values since they were always lower or equal to the thermodynamic value.

In order to remove all effects of fast protein - ligand rebinding in solution and accurately determinate dissociation constant values, is required the determination of the initial slope of the build up curve of STD-AF values with the saturation time. Once obtained, for each ligand concentration, a build up curve, data are fitted to obtain the initial build up slopes then plotted against the ligand concentration yielding a Langmuir isotherm from which the true dissociation constant K_D can be derived.

It is worth to note that STD effects depends largely on the off rate; the K_D range of STD method has been estimated to be $10^{-8} – 10^{-3}$ M (Mayer and Meyer 1999). For week binders (having $K_D > [L_0]$), the probability of the ligand to be in the receptor pocket is very low. As K_D increases further, the population of the complex decreases leading to a reduction and ultimately disappearance of the STD signals. On the other hand, if binding is very tight, and consequentially off rates are in the range of 0.1-0.01 Hz, decreasing K_D, increases the receptor - ligand lifetime and the saturation transfer is not very efficient. At sufficiently small K_D, the free state residence times of the ligand can exceed their free state R_1 values and the exchange is so slow that no STD signals are observable.

Fortunately, STD NMR method is well suited for the study of protein – carbohydrate complexes which exhibit excellent properties for STD NMR spectroscopy since the typical values of their K_D are in the $10^{-3} – 10^{-6}$ M range.

In conclusion, STD NMR is a solid and versatile technique which gives essential information, at a molecular level, on receptor – ligand interactions. It may be combined with any NMR spectroscopy pulse sequence generating a whole suite of STD NMR experiments, such as STD TOCSY and STD HSQC [27], fundamental for defying the epitope mapping of the ligand and determine K_D values.

References

[1] Silverstein, R. M.;. Webster, F. X.; KiemleJohn, D. J. *Wiley & Sons*, **2005**.
[2] Claridge, T. D. W. *Tetrahedron Organic Chemistry*, **2009**.
[3] Keeler, J. J. *Wiley & Sons*, **2010**.
[4] Sharon, N.; Lis, H. *Glycobiology*, **2004**. 14, 53R-62R.
[5] Bishop, J. R.; Gagneux, P. *Glycobiology*, **2007**. 17, 23R–34R.
[6] Imberty, A.; Varrot, A. *Curr Opin Struct Biol*.,**2008**. 18, 567–576.
[7] Rillahan, C. D.; Paulson, J. C. *Annu Rev Biochem*., **2011.** 80, 797–823. doi: 10.1146/annurev-biochem-061809-152236.
[8] Meyer, B.; Peters, T. *Angew Chem Int Ed*, **2003**. 42: 864-890.
[9] Lepre, A.C.; Moore, M. J.; Peng, W.J. *Chem. Rev*. **2004**. 104, 3641-3675.
[10] Angulo, J; Nieto, P. M. *Eur Biophys J*., **2011**. 40, 1357–1369.
[11] Calle ,L.P.; Canada, F.J.; Jimenez-Barbero, *J Nat Prod Rep*, **2011**. 28, 1118–1125.
[12] Johnson, M. A and Pinto, B. M. *Carbohydrate Research*, **2004**. 339, 907–928.
[13] Bhunia, A.; Bhattacharjya, S.; Chetterjee, *S. Drug Discov Today*, **2012**, doi: 10.1016/j.bbamem.2012.01.008.
[14] Carlomagno, T. *Annu. Rev. Biophys. Biomol. Struct*., **2005**. 34, 245–66.
[15] Hahn, E. L. and Maxwell, D. E. *Phys Rev*, **1952**. 88, 1070.
[16] McConnell, H. M. *J Chem Phys*, **1958**. 28, 430.
[17] Peng J. W.; Moore J. M. and Abdul-Manan, N. *Prog Nucl Magn Reson Spectrosc*, **2004**. 44, 225.
[18] Neuhaus, D. Wiley-VCH, **2000**.
[19] Poveda, A. and Jimenez-Barbero, J. *Chem Soc Rev*, **1998**. 27, 133.
[20] Asensio, L.; Canada, F. J.; Bruix, M.; Rodriguez-Romero, A. and Jimenez-Barbero, J. *Eur J Biochem*, **1995**. 230, 621.
[21]Keller, R.M.; Wuthrich, K. *Biochim. Biophys. Acta*, **1978.** *533*, 195–208.
[22] Mayer, M. and Meyer, C. *Angew Chem Int Ed Engl*, **1999**. 38, 1784.
[23] Angulo, J.; Díaz, I.; Reina, J. J.; Tabarani, G.; Fieschi, F.; Rojo, J.; Nieto P.M. *ChemBioChem* **2008**. *9*, 2225–2227.

[24] Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chem. Eur. J.* **2011.** *17*, 1547–1560.
[25] Mayer, M. and Meyer, C. *J. Am. Chem. Soc*. **2001**. 123,6108.
[26] Angulo, J.; Enriquez-Navas, P.M.; Nieto, P.M. *Chem Eur J*. **2010**. 16, 7803-7812.
[27] Vogtherr, M.; Peters, T. J. *AM. Chem. Soc*., **2000**.122, 6093-6099.

SECTION II

PROTEINS – GLYCOCONJUGATES INTERACTION INVOLVED IN BACTERIAL COMMUNICATION

Chapter 3

PrkC – Muropeptides interactions

3.1 Introduction

Bacterial Ser/Thr kinases modulate a wide number of cellular processes. In *Bacillus subtilis*, the Serine/Threonine protein kinase, PrkC, has been shown to induce the germination of bacterial spores in response to DAP-type, and not to Lys-type peptidoglycan fragments (Figure 3.1, cfr. 1.1.4) [1-2].

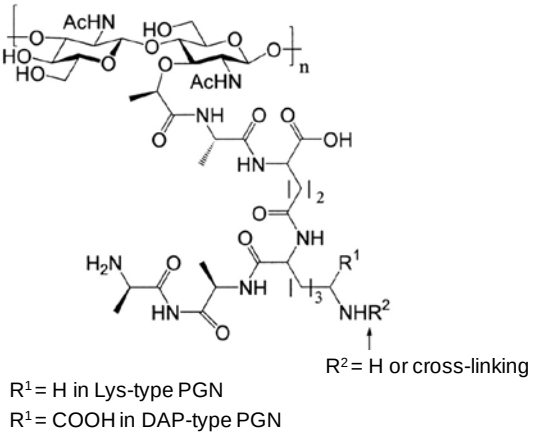


Fig. 3.1: General structure of DAP- and Lys- type PGN. That component of bacterial cell wall is formed by a network of glycan strands of alternating β-(1,4)-linked N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) possessing interconnected oligopeptide stems.

Muropeptides represent a clear signal that growing conditions are promising since they are produced during cell wall peptidoglycan remodeling associated to cell growth and division of neighboring bacteria. However, if muropeptides are able to physically bind the protein and how the extracellular region is able to distinguish the

two types of muropeptides remains unclear. In this study, we tackled the important question of how the extracellular region of PrkC (EC-PrkC) senses muropeptides. By coupling NMR techniques and protein mutagenesis, we prove that muropeptides bind to EC-PrkC through interactions, mediated by the *meso*-DAP moiety, with an arginine residue, Arg500, belonging to the protein C-terminal PASTA domain. Notably, mutation of this arginine completely suppresses muropeptide binding. These data provide the first molecular clues into the mechanism of muropeptides sensing by PrkC [1].

3.1.1 Bacteria and germination of dormant spores

Gram positive bacteria, such as *Bacillus* and *Clostridium* spp., can exist in metabolically inactive states that allow them to survive in case of nutrient starvation and to adverse environment, as severe physical and chemical conditions which are not conducive for growth. Such dormant cells, known as endospores, are highly resistant to a variety of stress factors, including heat, desiccation, radiation, pH extremes, antimicrobial factors and toxic chemicals, that rapidly kill vegetative cells [2-4].The high resistance of dormant cells is due to the peculiar structural organization characteristic of bacterial spores. They are rod shaped cells composed by an internal double membrane-bound compartment, named forespore, with protective structures assembled inside of and around the forespore. The coat, a protein multilayer, encases and protects the mature spore; some species, such as *Bacillus anthracis*, present an additional shell, the exosporium, that covers the completed spore, while in others, including *Bacillus subtilis*, the coat is the outermost layer [5]. The interior region of the dormant cell contains the spore chromosome and is known as core; its partial dehydration likely prevents enzymatic activity and influences the spore dormancy and

resistance, which are essential prerogatives for the ability of many pathogenic bacteria to cause disease.

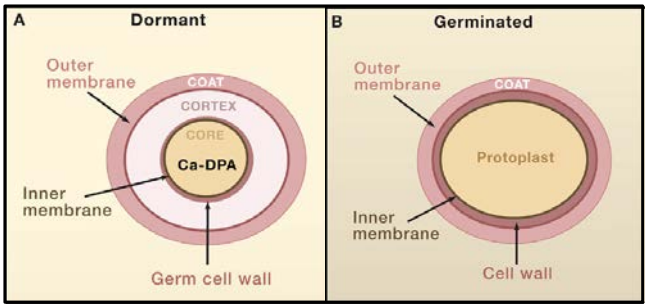


Fig. 3.2: Schematic structure of *Bacillus subtilis* spore when dormant (A) and vegetative (B). The coat, outer membrane, and cortex of dormant spores are not an effective permeability barrier to passage of small molecules, such as muropeptides, which are produced during the bacterial growth and allow the germination of dormant spores. (Adapted from Setlow, P. 2008).

The exit from dormancy is a process known as germination (Figure 3.2). Although spores are metabolically inactive, they are able to monitor the extracellular environment to identify growth-promoting conditions allowing them to reinitiate metabolism. The return to vegetative growth of bacterial spores can be triggered by nutrients or other agents, named germinants, that bind to specific receptors situated in the inner membrane of dormant cells signaling the improving of the environmental conditions. When germination is triggered, water enters the spore core, which swells, and ultimately the dormant spore converts into a vegetative cell. Germinants can be single amino acids, sugars, purine nucleosides, mixture of nutrients as asparagines, glucose, fructose and K+ (AGFK) and trigger *B. subtilis* germination [3-4]. An alternative and different mechanism for initiating germination has been recently demonstrated by Dworkin and coworkers [6]. They showed that metabolic products (≤ 1 pg/ml), released into the extracellular milieu during the bacterial growth, can act as a germinant indicating favorable conditions for cell growth, since other bacteria

are growing in the local environment. In detail, the authors [6] underlined that many microorganisms exit dormancy in response to cell wall muropeptides, spontaneously released by other growing microorganisms [2,6] following cell wall remodeling [7,8]. Around 30-50% of muropeptides are released by growing cells during each bacterial generation as peptidoglycan hydrolases and amidases partially digest the mature peptidoglycan to allow insertion of additional monomers; this causes releasing of peptidoglycan fragments, named muropeptides [1,3,7,9]. Turnover of cell wall proceeds proportionally to the growth rate. As a consequence, the bacterial growth in the local environment represents one of the factors able to induce the germination of dormant spores.

3.1.2 PrkC and PASTA domains

The exit from dormancy in response to the secreted PGN fragments requires the intervention of members of serine/threonine kinases (STPKs), characteristic only of Gram positive bacteria, which comprise a catalytic domain located in the cytoplasm, joined by a single transmembrane-spanning region to a large extracellular domain. The recent crystal structure of the extra-cellular region of a close homologue of *B. subtilis* PrkC from *Staphylococcus aureus*, determined in an our previous work [10], has shown that this region is formed by three PASTA (Penicillin binding Associated and Serine/Threonine kinase Associated) domains [6,11,12], also predicted by the PFAM database [13], and an unpredicted Immunoglobulin like domain [11] (Figure 3.3).
PASTA domains have been found in well conserved eukaryotic-like membrane serine/threonine kinases of many pathogens, including *Bacillus subtilis* and *Staphilococcus aureus,* but also exist in high molecular weight penicillin binding proteins (PBP) [14] in association with other domains like transpeptidases and glycosyltransferases. The STPKs serve as signal transducers and they are involved into the bacterial growth, the cell wall

biosynthesis, the regulation of the bacterial development and pathogenesis processes. On the other hand, the PBPs coordinate bacterial cell wall biosynthesis, remodeling and reproduction. Their extracellular region consists in one or more PASTA repeats acting as a signal binding sensor domain that, when bound, leads up to an intracellular conformational change with the activation of signaling cascade [15].

The four domains composing the EC-PrkC are arranged sequentially in a golf-club shape, such that only neighboring domains interact with each other (Figure 3.2B). Three of the four domains are, as predicted, PASTA repeats (Figure 3.2A). The structure shows that the three PASTA domains display a linear and regular organization. Indeed, each domain exhibits a two-fold symmetry with respect to its neighbouring domains (Figure 3.2B). In this organization, the sole α-helix of each domain is alternatively located on the two sides of the golf club (Figure 3.2B). Interestingly, the structure reveals the existence of a fourth domain, at the C-terminal end of the molecule, not predicted by searches in the PFAM database [11]. Furthermore, sequence analyses against the PDB do not identify any significant homologue for this domain.

In the crystal structure of the penicillin binding protein PBP2x from *S. peumoniae*, one of its two C-terminal PASTA domains binds cefuroxime, a β-lactam antibiotic mimicking the unlinked peptidoglycan [12,16]. This finding has suggested that PASTA domains of PrkC can bind muropeptides [2]. However, no experimental evidence has hitherto been provided to support this hypothesis. According to Dworkin and coworkers, the germination response to muropeptides requires a highly conserved eukaryotic-like membrane serine/threonine kinase, PrkC in *Bacillus subtilis*, that determines the bacterial ability to respond to micromolar concentrations of PGN fragments. The extracellular domain of *B. subtilis* PrkC contains three PASTA repeats likely responsible for peptidoglycan binding. However, it is worth to note that the extracellular domain of *B. subtilis* PrkC shows the ability to bind

the muropeptides deriving only from DAP-type PGN-while it does not recognize Lys-type, highlighting the potential importance of the third residue of the peptide stem in the recognition process.

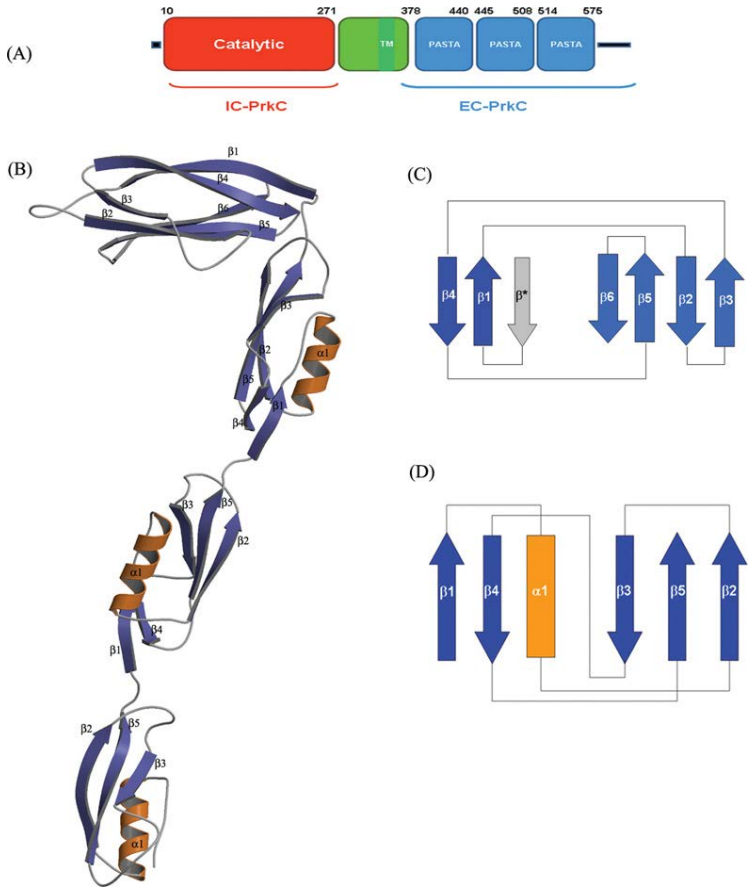


Fig. 3.3: The modular structure of EC-PrkC. (**A**) Domain prediction of PrkC, according to the PFAM database. IC-PrkC and EC-PrkC are the intracellular and extracellular regions of PrkC respectively. (**B**) Ribbon representation of EC-PrkC. Helices and β-strands are shown in orange and blue respectively. (**C**) Schematic diagram of canonical Ig domain topology. The grey strand in the topology sketch (strand β*) is lacking in PrkC-Ig. (**D**) Schematic diagram of PASTA domain topology.

On the other hand, it is important to underline that the PGN fragments deriving from germinating spores [17] do not act as a germinant, hence avoiding the possibility that spontaneously germinating spores induce neighboring dormant cells to growth, stimulating the germination under inappropriate conditions.
The mechanism of peptidoglycan-induced activation of PrkC is so far unknown. In order to demonstrate, under the biochemical point of view, that bacterial endospores retain an alert sensory mechanism to respond to released muropeptides and hence initiate germination, we have generated a blend of pure natural muropeptides from *Bacillus subtilis* PGN, and expressed and purified the EC-PrkC from *Bacillus subtilis.* Then, we performed NMR experiments describing, at molecular level, the interaction between the extracellular domain of eukaryotic-like membrane serine/threonine kinase (EC-PrkC) from the Gram positive *Bacillus subtilis*, using as ligands monomeric and dimeric muropeptides deriving from the DAP-type PGN. We were interested in applying NMR techniques, such as saturation transfer difference NMR spectroscopy and transferred NOE experiments, to investigate such interactions and to characterize the ligand's bioactive structures.
The results revealed the first full picture of the binding of muropeptides to the extra-cellular region of PrkC and provided the chemical basis for the discrimination of PGN types.

3.2 Binding of EC-PrkC from *Bacillus subtilis* to muropeptides

The study of the binding and the interaction modes between the extracellular domain of *Bacillus subtilis* PrkC and monomeric ($GlcNAcMurNAcAla_2GluDAP$) and dimeric ($GlcNAc_2MurNAc_2Ala_4Glu_2DAP_2$) muropeptides, deriving from DAP-type PGN, has been performed through NMR experiments.

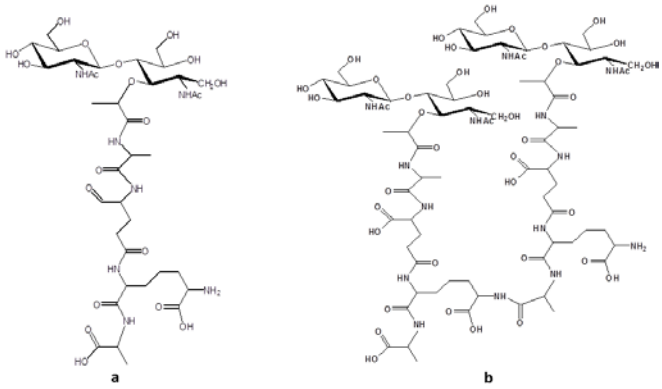


Fig. 3.4: Structures of monomeric (A, *m/z*= 942) and dimeric (B, *m/z*= 1865) muropeptides species.

Table 3.1: Chemical shifts of ^{1}H and ^{13}C of the muropeptide species **1**.

Residues	**H_a**	**H_b**	**H_g**	**H_d**	**H_e**				
A\A'	4.09	1.21							
Ala	49.3	16.57							
C	4.17	1.17							
Ala	49.5	18.3							
D	4.08	1.58/1.70	1.24	1.68	3.50				
A_2pm	53.3	29.9	21.4	30.0	54.2				
E	4.02	1.47/1.59	1.23	1.47/1.59	3.86				
A_2pm	53.7	30.8	21.0	30.8	54.8				
G	3.92	1.09							
Ala	50.6	17.3							
H/H'	3.94	1.94/1.71	2.12						
i-Glu	54.4	27.6	31.8						
	H-1/C-1	**H-2/C-2**	**H-3/C-3**	**H-4/C-4**	**H-5/C-5**	**H-6/C-6**		**CH_3 Ac**	
B	4.46	3.52	3.42	3.26	3.20	3.57/3.68		1.84	
GlcNAc	101.3	55.6	73.1	69.0	75.7	59.9		22.1	
	H-1/C-1	**H-2/C-2**	**H-3/C-3**	**H-4/C-4**	**H-5/C-5**	**H-6/C-6**	**H'-2/C'-2**	**Me'-2/C'-2**	**CH_3 Ac**
F	3.41/3.46	4.21	3.70	3.6	3.62	3.39/3.59	4.03	1.21	1.75
MurNAc	61.0	51.4	76.3	78.0	70.3	61.8	77.9	18.8	21.1

Table 3.2: Chemical shifts of ^{1}H and ^{13}C of the muropeptide species **2**.

Residues	**H_a**	**H_b**	**H_g**	**H_d**	**H_e**				
A'	4.09	1.21							
Ala	49.3	16.57							
E	4.02	1.47/1.59	1.23	1.47/1.59	3.86				
A_2pm	53.7	30.8	21.0	30.8	54.8				
H'	3.94	1.94/1.71	2.12						
i-Glu	54.4	27.6	31.8						
	H-1/C-1	**H-2/C-2**	**H-3/C-3**	**H-4/C-4**	**H-5/C-5**	**H-6/C-6**		**CH_3 Ac**	
B	4.46	3.52	3.42	3.26	3.20	3.57/3.68		1.84	
GlcNAc	101.3	55.6	73.1	69.0	75.7	59.9		22.1	
	H-1/C-1	**H-2/C-2**	**H-3/C-3**	**H-4/C-4**	**H-5/C-5**	**H-6/C-6**	**H'-2/C'-2**	**Me'-2/C'-2**	**CH_3 Ac**
F	3.41/3.46	4.21	3.70	3.6	3.62	3.39/3.59	4.03	1.21	1.75
MurNAc	61.0	51.4	76.3	78.0	70.3	61.8	77.9	18.8	21.1

Peptidoglycan from growing *B. subtilis* cells was extracted, purified and digested into a muropeptide blend using mutanolysin, an enzyme that hydrolyzes the β-1,4 bond between the MurNAc and GlcNAc sugars. The generated peptidoglycan fragments were reduced with sodium borohydride, then identified by RP-HPLC and LC-MS; dimeric and monomeric muropeptides were isolated and analysed.

The main isolated compounds, a dimeric species (m/z = 1865.1 Da) and a monomeric species (m/z = 942 Da), were accumulated by reiterated HPLC injections of both muropeptide preparations and analyzed by a combination of two dimensional NMR and tandem MS. The muropeptide species are shown in Figure 3.4.

The complete assignment of ^{1}H and ^{13}C resonances was achieved combining the information from DQF-COSY, TOCSY, ROESY, NOESY and ^{1}H-^{13}C-HSQC 2D-NMR experiments (Table 3.1 and 3.2). The amino acid sequence was univocally established by

comparison of inter-residual dipolar NMR correlations obtained by ROESY and NOESY experiments.

3.2.1 Binding of EC-PrkC from *Bacillus subtilis* to monomeric muropeptide **1**

At first, using tr-NOE and STD NMR techniques, we analyzed the interaction between EC-PrkC of *B. subtilis* and the isolated monomeric muropeptide (**GlcNAcMurNAcAla$_2$GluDAP**).
ROESY and trROESY experiments were performed to obtain information on the bioactive conformation of the ligand. To observe transferred NOEs, a protein-ligand ratio at 1:20 in binding sites was used. A qualitative analysis of the tranferred NOEs observed for the monomeric acceptor ligand reveals only minor changes respect to the ligand in the free state, allowing to deduce a virtually identical conformation between the free and the bound state, and thus indicating that, upon binding to the protein, the bioactive conformation of the acceptor was very similar to the most populated conformation of the free ligand. On the other hand, we obtained interesting results performing STD NMR experiments. To optimize the experimental parameters and choose the optimal conditions for executing STD spectra, we performed several experiments at different temperatures, protein selective irradiation frequencies and saturation time; the best magnetization transfer was obtained using an on resonance frequency of 0 ppm. Once these parameters were optimized we performed STD NMR on the protein in the presence of the ligand using a molar ratio of 1:50 (Figure 3.5).
Because of considerable signal overlapping that significantly complicated the analysis of the STD NMR spectrum, we estimated the contribution for each proton also from STD-2D-TOCSY NMR spectra (data not shown).
To determine the binding epitope of the monomeric muramyl peptides, relative STD effects have been calculated from the STD

amplification factors that were obtained from the STD NMR experiments. By comparing STD NMR spectrum with its corresponding reference, it could be clearly inferred that the monomer of PGN bound to the EC-PrkC protein, as showed by several enhancements in the STD NMR spectrum (Figure 3.5a/b).

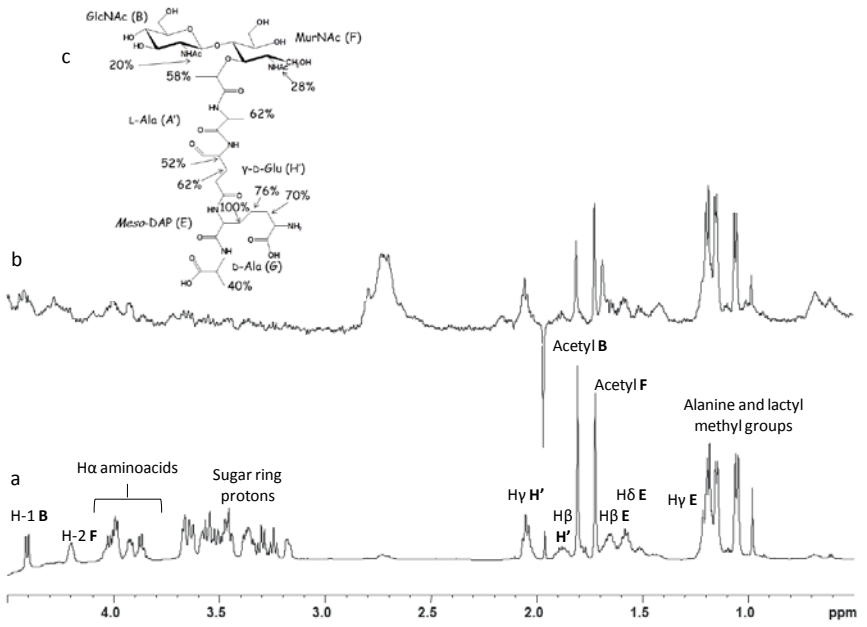


Fig. 3.5: a) ^{1}H NMR spectrum of **1**; b) 1D STD NMR spectra of **1** in the presence of EC-PrkC (1:50 ratio). c) Chemical structure and epitope binding of **1** to ECPrkC; the percentages refer to the relative STD effects.

In detail, the most relevant STD signals have been observed between 1-2 ppm, indicating that this region, mainly corresponding to the lateral chain of the amino acid moieties, was the main involved in the recognition process, since it makes the closest contacts with residues in the active site of EC-PrkC. STD signals appeared even at lower field were both α protons signals of the amino acid stem and those belonging to the sugar backbone resonated, but they presented a significantly lower intensity

compared to the corresponding signals in the reference spectrum. A quantitative analysis of the STD NMR allowed us to confirm a strong involvement of the peptide stem in the binding. Indeed, the analysis of the relative signals allowed us to estimate the contribution of each proton and thus to obtain the STD-NMR derived epitope mapping (Figure 3.5c and Table 3.3).
The highest signal was observed for the H_β proton of *meso*-diaminopimelic acid, resonating at 1.65 ppm. A medium-high distribution of % of STD was also observed for side-chain protons of other aminoacids of the peptide stem, whereas a low intensity of the STD signals was recorded for the carbohydrate moieties.

Table 3.3 Relative STD effects of **1** bound to EC-PrkC.

	STD %
H_β meso-DAP (E)	100
Hγ meso-DAP (E)	76
H_δ meso-DAP (E)	70
H_α i-Glu (H')	52
H_β i-Glu (H')	62
Hγ i-Glu (H')	50
H_β Ala (G)	40

3.2.2 Binding of EC-PrkC from *Bacillus subtilis* to dimeric muropeptide **2**

STD NMR experiments, carried out to investigate the direct binding of the dimeric muropeptide **2** to EC-PrkC, confirmed a major involvement of the peptide stem of the muropeptide in the binding. Indeed, strongest STD signals characterize the amino acidic portion of the dimeric ligand whereas weak albeit

quantifiable signals were observed for the sugar backbone (Figure 3.6 and Table 3.4).

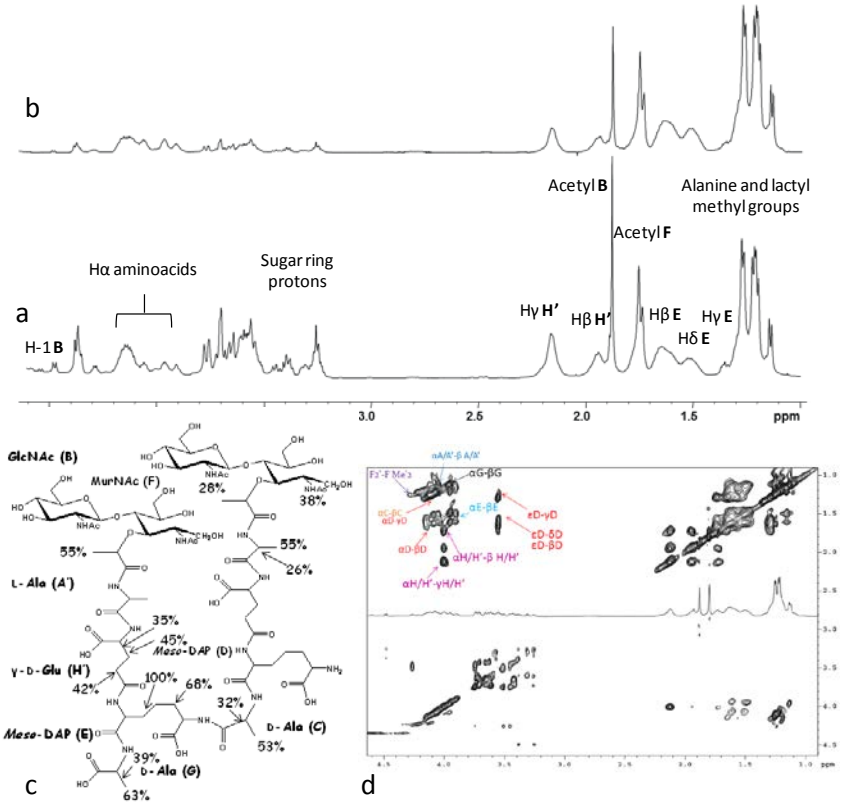


Fig. 3.6: a) ^{1}H NMR spectrum of 2; b) 1D STD NMR spectra of **2** in the presence of EC-PrkC (1:50 ratio). c) Chemical structure and epitope binding of **2** to ECPrkC; the percentages refer to the relative STD effects. d) 2DSTD TOCSY experiment.

In detail, spectra showed the highest STD contribution for protons of the *meso*-diaminopimelic acid (H_β meso-DAP I_{STD} 100%) and only a small involvement, with STD signals below 20%, of the sugar backbone (Figure 3.6c). It is worth to note that, owing to severe signals overlapping in the region of the spectrum between 3-5.2 ppm (Figure 3.7), the univocal individuation of the proton

signals in contact with the protein was achieved by two dimensional STD experiments (Figure 3.6d).

Table 3.4 Relative STD effects of **2** bound to EC-PrkC.

	STD %		STD %
H_β meso-DAP (E)	100%	H_α i-Glu (H/H')	35%
H_δ meso-DAP (E)	68%	H_β i-Glu (H/H')	45%
H_β meso-DAP (D)	68%	Hγ i-Glu (H/H')	42.5%
Hγ meso-DAP (D)	55.2%	H_α Ala (G)	38.8%
H_δ meso-DAP (D)	43.8%	H2 MurNAc (F)	15.2%
H_β Ala (G)	63.6%	H1 GlcNAc (B)	15.2%
H_α Ala (G)	39%	H3 MurNAc (F)	15.1%
H_α Ala (C)	32 %	H6 MurNAc (F)	14.7 %
H_β Ala (C)	53.4 %	H3 GlcNAc (B)	12.9%
H_α Ala (A)	26 %	CH_3Ac GlcNac (B)	27.8%
H_β Ala (A)	55 %	CH_3Ac MurNAc (F)	38%

Consistent with data obtained for the monomeric muropeptide, ROESY and tr-ROESY spectra showed that the bioactive conformation of **2** was not modified upon protein binding (Figure 3.8).

Altogether, NMR data, collected on the two germinant muropeptides, showed that the *meso*-diaminopimelic acid plays a key role in the binding of the extracellular domain of the *B. subtilis* PrkC, while the carbohydrate unit takes part in the recognition protein but contributes in minor extent to the interaction.

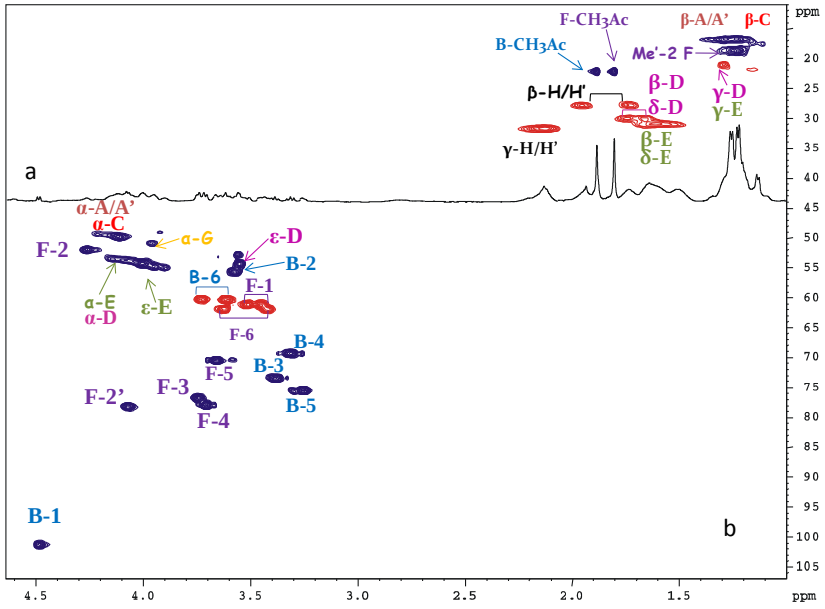


Fig. 3.7: a) 1D STD NMR spectrum of the mixture protein-ligand; b) HSQC of the muropeptide **2**.

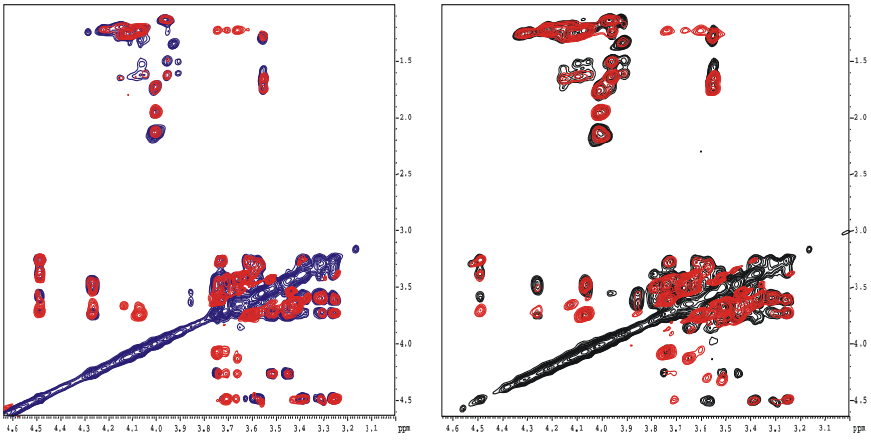


Fig. 3.8: ROESY(left) and tr-ROESY(right) respectevely on PGN dimer and on the mixture EC-PrkC from *B. subtilis* – PGN dimer 1 : 25.

3.2.3 Binding of EC-PrkC from *Bacillus subtilis* to DAP-type and Lys-type tripeptide

In order to further investigate the interaction between EC-PrkC of *B. subtilis* and PGN fragments we studied the binding of DAP-type tripeptide (L-Ala-γ-D-Glu-*m*-DAP) to EC-PrkC. The signal enhancements present in the STD NMR spectrum (Figure 3.9) confirmed that the DAP-type tripeptide was able to bind to EC-PrkC. As already found exploring the binding to muropeptides **1** and **2** to EC-PrkC, the most relevant STD signals have been observed between 1-2 ppm, which is the region corresponding to amino acids side chains.

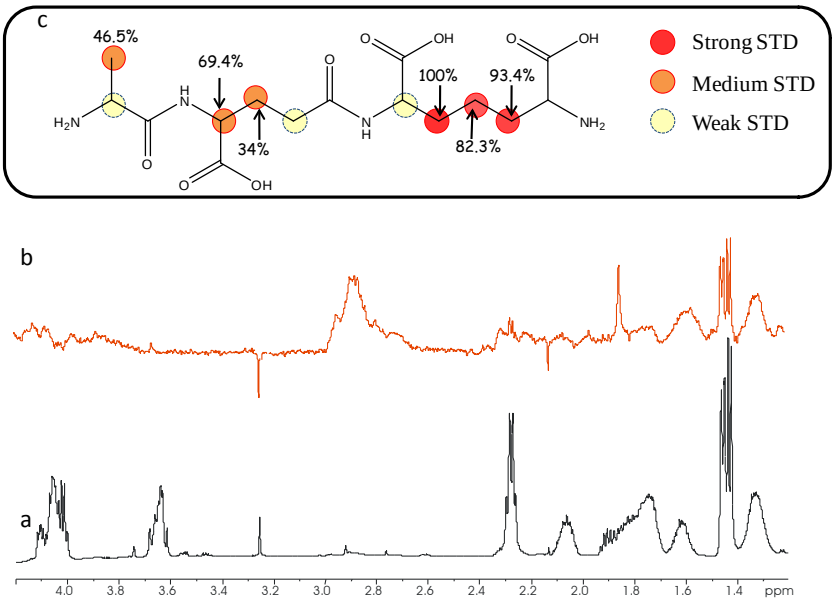


Fig. 3.9: a) Reference ^{1}H NMR spectrum and b) STD 1D NMR spectrum of *B. subtilis* EC-PrkC – tri-DAP mixture, 1:100. On resonance irradiation frequency at 0.5 ppm, saturation time of 3 seconds.

More specifically, largest STD signals were observed for protons belonging to the lateral chain of the DAP residue, confirming its key role in the recognition process.

The inability of EC-PrkC to bind to a lysine containing tripeptide, tri-Lys (L-Ala-γ-D-Glu-Lys), further confirmed the above conclusions (Figure 3.10). STD NMR experiments performed on the LYS-type tripeptide, indeed, did not show any enhancements, demonstrating that the recognition of EC-PrkC is DAP-driven.

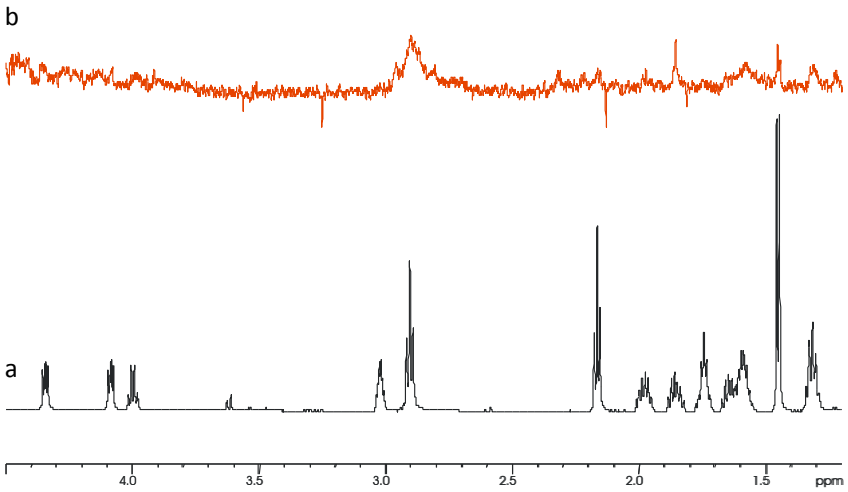


Fig. 3.10: a) Reference ^{1}H NMR spectrum and b) STD 1D NMR spectrum of *B. subtilis* EC-PrkC– tri-LYS mixture, 1:100. On resonance irradiation frequency at 0.5 ppm, saturation time of 3 seconds.

3.3 Identification of the muropeptide binding site on EC-PrkC

In collaboration with the group of professor Rita Berisio, CNR of Naples, we reasoned on the key role played by DAP in EC-PrkC binding, as determined by STD NMR experiments (Figure 3.5 and 3.6). Indeed, the massive involvement of DAP in binding well agrees with the previous finding that only peptidoglycan containing DAP at the third position of the peptide stem germinate *B. subtilis* spores, whereas peptidoglycan containing L-Lys at this position does act as a germinant [2]. Prompted by this observation, we

carried out a statistical survey in the Protein Data Bank (PDB) to identify the structural determinants responsible for DAP binding to proteins. Four crystal structures of protein complexes with DAP containing muropeptides are available in the PDB. Notably, in all of the structures, the carboxylate end of DAP forms a salt bridge with an arginine side chain (Figure 3.11).

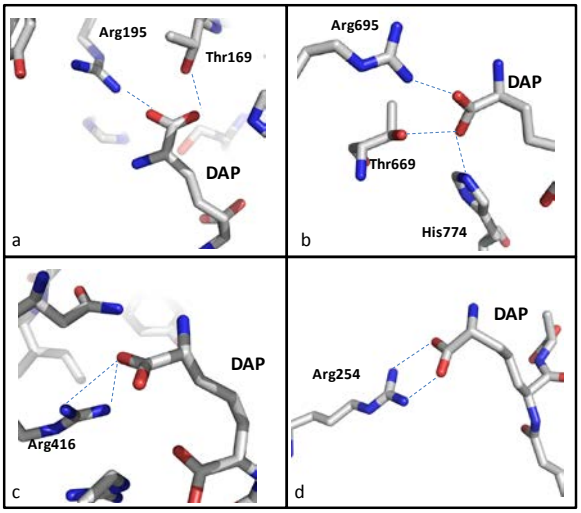


Fig. 3.11: Interaction of DAP with protein structures available in the Protein Data Bank. The four panels represent the binding of DAP to (a,b) Corynebacterium glutamicum diaminopimelate dehydrogenase (PDB codes 2DAP and 1F06, respectively), (c) UDP-N-acetylmuramoyl-L-alanyl-D-glutamate: meso-diaminopimelate ligase from *Escherichia coli* (PDB code 1E8C), (d) Drosophila peptidoglycan recognition protein (PGRP)-LE (PDB code 2CB3).

This finding, together with the STD NMR data, suggested that the muropeptide binding occurs through an interaction of DAP with an arginine residue of EC-PrkC.

EC-PrkC from *B. subtilis* contains two arginine residues, Arg500 and Arg614, which are located in the third PASTA domain and in the C-terminal IG-like domain, respectively. The model of EC-PrkC structure shows that Arg614 is involved in salt bridge

formation with Glu604 whereas Arg500 is fully solvent exposed (Figure 3.12).

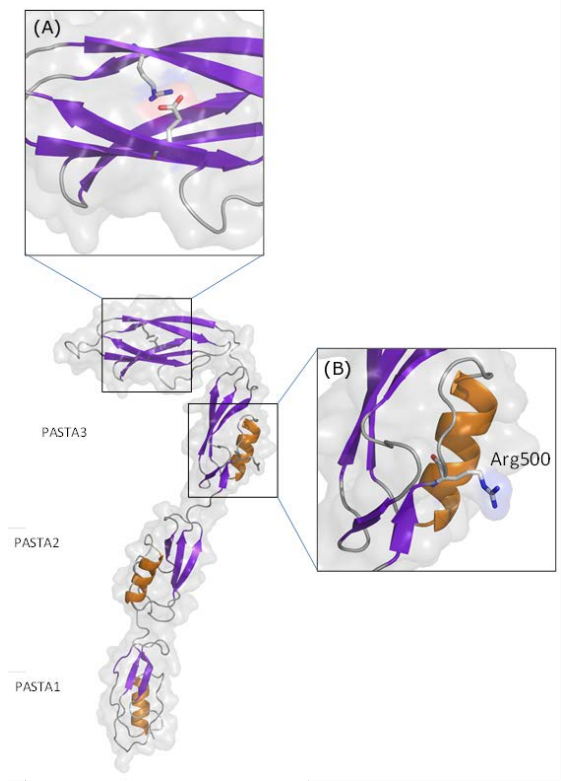


Fig. 3.12: Structure of EC-PrkC from *B. subtilis* derived by homology modeling. The two insets show enlargements of (A) the salt bridge between Arg614 and Glu604 in the IG-like domain and (B) the location of Arg500 in PASTA3.

These considerations led us to hypothesize that Arg500 is involved in muropeptide binding, in a fashion similar to that observed in the PDB (Figure 3.11). To corroborate this hypothesis, we mutated Arg500 to alanine (R500A) and to glutamic acid (R500E), which was predicted to be more disruptive of the Arg-DAP interaction. Surprisingly, STD NMR experiments performed in the same conditions used for wild type EC-PrkC unequivocally show that

neither R500E nor R500A are able to bind muropeptides. Indeed, no STD signals were observed in either cases (Figure 3.13).

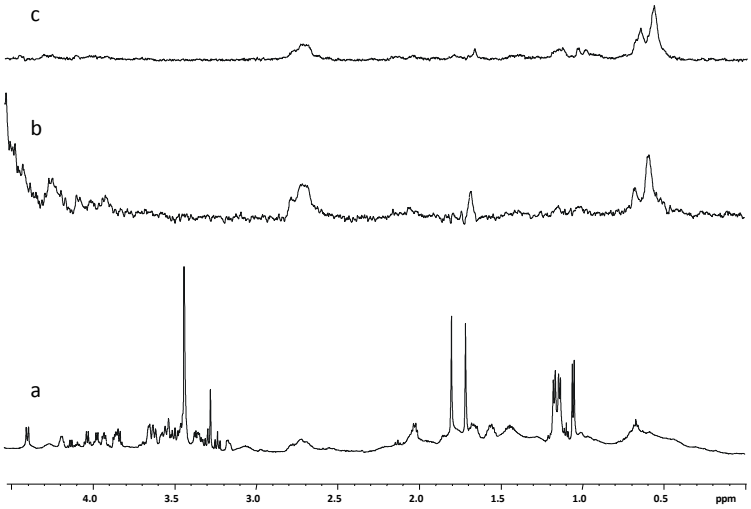


Fig 3.13: a) ^{1}H NMR spectrum of 2; b) 1D STD NMR spectra of **1** in the presence of R500A and (c) R500E mutants (1:100 ratio). Signals at about 2.9 and 0.5 ppm in (b-c) are residual protein resonances.

To exclude that differences in the muropeptide binding behaviour of the two mutants were due to defects in protein folding induced by mutations, CD spectra of R500E and R500A were compared to that of the wild type structure. Results show that mutations did not alter the structural integrity of EC-PrkC (Figure 3.14).
The inability of R500E and R500A mutants to bind muropeptides definitively shows that Arg500 is the main anchorage of muropeptides. The combination of these results with STD NMR experiments on the wild type EC-PrkC proves that PrkC senses muropeptides through interactions of the carboxyl group of the DAP moiety with the side chain of Arg500.

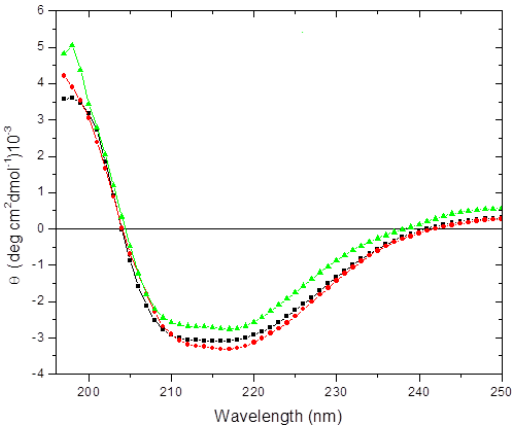


Figure 3.14: CD spectra of R500A (black) and R500E (red) mutants of EC-PrkC compared to the CD spectrum of wild-type EC-PrkC (green).

Furthermore, the observed inability of the non-disruptive R500A mutant to bind muropeptides shows that the contribution of sugar moieties is not sufficient for binding in the absence of DAP-Arg interactions.

3.4 Oligomerization state of EC-PrkC

The identification of direct binding between DAP-type muropeptides and EC-PrkC raised the possibility of physical interactions between EC-PrkC monomers, as previously suggested for the Ser/Thr kinase PknB from *Mycobacterium tuberculosis* [18]. In collaboration with the group of professor R. Berisio, CNR of Naples, we addressed the question of whether muropeptides alter the oligomerization state of EC-PrkC, using a set of complementary techniques and various experimental conditions.

First, different pH (pH=7.5 and 4.5) and ionic strength conditions have been screened using analytical size-exclusion chromatography, coupled with triple-angle light scattering (SEC-LS). The on-line measurement of the intensity of the Rayleigh scattering as a function of the angle as well as the differential refractive index of the eluting peak in SEC was used to determine

Mw. SEC-LS experiments were carried out also after overnight incubation of EC-PrkC with 8-molar excess of PGN muropeptide **2**. All experiments produced Mw values which corresponded to a monomeric organization of the molecule (Figure 3.14).

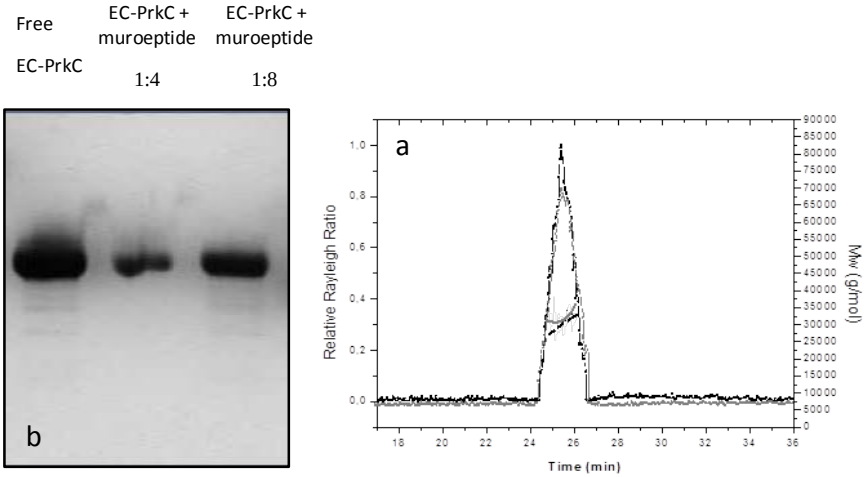


Fig 3.14: Effect of muropeptides on EC-PrkC oligomerisation state. (a) Analytical SEC-MALS. Rayleigh ratios and molecular masses (left and right scales, respectively) are plotted against the elution time. The black curve corresponds to EC-PrkC in 30 mM NaCl and 8 mM Tris–HCl, pH 8.0. The grey curve was measured in the same buffer after the incubation of EC-PrkC with 8-molar excess of dimeric muropeptide. In both experiments, average Mw values correspond to a monomeric state of the protein, 31840 (±955) Da and 32570 (±1302) before and after incubation with the muropeptide, respectively. (b) Native gel electrophoresis (Lane 1, Free EC-PrkC 25μg; Lanes 2 and 3, EC-PrkC after incubation with dimeric muropeptide ($GlcNAc_2MurNAc_2Ala_4Glu_2DAP_2$) with ratios 1:4 and 1:8, respectively.

Furthermore, we sought for non-discrete oligomers, which may be not easily detectable by SEC-LS, as previously observed for Peptidoglycan Recognition Proteins [19]. Non-discrete oligomers, typically detected by native gel electrophoresis, result in the formation of smear bands on the native gel. As a result of these experiments, a single band accumulates on the gel; this band does

not migrate, neither smears when a sample containing EC-PrkC incubated with **2** is loaded. Similar results were obtained by Surface Plasmon Resonance (Biacore 3000), using two different immobilization methods (CM5 sensor chip by covalent EDC-NHS amine coupling and on an NTA (nitrilotriacetic acid) sensor chip) to ensure different experimental conditions and protein orientations on the chip (Figure 3.15).

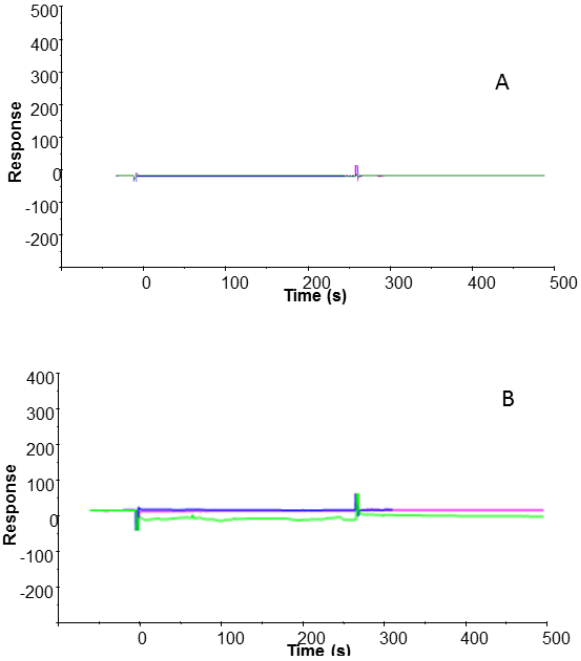


Fig 3.15: Surface Plasmon Resonance experiments. Overlay of SPR sensorgrams recorded (A) using amine coupling (CM5) exploring protein concentration ranges between 0 µM (blue) and 200 µM (green), (B) after EC-PrkC incubation with increasing equivalents of muropeptides, using a fixed protein concentration of 20 µM. Blue, magenta and green sensograms refer to 0, 10 and 100 muropeptide equivalents, respectively. Panel C reports the same experiment described in panel B, using the His-tag capture (NTA).

Last, the change of the molecular size distribution of EC-PrkC at increasing protein concentrations has been checked using Small

Angle X-ray Scattering (SAXS) experiments. In all screened conditions (data not shown), results indicate that the particles are elongated and the estimated Mw as well as computed volumes are compatible with a monomeric organization in solution. In addition, incubation of the protein with a 8-molar excess of **2** does not significantly affect the protein scattering pattern (Figure 3.16).

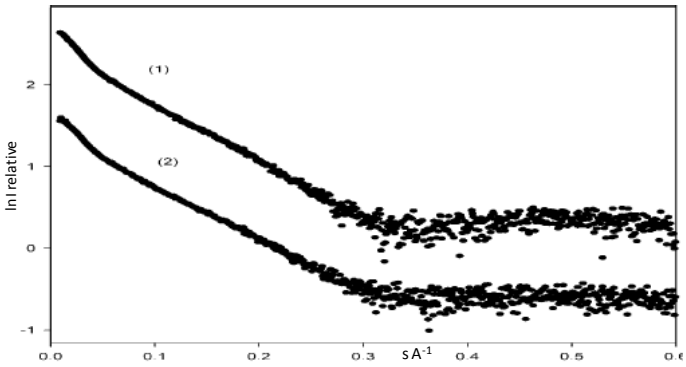


Fig. 3.16: Small-angle X-ray scattering profiles of EC-PrkC (1) without and (2) with addition of **2**. The plot displays the logarithm of the scattering intensity as a function of momentum transfer $s = 4\pi \sin(\theta)/\lambda$, where 2θ is the scattering angle and $\lambda = 0.15$ nm is the X-ray wavelength. Curves have been arbitrarily displaced along the logarithmic axis for better visualization. Altogether, all adopted biophysical techniques show that EC-PrkC has no tendency to form oligomers in solution, independent of the protein concentrations, and that the binding of muropeptides does not affect the protein oligomerization state.

3.5 Conclusion

Bacteria use an extensive network of post-translational modifications to transmit signals and to coordinate cellular functions. Phosphorylation frequently occurs in response to environmental signals and is mediated by specific protein. Among these, eukaryotic-type serine/threonine kinases are expressed in many prokaryotes including a broad range of pathogens. Despite a discrete amount of evidence on downstream actions, phosphorylation of protein substrates and subsequent consequences

for the bacterial physiology, we know precious little of the upstream part of the cascade, the ligand binding. In *B. subtilis*, the kinase PrkC has been shown to sense muropeptides of the cell wall and to control spore germination [2]. The mechanism of peptidoglycan-induced activation of PrkC is, however, so far unknown. We reported the x-ray characterization of the entire extra-cellular region of the homologous PrkC kinase from *Staphylococcus aureus*. The sole structure determination could not, however, predict that protein interactions with muropeptides exists. Thus, by combining an array of state-of-art biophysical and biochemical techniques, we have addressed the important question of how EC-PrkC senses muropeptides. Saturation Transfer Difference NMR results unambiguously showed that EC-PrkC from *B. subtilis* was able to bind DAP-containing muropeptides. Indeed, STD signal was particularly strong for the muropeptide DAP moiety, indicating that PGN fragments were anchored to the protein principally through *meso*-diaminopielic acid. This result was consistent with the observed ability of DAP-containing muropeptides, but not Lys-containing ones, to germinate *B. subtilis* spores. It is worth to note that the *meso*-diaminopimelic acid is a peculiar component of PGN of all Gram negative bacteria and of few Gram positive bacteria, such as bacilli. Thus, our results explain the specificity of the interaction between EC-PrkC and exclusively DAP-type PGN.

By coupling protein mutagenesis and NMR techniques we succeeded also in defining the structural requirements of EC-PrkC protein necessary for recognition and binding of muropeptides. In particular, an arginine residue belonging to the third PASTA domain played a key role in the interaction with *meso*-diaminopimelic moiety of PGN.

Therefore, our results demonstrated that the sugar backbone of PGN contributed to the interaction with the receptor protein although it was less involved in the binding process respect to the peptide region. On the other hand, the third PASTA domain was

essential for the recognition process, since it was the protein region in closer contact to PGN fragments, allowing the fundamental ARG-DAP interaction.

References

[1] Squeglia, F.*; Marchetti, R.*; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *JACS.* **2011.** 133, 20676–20679. *equal contribution.
[2] Dworkin, J.; Shah I.M. *Nature Reviews Microbiology*, 2010. **8,** 890-896.
[3] Setlow, P. *Cell*, **2008**. 135, 410-412.
[4] Setlow, P. *Current Opinion in Microbiology*, **2003**. 6, 550-556.
[5] Aronson, A. I. and Fitz-James**,** P. *Bacteriol. Rev*, 1976. **40,** 360–402.
[6] Shah, I.M. et al. *Cell*, **2008**. 135: 486-496.
[7] Doyle, R. J.; Chaloupka, J.; Vinter, V. *Microbiol. Rev*, **1988**. 52, 554–567.
[8] Goodell, E.W. *J Bacteriol*., **1985**. 163, 305-310.
[9] Boudreau, M. A.; Fisher, J. F.; Mobashery, S.. *Biochemistry*, **2012**. 51, 2974-90. DOI:10.1021/bi300174x.
[10] Ruggiero, A.; Squeglia, F.; Marasco, D.; Marchetti, R.; Molinaro, A.; Berisio, R. *Biochem J*. **2011.** 435, 33-41.
[11] Jones, G. and Dyson P. *Journal of Bacteriology*, **2006**. 188, 7470-7476.
[12] Maestro, B. et al. *FEBS letters*, **2010**. 585, 357-363.
[13] Finn, R. D.; Tate, J.; Mistry, J.; Coggill, P. C.; Sammut, S. J.; Hotz, H. R.; Ceric, G.; Forslund, K.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. *Nucleic Acids Res* **2008**, *36*, D281-8.
[14] Gordon, E.; Mouz, N.; Duee, E.; Dideberg, O. *J Mol Biol.* **2000**. *299*, 477-85.
[15] Yeats et al. *TRENDS in Biochemical Sciences*, **2002**. 27, 438-440.
[16] Dessen, A.; Mouz, N.; Gordon, E.; Hopkins, J.; Dideberg, O. *J Biol Chem* **2001**. *276*, 45106-12.
[17] Meador-Parton, J.; Popham, D.L. *Journal of Bacteriology*, **2000**. 182, 4491-4499.
[18] Barthe, P.; Mukamolova, G. V.; Roumestand, C.; Cohen-Gonsaud, M. *Structure* **2010**. *18*, 606-15.
[19] Lim, J. H.; Kim, M. S.; Kim, H. E.; Yano, T.; Oshima, Y.; Aggarwal, K.; Goldman, W. E.; Silverman, N.; Kurata, S.; Oh, B. H. *J Biol Chem* **2006**. *281*, 8286-95.

Chapter 4

Bc2lA – carbohydrates from bacterial LPS interaction

4.1 Introduction

Bacteria from the Burkholderia cepacia complex (Bcc) cause highly contagious pneumonia in cystic fibrosis (CF) patients. Among them, *Burkholderia cenocepacia* is one of the most dangerous and is the most frequent cause of morbidity and mortality in CF patients. Indeed, it is responsible of "cepacia syndrome", a deadly exacerbation of infection, that is the main cause of poor outcomes in lung transplantation. *Burkholderia cenocepacia* bacterium produces several soluble lectins with specificity for fucosylated and mannosylated glycoconjugates [1]. These proteins are present on the bacterial cell surface and it has been proposed that they bind to lipopolysaccharide epitopes. In this work, we report on the interaction of one *B. cenocepacia* lectin, BC2L-A, with heptose and other *manno* configured sugar residues. Saturation Transfer Difference NMR spectroscopy studies of BC2L-A with different mono- and disaccharides demonstrated the requirement of *manno* configuration with hydroxyl or glycol group at C6 for the binding process. The crystal structure of BC2L-A complexed with the methyl-heptoside confirmed the location of the carbohydrate ring in the binding site and elucidated the orientation of the glycol tail, in agreement with NMR data. Titration calorimetry performed on monosaccharides, heptose disaccharides and bacterial heptose-containing oligosaccharides and polysaccharides confirmed that bacterial cell wall contains carbohydrate epitopes that can bind to BC2L-A. Additionally, the

specific binding of fluorescent BC2L-A lectin on *B. cenocepacia* bacterial surface was demonstrated by electron microscopy [2].

4.1.1 Lectins: ubiquitous proteins with a taste for sugars

As mentioned in Section I, whilst carbohydrates are a source of biological information written in a specific "language", lectins are proteins that have the ability to decipher this glycocode and to translate it for the biological community, without modifying sugars or invoking an immune response [3,4].

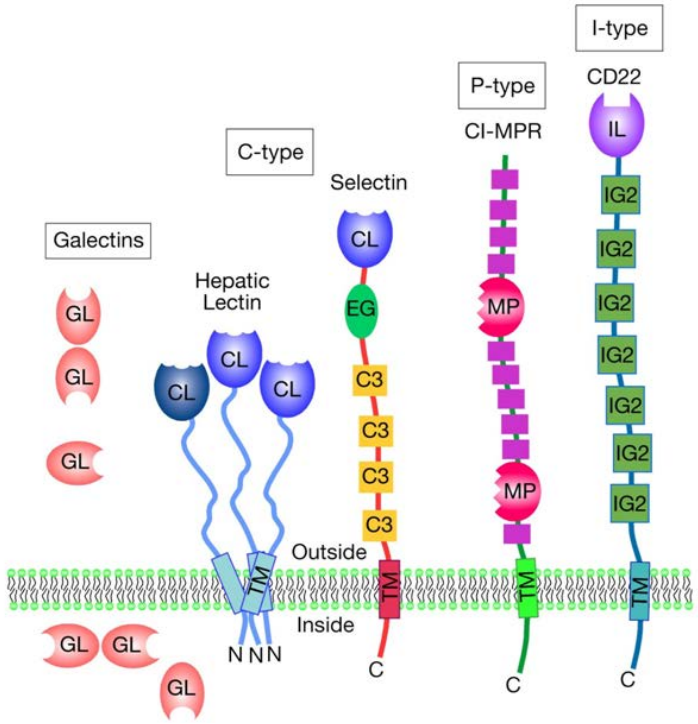


Fig. 4.1: Schematic examples of major types of animal lectins, based on protein structure. The carbohydrate-binding domains (CRD) can vary among family members. In figure are shown some defined CRD: (CL) C-type lectin CRD; (GL) S-type lectin CRD; (MP) P-type lectin CRD; (IL) L-type lectin CRD. Other domains are (EG) EGF-like domain; (IG2) immunoglobulin C2-set domain;

(TM) transmembrane region; and (C3) complement regulatory repeat. (Adapted from “Essential of Glycobiology”, second edition).

In general, lectins are oligomeric proteins composed by several sub-units, at least one of which will contain a carbohydrate binding site (CRD: carbohydrate recognition domain) (Figure 4.1). They can be of different sizes, specificities and exhibit dependencies on other complimentary elements (e.g. the presence of metal ions) in order to facilitate carbohydrate recognition [5,6], as well as different 3D structural organizations of their sub-units (tetrahedral, planar etc). These proteins interact with carbohydrates non-covalently in a manner that is usually reversible and highly specific [7,8]. According to the monosaccharide ligand toward which they exhibit the highest affinity, Sharon classified lectins into five groups: mannose, galactose/N-acetylgalactosamine, Nacetylglucosamine, fucose and N-acetylneuraminic acid [9] (Figure 4.2).

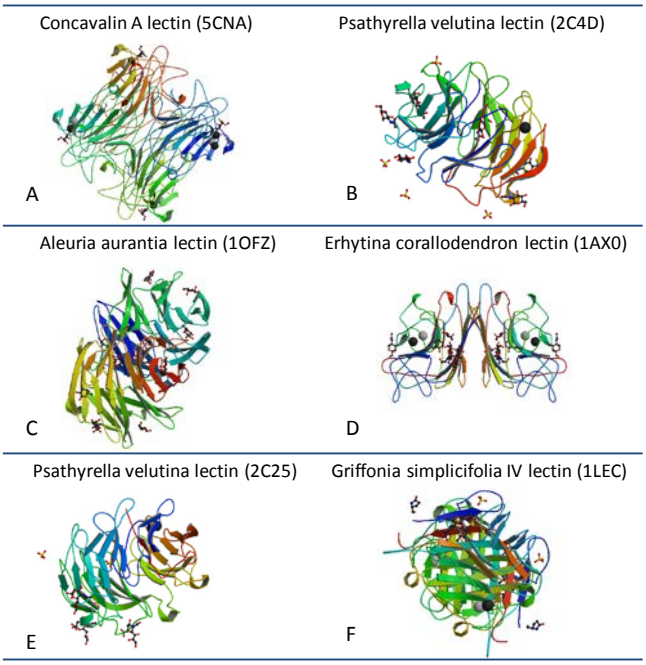


Fig. 4.2: Examples of lectins that specifically recognize mannose (**A**, (48)), N-acetylglucosamine (**B**,(49)), fucose (**C**,(50)), galactose/N-acetylgalactosamine

(**D**,(51)), N-acetylneuraminic acid (**E**,(49)). The lectin represented in **F** ,(52) is specific for the oligosaccharide Fucα2Galβ3(Fucα4)GlcNAc. PDB codes are reported in brackets.

Lectins are found all through nature and exist in most living organisms, ranging from viruses and bacteria to plants and animals. They were first thought to be unique to plant cells, where they are proposed to be involved in protection against pathogens or feeders; however, lectins associated with bacterial, viral and higher organisms are now well documented [7,8] (Figure 4.3). A classification according to their origin is reported in Table 4.1.

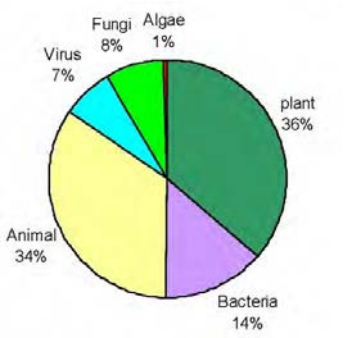


Fig. 4.3: Repartition of the 187 different lectins with structures available in the Protein data bank and accessible in the lectin database (http://www.cermav.cnrs.fr/lectines/), classified as a function of origin.

In pathogenic micro-organisms, lectins are mostly involved in cell-cell signaling and host tissue adhesion [10]. They can be present as subunit parts of toxins, such as in *Vibrio cholera* [11], in which the role of the lectin is to explore and to bind the host cell surfaces, while the toxin subunit invades the cells upon lectin association. They can also be part of multiprotein organelles, such as bacterial fimbria (pili), flagella and viral capsids; therefore, they participate to the adhesion process at the early stage of infection. Soluble lectins are also expressed as virulence factors by opportunistic bacteria; they were first evidenced in *Pseudomonas aeruginosa* [12] and later in other Gram negative bacteria (Table 4.2).

Table 4.1 Classification and functions of lectins according to their origin.

Lectin	Function
MICROORGANISMS	
Bacteria	
Pili or fimbriae	Adhension, infection
Soluble lectins	Adhesion, infection, biofilm formation
Virus	
Toxin	Adhesion, infection
Hemagglutinins	Adhesion, infection
Amoeba	
Surface lectin	Adhesion
ANIMALS	
Calnexin	Assisting protein folding of glycoprotein
M-type	E.R. protein degradation
L-type	Regulation of biosynthesis of glycoproteins
P-type (Mannose 6P-receptor)	Protein sorting post Golgi (apoptosis)
C-type (calcium dependant lectins)	Adhesion of leucociti to cells of bood vessels (selectins); Immunity regulation (collectins)
I-type	Adhesion (siglecs)
R-type	Target for enzymes, hormane regulation
Galectins	Glycan recognition in extracellular matrix
PLANTS	
Leguminosae lectins	Defence, symbiosis
Others	Defence

In detail, two soluble lectins, LecA (PAIL) and LecB (PAIIL), specifically binding galactose and fucose, were initially identified and characterized in the cytoplasm of the opportunistic pathogen *P. aeruginosa* [13]. Lec A and Lec B are C-type bacterial lectins, which form tetramers under physiological conditions. Each monomer, exhibiting a molecular weight of ~12.8 kDa, consists of a β-sandwich and contains two calcium ions that interact directly with the carbohydrate ligand.

PA-IIL-like lectins have been characterized in other opportunistic bacteria such as *Ralstonia solanacearum* [16-17], *Chromobacterium violaceum* [15], and *Burkholderia cenocepacia* [18-19], albeit with some variations in the fine specificity.

Table 4.2 Different classes of soluble bacterial lectins.(Adapted from ref. [9]). These lectins are all virulence factors produced under the control of quorum sensing.

Bacterium	Lectin	Higher affinity ligand	Structure PDB	Oligomer (sites number)
Pseudomonas aeruginosa [13]	PA-IL (LecA)	Globoside αGal4Gal	2VXJ	Tetramer
Pseudomonas aeruginosa [14]	PA-IIL (LecB)	Lewis a	1W8H	Tetramer
Chromobacterium violaceum [15]	CV-IIL	Fucosylated blood group epitopes	2BOI	Tetramer
Ralstonia solanacearum [16]	RSL	Fucosylated blood group epitopes	2BS5	Trimer (6sites)
Ralstonia solanacearum [17]	RS2L	Oligomannose	1UQX	Tetramer
Burholderia cenocepacia [18]	BC2L-A	Oligomannose	2WRA	Dimer
Burholderia cenocepacia[19]	BC2L-C N-ter domain	Lewis Y	2WQ4	Trimer

4.1.2 Bacterial lectins binding to heptoses

In this chapter the focus will be pointed out on a bacterial lectin, Bcla (*B. cenocepacia* lectin A), secreted from the opportunistic pathogen *Burkolderia cenocepacia*. In this Gram negative bacterium, three soluble lectins have been identified, BC2L-A, BC2L-B and BC2L-C, all containing a PA-IIL like domain (cfr 4.1.1), while in two of them there is an additional N-terminal domain without any similarities with known proteins [18, 19].

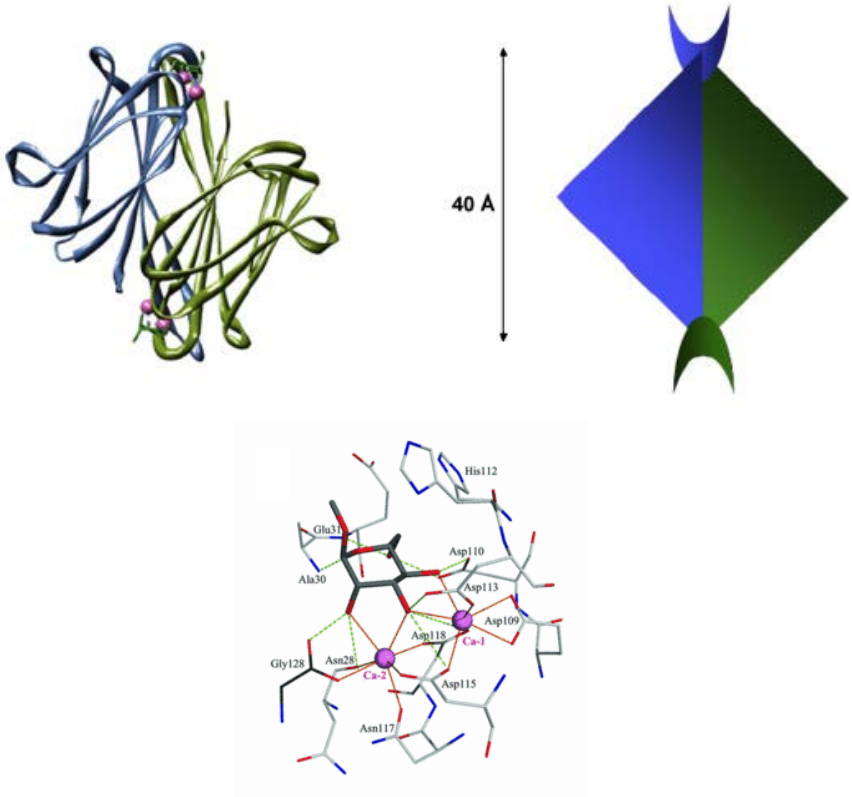


Fig. 4.4: Crystal structure of BclA complexed with Me-α-D-Man (left) and its dimeric architecture (right) [1]. Details of the binding site with co-ordination bonds represented by solid pink lines and hydrogen bonds represented by broken green lines. Ca-1 and Ca-2 indicate the two Ca^{2+} ions.

The smallest *B. cenocepacia* lectin is BC2L-A; it is a recently-characterized soluble bacterial lectin implied in bacterial biofilm formation, particularly in patients suffering from cystic fibrosis. This is a C-type lectin that forms homodimers in physiological conditions with a molecular mass of 13.8 kDa for each monomer, containing one CRD which includes two calcium ions directly involved in ligand binding. As shown in Figure 4.4, the inter-binding site distance is of ~40 Å.

Bcl2A lectin is specific for *manno*-configured sugars and is strictly limited to D-mannose containing oligosaccharides since the particular arrangement of hydroxyl groups in positions 2, 3 and 4 is fundamental for the establishment of hydrogen bonds (taken together to the position 6) to amino acids of the protein binding pocket, and for the coordination to the metal ions (Figure 4.4). Such involvement of a pair of calcium in carbohydrate binding has only been observed for the related lectin PA-IIL from *P. aeruginosa* [20], CV-IIL from *Chromabacterium violaceum* [15] and RS-IIL from *Ralstonia solanacearum* [17]. These three lectins assembled as tetramers, while the *Burkholderia* lectins BC2L-A and the lectin C-terminal domain of BC2L-C are dimeric [19].

Attempts to determine the localization of *Burkholderia* lectins concluded that the three proteins are secreted or released into the extracellular medium [21]. BC2L-B and -C lectins are specifically released from the intact cells upon mannose treatment, suggesting that an important portion is attached to the external envelope of bacteria. BC2L-A was not detected on the surface, but this may be due to its much lower expression level in *B. cenocepacia* [19]. It was further demonstrated that BC2L-C binds efficiently to L-*glycero*-D-*manno*-heptose (L,D-Hep), an abundant component of the *B. cenocepacia* lipopolysaccharide [22]. Heptoses are very important residues, present as components of the core region of LPS from Gram-negative bacteria, covalently linked to the glycolipid portion, the lipid A [23] (cfr 1.1.3, Figure 1.7). In case of enterobacterial

LPS, a heptose trisaccharide fragment: α-L,D-Hep*p*^{III}-(1-7)-α-L,D-Hep*p*^{II}-(1-3)-α-L,D-Hep*p*^{I}) is always present in the inner core region of LPS. In *Burkholderia*, heptoses are ubiquitous and found in different glycosydic linkages. In addition to or instead of L,D-Hep, LPS may contain D-*glycero*-D-*manno*-heptopyranose (D,D-Hep) which is the biosynthetic precursor of L,D-Hep. In a limited number of cases, heptoses have been detected as components of repeating units of O-antigens, such as in *Yokenella regensburgei* [24], whereas in a strain of *Agrobacterium* it is the only LPS O-antigen carbohydrate component [25]. Interaction between plant lectins and LPS has been proposed to play a role in the establishment of symbiosis between legume plants and nitrogen-fixing bacteria [26]. The interaction between human lectins of the innate immune system, such a as surfactant protein D (SP-D) and mannan-binding protein (MBP), and pathogenic bacteria is mediated by heptose [27].
In *P. aeruginosa* and *B. cenocepacia*, bacterial lectins interact with LPS heptose on the cell surface and participate in host tissue binding and/or colony aggregation [21]. Despite of these important roles, the number of investigations on lectin-heptose interaction is very small. Heptose has been demonstrated to bind efficiently to Concanavalin A, a classical mannose specific lectin [28] and the crystal structure of surfactant protein D has been solved in complex with heptose [29].

Here we studied the interaction of BC2L-A with heptose and other *manno* configured sugar residues by a combination of structural and thermodynamic approaches [2]. Saturation Transfer Difference NMR spectroscopy, together with X-Ray cristallografy, Titration calorimetry, Microscopy study studies were performed in order to shed light on the BC2L-A recognition properties.

4.2 Epitope mapping of monosaccharides bound to BC2L-A

The analysis of the binding of carbohydrate ligands to BC2L-A was performed by one-dimensional STD-NMR experiments.
STD NMR data obtained from D-mannose (Figure 4.5, see also Table 4.3) and the corresponding α-methyl glycoside (data not shown) show that all protons of the monosaccharide exhibited some degree of enhancement, demonstrating binding of the mannose to the protein. Actually, since the ligand was small enough to completely enter the binding pocket of the protein, all carbohydrate ring hydrogens were close to receptor atoms and gave rise to STD signals. The most prominent STD effects were observed for H-2, H-3 and H-4 protons, all strongly involved in the binding, while the anomeric protons were slightly visible.

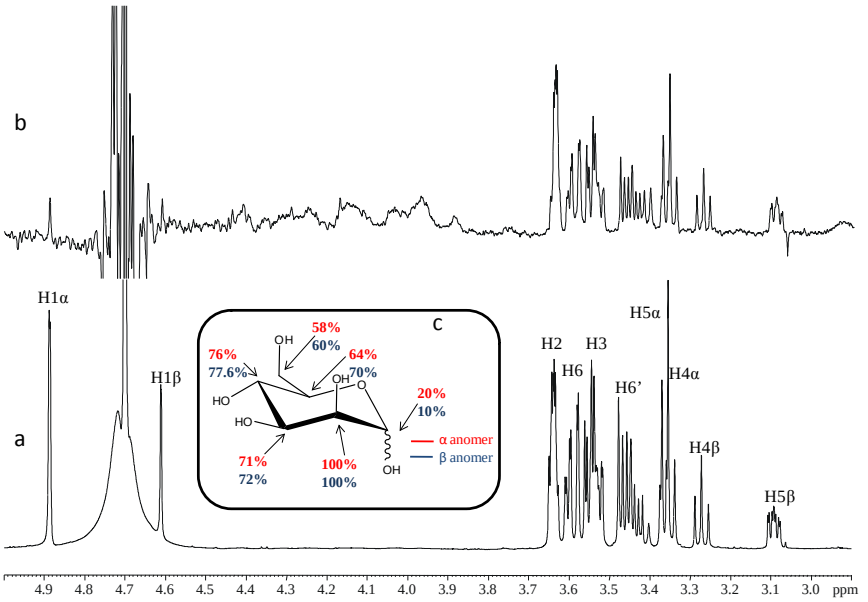


Fig. 4.5: a) Reference ^{1}H NMR spectrum and b) STD 1D NMR spectrum of Bc2lA – D-mannose mixture, 1:60; on resonance irradiation frequency at 0.54 ppm, saturation time of 1.5 seconds. c) Epitope mapping of the mannose in the binding site of BclA; only data from the more abundant anomer have been reported.

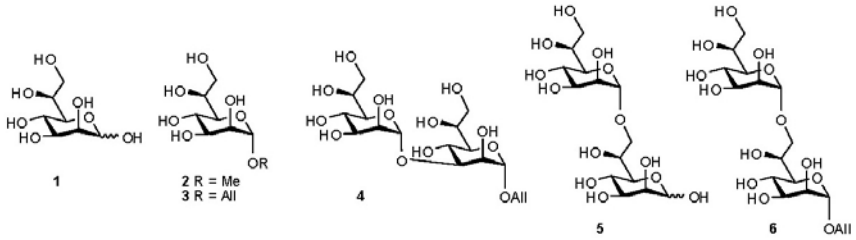
Table 4.3 Chemical shifts of ^{1}H and ^{13}C of the D-mannose.

α-D-Man	1	2	3	4	5	6
^{1}H	4.88	3.64	3.55	3.35	3.36	3.60/3.47
	93.8	70.5	69.9	66.5	72.8	60.8
β-D-Man	1	2	3	4	5	6
^{1}H	4.62	3.64	3.53	3.27	3.09	3.57/3.43
	93.5	70.5	72.2	66.3	75.9	60.7

These data were in agreement with the crystal structures of BC2L-A complexed with αMeMan (Figure 4.3) [1]. Hydrogens at C-5 and C-6 were close to His112 but this residue displayed conformational flexibility, which may explain why the saturation transferred to this proton was not as efficient as for the other ones.

In order to investigate the effect of different groups at C-5, STD-NMR experiments were conducted on methyl L-*glycero*-α-D-*manno*-heptoside (αMeHep), that presented a glycol group at C-5 (Scheme 4.1), and rhamnosides with a methyl group at the same position.
Figure 4.6 showed ^{1}H and 1D STD NMR spectra obtained with αMeHep (see Table 4.4). By comparing the STD spectrum with its corresponding reference experiment, it could be clearly inferred that the monosaccharide was able to bind to BC2L-A, as showed by the enhancement of several signals. In fact, the epitope mapping revealed that, beside the expected interacting H-2, H-3 and H-4 proton signals, the region comprising H-5/ H-7 was close to the receptor and contributed to the overall binding.

Scheme 4.1. Schematic representation of heptose and derivatives.



Oligosaccharide 1 (from *B. cenocepacia* K56-2*Waal* mutant)

L,D-Hep-(1→7)-L,D-Hep-(1→3)-L,D-Hep-(1→5)-D-Kdo
3 4 4
↑ ↑ ↑
1 1 2
L,D-Hep-(1→2)- α-D-Glc β-D-Glc D-Ko-(8←1)Ara4N

Oligosaccharide 2 (from *Burkholderia multivorans*)

α-L,D-Hep
1
↓
7
β-D-QuiNAc-(1→3)-α-D-Rha-(1→2)-α-L,D-Hep-(1→3)-α-L,D-Hep-(1→5)-α-D-Kdo
3 4 4
↑ ↑ ↑
1 1 2
α-L-Rha β-D-Glc α-D-Ko
3 6
↑ ↑
1 1
β-D-Glc-(1→3)-α-D-GalNAc -(1→3)-β-D-GalNAc α-L,D-Hep

Polysaccharide (*Agrobacterium radiobacter*)

[α-L,D-Hep-(1→3)-]$_n$

Table 4.4 Chemical shifts of ^{1}H and ^{13}C of the O-α-methyl-L,D-heptose

α-Met-L-D-Hep	1	2	3	4	5	6	7	CH3
^{1}H	4.47	3.62	3.46	3.56	3.28	3.75	3.47/3.41	3.08
^{13}C	100.7	69.7	70.5	65.5	70.7	68.4	62.2	54.3

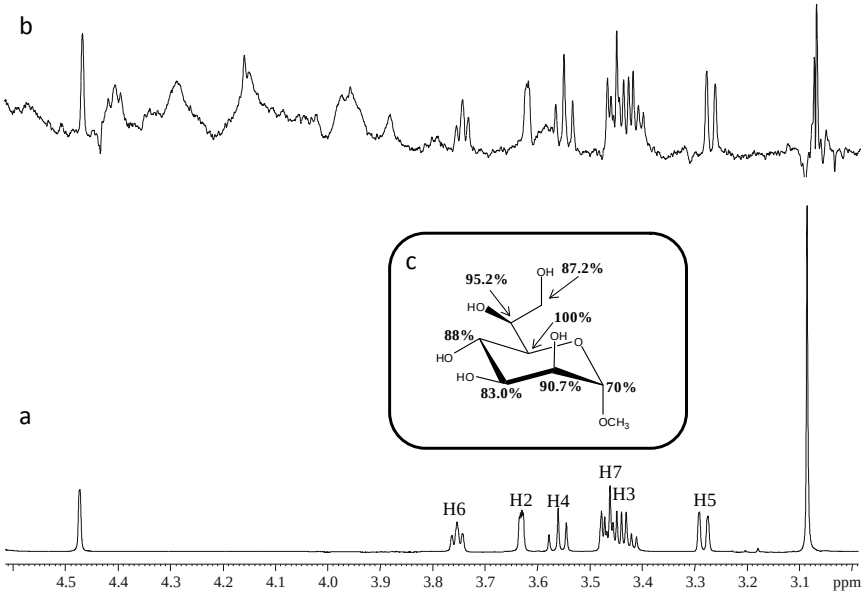


Fig. 4.6: a) Reference ^{1}H NMR spectrum and b) STD 1D NMR spectrum of BC2LA/αMeHep mixture, 1:60; on resonance irradiation frequency at 0.54 ppm, saturation time of 2.5 seconds. c) Epitope mapping of the αMeHep in the binding site of BclA.

Further STD NMR experiments were carried out to investigate the direct binding of 6-deoxy-sugars to BC2L-A, confirming that the terminal ring protons represented a structural requirement for binding to the lectin. In fact, the STD NMR spectrum obtained with methyl-α-D-rhamnoside did not display signals (Figure 4.7 and Table 4.5), indicating that the interaction was much weaker when no hydroxyl groups are present on position 6. No STD signals were observed in the interaction of BC2L-A with L-rhamnose (Figure 4.8 and Table 4.5), that exists in the $^{1}C_{4}$ chair in solution.

Table 4.5 Chemical shifts of ^{1}H and ^{13}C of the L-rhamnose.

L-α-Rha	1	2	3	4	5	6
^{1}H	4.82	3.63	3.50	3.59	3.14	1.94
^{13}C	93.7	70.52	69.6	68.0	71.8	30.2
L-β-Rha	1	2	3	4	5	6
^{1}H	4.58	3.63	3.31	3.57	3.06	1.94
^{13}C	93.3	70.52	72.5	68.0	71.5	30.2

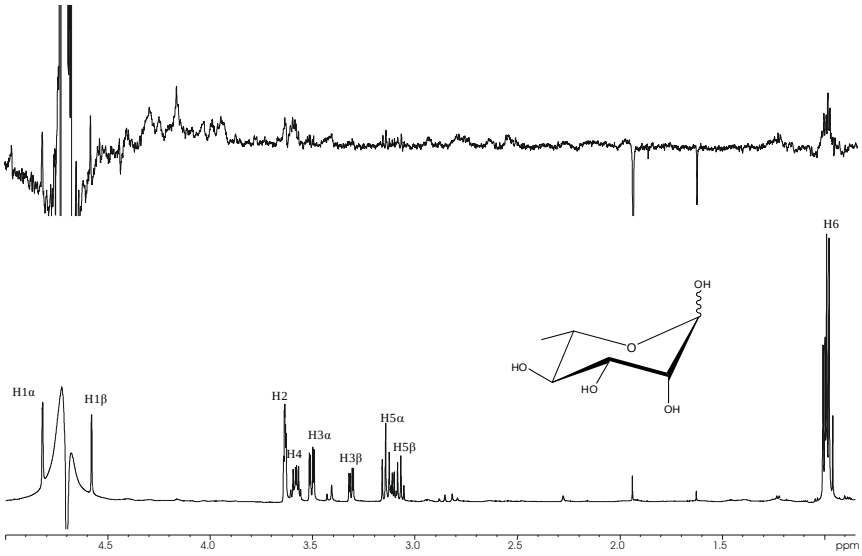


Fig. 4.7: a) Reference ^{1}H NMR and b) STD 1D NMR spectra of Bc2lA L-rhamnose mixture,1:60; on resonance irradiation frequency at 6.8 ppm ppm, saturation time of 1.5 seconds. c) Structure of L-rhamnose, that exists in $^{1}C_{4}$ chair in solution.

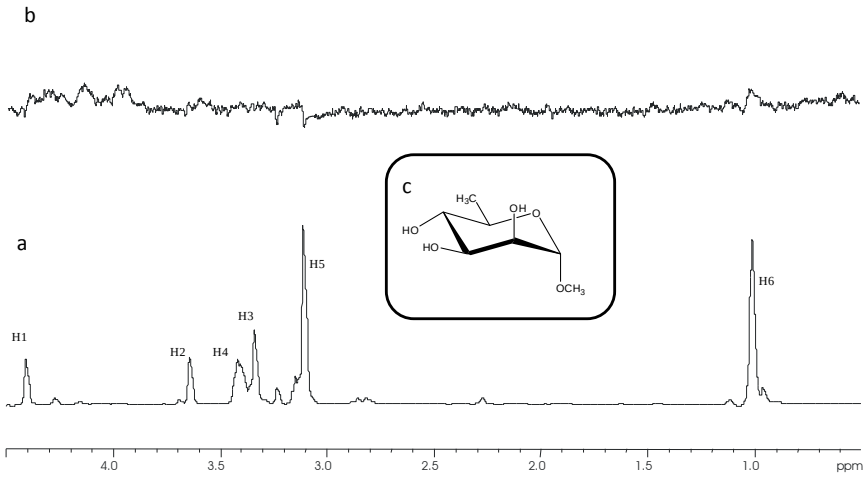


Fig. 4.8: a) Reference ^{1}H NMR and b) STD 1D NMR spectra of Bc2lA Methyl-α-D-rhamnoside mixture,1:60; on resonance irradiation frequency at 0 ppm ppm, saturation time of 2 seconds. c) Structure of L-rhamnose, that exists in $^{4}C_{1}$ chair in solution.

4.3 Epitope mapping of disaccharides bound to BC2L-A

We have also analyzed the interaction between BC2L-A protein and disaccharides containing L-*glycero*-D-*manno* heptoses. In Figures 4.9 and 4.10, STD NMR spectra performed using α-(1→3)-heptoside and α-(1→7)-heptose disaccharides have been reported. In both cases, the observed line broadening was attributable to an increased relaxation rate of the protein-ligand complex, indication that the disaccharides were in contact with the receptor. For αHep1-3Hep (Figure 4.9, Table 4.6), due to a significant spectral overlap, a quantitative analysis of the STD signal intensities was difficult, whereas 2D STD NMR experiments were impaired by physical characteristic of protein, as for example its low molecular weight.

Table 4.6: Chemical shifts of ^{1}H and ^{13}C of the α-(1→3)-heptose disaccharide.

Ring A	1	2	3	4	5	6	7			
^{1}H	4.86	3.76	3.60	3.59	3.35	3.75	3.45/3.39			
^{13}C	102.4	69.7	70.04	65.7	71.2	68.4	62.5			
Ring B	1	2	3	4	5	6	7	O1 CH_2	O2 CH	O3 CH_2
^{1}H	4.60	3.76	3.61	3.69	3.40	3.75	3.45/3.39	3.92/3.76	5.68	5.07/4.99
^{13}C	98.8	69.7	77.6	65.4	71.4	68.4	62.5	67.9	133	117.9

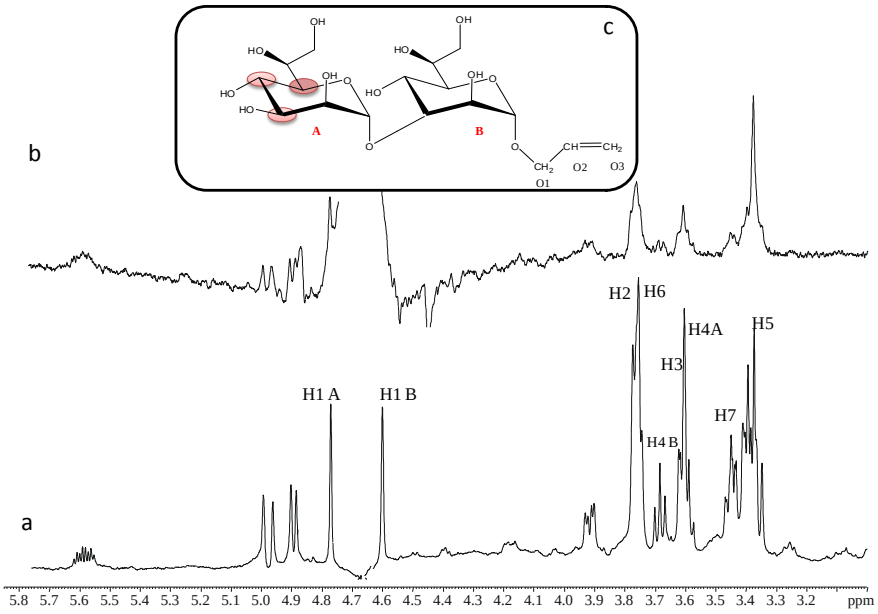


Fig. 4.9: a) Reference ^{1}H NMR and b) STD 1D NMR spectra of Bc2lA/Hepα1-3Hep disaccharide mixture, 1:60; on resonance irradiation frequency at 7.5 ppm, saturation time of 2 seconds. c) Qualitative epitope mapping of Hepα1-3Hep. The STD intensity increases with the color intensity of the circle.

Nevertheless, from a qualitative analysis of the spectrum, performed comparing the STD NMR spectrum with its

corresponding reference, it was evident that H-5, H-4 and H-3 protons, belonging to ring **A** and almost overlapping, were those receiving highest transfer of magnetization and were thus strongly involved in the interaction.

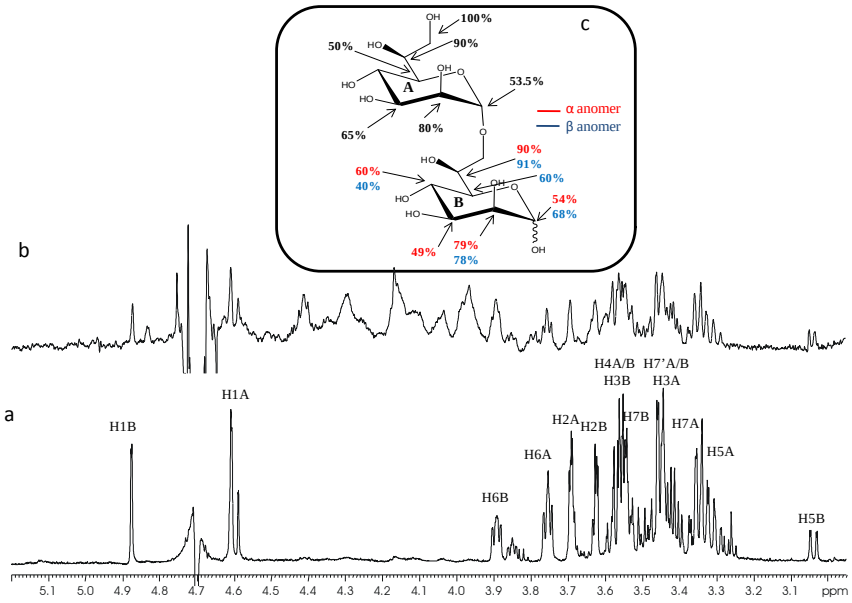


Fig. 4.10: a) Reference ^{1}H NMR and b) STD 1D NMR spectra of Bc2lA O-Methyl-α-(1→7)-heptoside disaccharide mixture, 1:60; on resonance irradiation frequency at -0.5 ppm, saturation time of 2 seconds. c) Epitope mapping of the α-(1→7)-heptose disaccharide.

Table 4.7 Chemical shifts of ^{1}H and ^{13}C of the O-Metyl-α-(1→7)-heptose disaccharide.

Dis 1→7 Residuo A	1	2	3	4	5	6	7	
^{1}H	4.61	3.70	3.45	3.56	3.32	3.75	3.44/3.39	
^{13}C	100.56	69.52	70.4	65.8	71	68.4	62.6	
Dis 1→7 Residuo B α	1	2	3	4	5	6	7	CH3
^{1}H	4.47	3.63	3.54	3.57	3.25	3.89	3.43/3.49	3.09
^{13}C	100.72	69.9	70.5	65.7	71.9	67.04	69.2	54.46

Results obtained from a quantitative analysis performed on the αHep1-7Hep indicated that the largest fraction of saturation in the

disaccharide moiety was received by protons H-6 and H-7 belonging to ring **A** (Figure 4.10, Table 4.7). In both disaccharides, it can be therefore concluded that the non reducing unit is the one that plays the stronger role in the binding event.

4.4 Affinity measurements and thermodynamics by titration microcalorimetry

The association constant, Ka, of the equilibrium of binding can be related to the thermodynamic parameter ΔG that corresponds to the Gibbs free energy:

$$\Delta G = -RT\ln Ka = \Delta H - T\Delta S \quad \text{[Equation 4.1]}$$

The free energy of binding between a carbohydrate and a lectin, ΔG, is characterized by an enthalpic ΔH term and an entropic ΔS term, dependent to the temperature of the system T, expressed in Kelvin (R represents the gas constant). ΔG must be negative for interactions that occur spontaneously, with a negative, exothermic, ΔH term and a positive ΔS entropic term. When the exothermic enthalpy term is dominant, like in almost all carbohydrate-lectin interactions, the reaction is called enthalpy-driven [3]. It means that the formation of the complex is exothermic, favored by the establishment of polar/non polar contacts between ligand and protein, and, in some cases, by the formation of ion-mediated contacts.

L,D-Heptose and derivatives were tested by titration microcalorimetry in order to determine the affinity for BC2L-A (Table 4.8). In all cases, the interaction resulted in exothermic peaks as displayed on the examples in Figure 4.11.

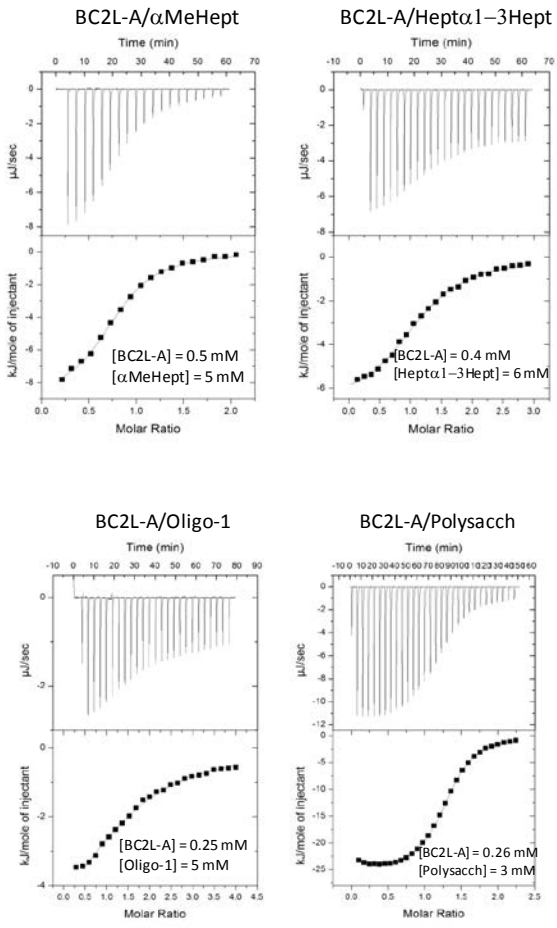


Fig. 4.11: Microcalorimetry data. Upper paneln of each figure: thermogram of monosaccharides, disaccharides, LPS oligosaccharides and polysaccharides titrating BC2L-A. Lower panel: integration of data with curve fitted for "one binding site" model.

All curves could be fitted with a one-site model, although the binding of the reducing monosaccharide L,D-heptose showed some cooperativity beween the two monomers, as previously described for mannose binding [1]. The reducing monosaccharide displayed medium range affinity (Kd = 137 µM), while the introduction of a methyl group in α-configuration enhanced the affinity (Kd = 53.8

μM). The configuration of the glycol tail is crucial, since the D,D-heptose monosaccharide was not recognized by BC2L-A. The α1-3 and α1-7 linked disaccharides behaved approximately as the αMeHep monosaccharides, with a slightly higher affinity of BC2L-A for the Hep α1-7 Hep disaccharide. The α-methyl derivative of this disaccharide was the best ligand of this series with a Kd of 18 μM. The observed affinity for heptose residues was one order of magnitude weaker to what was observed for the binding of BC2L-A to αMeMan [1].

Table 4.8 Microcalorimetry data for the interaction of Bc2lA with monosaccharides and disaccharides.

Ligand	K_A (10^4 M^{-1})	K_D (μM)	n	-ΔG (kJ/mol)	-ΔH (kJ/mol)	TΔS (kJ/mol)
D-Man[a]	19.4	5.1	0.91	30.2	27.0	3.2
αMeMan[a]	36.4	2.7	0.83	31.8	23.0	8.8
L,D-Hep	0.73 +/-0.3	137	0.91 +/-0.05	22.1	10.4 +/- 0.7	11.7
LαD-MeHep	1.86 +/-0.09	53.8	0.79 +/-0.03	24.4	9.7 +/- 0.2	14.7
LαD- AllylHep	1.35+/-0.01	74.1	0.94 +/-0.04	23.5	12.7+/-1.1	10.8
D,D-Hep	No binding					
Hep α1-3Hep-O-Allyl	1.41+/-0.01	70.7	1.07+/-0.05	23.7	6.9 +/- 0.3	16.8
Hep α1-7Hep	2.29+/-0.07	43.8	01.11+/-0.02	24.9	10.5+/- 0.5	14.4
Hep α1-7Hep-O-Met[b]	5.5+/-0.3	18.3	0.87+/-0.09	27.0	12.6+/- 0.8	14.4

[a] From 17.[b] Large standard deviations and poor reproducibility due to strong cooperative effect.

Analysis of the thermodynamic contribution to the free energy of binding demonstrated a favorable role of the entropy term, with negative values of -TΔS. The binding was equally driven by enthalpy and entropy term for the heptose monosaccharides and derivatives and was clearly entropy driven for the disaccharides. Such favorable entropy of binding is unusual for protein-carbohydrate interactions where a strong entropy barrier is

generally observed [30]. This appears to be a signature of this two-calcium family of lectins since favorable entropy of binding has been observed for all of the members, with a particularly strong contribution for CV-IIL [15].

Table 4.9 Microcalorimetry data for the interaction of Bc2lA with oligosaccharides and polysaccharides.

Ligand	K_A (10^4 M^{-1})	K_D (μM)	n	-ΔG (kJ/mol)	-ΔH (kJ/mol)	TΔS (kJ/mol)
Oligosaccharide 1 ***B. cenocepacia***	0.31+/-0.01	325	1[a]	19.9	10.2 +/-0.4	9.7
Oligosaccharide 2 ***B. multivorans***	0.281[b]	356	1[a]	19.7	11.7	8.0
Polysaccharide (L,D Hep1-3)n ***A. radiobacter***	26.6[b]	3.8	0.951	30.9	24.3	6.7

[a] Stochiometry fixed during the fitting procedure

[b] only one experimental measure

BC2L-A was also tested for its ability to bind to major oligosaccharide fragments from Burkholderia LPS (Table 4.9). Direct binding to whole LPS could not be observed by microcalorimetry, presumably because of solubility issue and/or accessibility of the heptose residues. When working with delipidated LPS, such as core oligosaccharides from *Burkholderia multivorans* and *Burkholderia cenocepacia*, a good affinity was observed with heptose-carrying oligosaccharides with Kd in the range of 300 μM. Interestingly, BC2L-A could also interact with polysaccharides from other bacterial species. The linear polysaccharide from *Agrobacterium radiobacter* is a homopolymer of α1-3 linked L,D-heptose residues. The stoichiometry of binding indicated that BCL2-A bound to only one residue in the whole polysaccharide, presumably the one at the non-reducing end since

free O-3 is required for coordination of calcium ions in the lectin binding site. A strong affinity was observed (Kd = 3 μM) with an enthalpy of binding significantly higher than for the mono or disaccharides, suggesting a structured conformation of the polysaccharides creating additional contacts with the protein.
Thus, all the results showed that BC2L-A was able to bind to LPS derived from different bacteria and underlined that the αHep1-7Hep disaccharide and a $[\alpha Hep1\text{-}3]_n$ polysaccharide were high affinity ligands.

4.5 Crystal structure of BC2L-A complexed with MeHep

The crystal structure of BC2L-A complexed with α-MeHep confirmed the orientation of the *manno* configured ring in the binding site and the coordination to calcium ions [2]. Indeed, the overall crystal structure and the dimer arrangements (Figure 4.12a) were very similar to the previously described complex between BC2L-A and αMeMan (Figure 4.4) [1].
Clear electron density for one molecule of αMeHep and two calcium ions was observed in each binding site (Figure 4.12b). As observed in other crystal structures of PAIIL-related lectins, three hydroxyl groups of the sugar ring were engaged in coordination contact with the two calcium ions. A dense network of hydrogen bonds involved the O2, O3, O4 and O6 hydroxyl groups of heptose similar to what has been observed in BC2LA/αMeMan complexes and corresponded to a conformation of the ω1 torsion angles (O5-C5-C6-O6) close to -40°, corresponding to energetically favored *gauche-gauche* conformation (gg) according to Marchessault and Pérez [31].

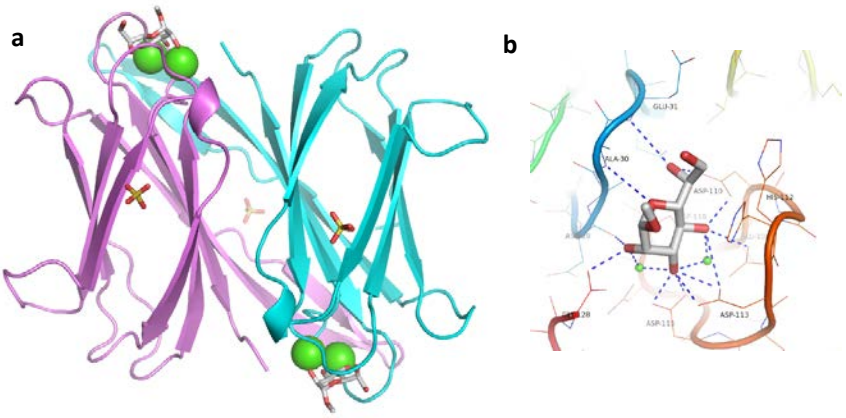


Fig. 4.12: (a) representation of ribbon diagram of BclA dimer complexed with αMeHep (sticks) and calcium (spheres). Sulfate ions are represented by sticks. (b) Electron density (1σ = electron per $Å^2$) around the αMeHep with hydrogen bonds to amino acids in binding site represented by dashed lines.

4.6 Binding of BC2L-A to surfaces of *Burkholderia cenocepacia*

BC2L-A labeled with FITC was used in order to determine its ability to bind to surfaces of *B. cenocepacia* LMG 16656 (J2315). Cells and biofilms were incubated with BC2L-A-FITC and examined using fluorescence microscope. A visible fluorescence at the cell surfaces was observed (Figure 4.13a). *Escherichia coli* strain BL21(DE3) was used as negative control and BC2L-A-FITC was not able to bind to and stain these cells. When mixed together, *B. cenocepacia* cells were clearly distinguished from *E. coli* by BC2L-A-FITC staining. Experiment was performed also with *Burkholderia cenocepacia* biofilm scraped from cultivation flasks (Figure 4.13b) and with undisturbed biofilm growing directly on microscope slide. Both *Burkholderia cenocepacia* biofilms also provided clear fluorescent signal, distribution of fluorescence within biofilms was uniform.

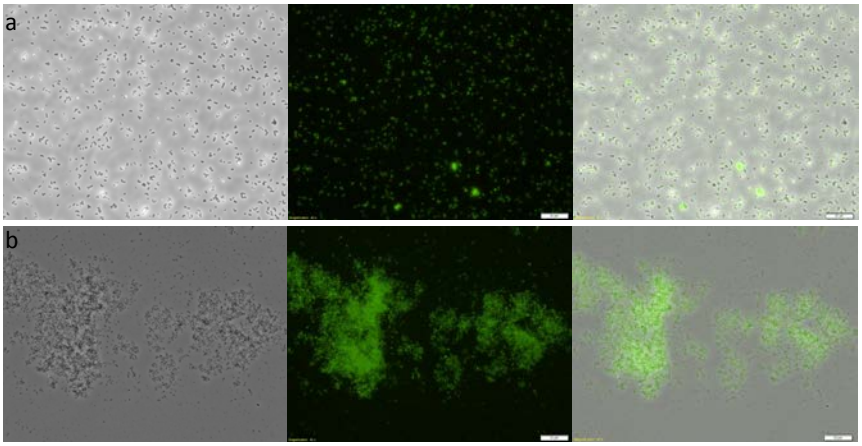


Fig. 4.13: (a) cells and (b) biofilm on glass slides or *B. cenocepacia* LMG 16656 (J2315) stained by BC2L-A-FITC. Bright field (left), green fluorescence (middle), superposition (right). Bar 20 µm.

Therefore, microscopy studies confirmed the binding of fluorescent lectin on both bacterial surface and biofilm.

4.7 Conclusion

In this study, a sequence of advanced techniques such as NMR spectroscopy, X-ray crystallography and titration microcalorimetry has disclosed the binding and epitope patterns of LPS heptose and other *manno* configured sugar residues to the lectin BC2L-A, produced by the opportunistic pathogen *Burkholderia cenocepacia.* LPS are the major components of the outer membrane of Gram-negative bacteria and exert a variety of biological activities in animals and plants. Their biosynthesis is indispensable for growth and survival of Gram-negative bacteria and comprise about 10-15% of the total molecules in bacterial outer membranes covering about 75% of bacterial surface area, thus they are exposed to the external environment. They represent a defensive barrier which helps bacteria to resist to antimicrobial compounds and environmental stresses and are involved in many aspects of host–bacterium interactions as recognition, adhesion and colonization. Together

with other indispensable microbial products they are taken by the host organism as hallmark of microbial presence, and are hence referred to as Microbe Associated Molecular Patterns.

Both X-ray crystallography and titration calorimetry showed a good affinity of heptose(s) residues for the BclA lectin, in particular of the non reducing end heptose residue, which arranges the angles (O5-C5-C6-O6) to the energetically favored *gg* conformation. By titration microcalorimetry it was also possible to observe binding of the lectin to major oligosaccharides such as those deriving from R-LPS from two *Burkholderia* strains, and to a another bacterial polysaccharide fully composed of heptose residues. In accordance, STD NMR spectroscopy studies of BC2L-A with different mono- and disaccharides indicated that methyl L-*glycero*-α-D-*manno*-heptoside (αMeHep) and D-mannose were both in contact with the protein binding pocket. No STD signals were observed for of 6-deoxy-sugars L- and D-rhamnose, highlighting the inability of the protein to recognize and bind substrates lacking hydroxyl groups on position 6, independently on the chair conformation ($^1C_4$ or $^4C_1$). These data were confirmed by STD NMR experiment on α-(1→3) and α-(1→7) heptose disaccharides, both present in the LPS of *Burkholderia*, in which, the binding of the terminal heptose was more effective, testifying a major role in the interaction with the protein.

To conclude, we have here demonstrated that a lectin secreted by *Burkholderia cenocepacia* was able to bind particular and peculiar saccharide residues present in the majority of LPS core oligosaccharides. This recognition can be pivotal in bacterial social life such as quorum sensing, exit from dormancy, formation of microcolony and/or any other processes in which the detection of specific bacterial strain is necessary.

References

[1] Lameignere, E.; Malinovska, L.; Slavikova, M.; Duchaud, E.; Mitchell, E. P.; Varrot, A.; Sedo, O.; Imberty, A.; Wimmerova, M. *Biochem. J.*, **2008,** 411, (2), 307-18.
[2] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*, **2012.** 10,1387-98.
[3] Sharon, N.; Lis, H., *Lectins*. Second ed.; Kluwer Academic Publishers: Dordrecht, **2003**.
[4] Imberty, A. *Synthesis and Biological Applications of Glycoconjugates*, **2011**. 3-11
[5] Aragao, K. S.; Satre, M.; Imberty, A.; Varrot, A., *Proteins*, **2008.** 73, (1), 43-52.
[6] Cioci, G.; Mitchell, E. P.; Gautier, C.; Wimmerova, M.; Sudakevitz, D.; Perez, S.; Gilboa-Garber, N.; Imberty, A., *FEBS Lett.*, **2003.** 555, (2), 297-301.
[7] I. E. Liener, N. Sharon, and I. J. Goldstein, The Lectins: Properties, Functions and Applications in Biology and Medicine, Academic Press, Orlando, **1986**.
[8] Sharon, N. and Lis, H. Lectins, Chapman and Hall, London, **1989**.
[9] Lis, H and Sharon, N *Chem. Rev*., **1998**. 98, 637.
[10] Imberty, A.; Varrot, A. *Curr. Opin. Struct. Biol.*, **2008**. *18*, 567-576.
[11] Turnbull, W. B.; Precious, B. L.; Homans, S. W., *JACS*, **2004.** 126, (4), 1047-54.
[12] Gilboa-Garber, N. *Pseudomonas aeruginosa* lectins. *Methods Enzymol.*, **1982**, *83*, 378-385.
[13]Blanchard, B.; Nurisso, A.; Hollville, E.; Tétaud, C.; Wiels, J.; Pokorná, M.; Wimmerová M.; Varrot, A.; Imberty, A. *J. Mol. Biol.*, **2008**. *383*, 837-853.
[14] Perret, S.; Sabin, C.; Dumon, C.; Pokorná, M.; Gautier, C.; Galanina, O.; Ilia, S.; Bovin, N.; Nicaise, M.; Desmadril, M.; Gilboa-Garber, N.; Wimmerova, M.; Mitchell, E.P.; Imberty, A. *Biochem. J.*, **2005**. *389*, 325-332.
[15] Pokorná, M.; Cioci, G.; Perret, S.; Rebuffet, E.; Kostlánová, N.; Adam, J.; Gilboa-Garber, N.; Mitchell, E.P.; Imberty, A.; Wimmerová, M. *Biochemistry (Mosc)*, **2006**. *45*, 7501-7510.

[16] Kostlanová, N.; Mitchell, E.P.; Lortat-Jacob, H.; Oscarson, S.; Lahmann, M.; Gilboa-Garber, N.; Chambat, G.;
Wimmerová, M.; Imberty, A. *J. Biol. Chem.*,**2005.** *280*, 27839-27849.
[17] Sudakevitz, D.; Kostlanova, N.; Blatman-Jan, G.; Mitchell, E.P.; Lerrer, B.; Wimmerova, M.; Katcoff, D.J.; Imberty, A.; Gilboa-Garber, N. *Mol. Microbiol.*, **2004**. *52*, 691-700.
[18] Lameignere, E.; Shiao, T.C.; Roy, R.; Wimmerová, M.; Dubreuil, F.; Varrot, A.; Imberty, A. *Glycobiology*, **2010**. *20*, 87-98.
[19] Šulák, O.; Cioci, G.; Delia, M.; Lahmann, M.; Varrot, A.; Imberty, A.; Wimmerová, M. *Structure*, **2010**. *18*, 59-72.
[20] Mitchell, E.; Houles, C.; Sudakevitz, D.; Wimmerova, M.; Gautier, C.; Pérez, S.; Wu AM, Gilboa-Garber, N.; Imberty, A. *Nature Struct Biol*, **2002**. 9, 918–921.
[21] Šulák, O.; Cioci, G.; Lameignère, E.; Balloy, V.; Round, A.; Gutsche, I.; Malinovská, L.; Chignard, M.; Kosma, P.; Aubert, F. et al. *PLoS Pathog*., **2011.** 7, e1002238.
[22] De Soyza, A.; Silipo, A.; Lanzetta, R.; Govan, J.R.; Molinaro, A. *Innate Immun.* **2008**. 14,127–144.
[23] Silipo, A.; Molinaro, A. *Subcell Biochem*., **2010**. 53, 69–99.
[24] Kosma, P. *Curr Org Chem*., **2008**.12,1021–1039.
[25] De Castro, C.; Sturiale, L.; Parrilli, M. *Eur J Org Chem*., **2004**. 2436–2440.
[26] Agrawal, P.; Kumar, S.; Jaiswal, Y.K.; Das, H.R. *Biochimie*., **2011**. 93, 440–449.
[27] Kawasaki, N.; Kawasaki, T.; Yamashina, I. *J Biochem*., **1989**.106, 483–489.
[28] Jaipuri, F.A.; Collet, B.Y.; Pohl, N.L. *Angew Chem Int Ed Engl*. **2008**. 47, 1707–1710.
[29] Wang, H.; Head, J.; Kosma, P.; Brade, H.; Muller-Loennies, S.; Sheikh, S.; McDonald, B.; Smith, K.; Cafarella, T.; Seaton, B. et al. *Biochemistry*. **2008**. 47, 710–720.
[30] Dam, T.K.; Brewer, C.F. *Chem Rev*., **2002**.102, 387-429.
[31] Marchessault, R.H.; Pérez, S.. *Biopolymers*. **1979**. 18, 2369-2374.

SECTION III

CELL WALL GLYCOCONJUGATES AS KEY MEDIATORS IN THE INTERACTION WITH EUKARYOTIC HOST

Chapter 5

rhMBL – carbohydrates from Ebola virus interaction

5.1 Introduction

Recognition of pathogen-specific glycostructures by human lectins on immune cells is an important means of host immune defence. The mannan binding lectin, MBL, makes an important contribution to innate immune protection and works along with epithelial barriers, cellular defenses such as phagocytosis, and patter recognition receptors that trigger pro-inflammatory signaling cascade [1]. Several investigators have studied and demonstrated the therapeutic potential of rhMBL against many pathogens [2-7]. However, due to its extremely complex quaternary structure, unsuitable for industrial scale production given the cost and the difficulty, its possible use in successful anti-infective therapy is limited. In order to gain better and inexpensive therapeutic agents, less complex chimeric fusion proteins have been developed with similar ligand recognition and enhanced effector functions. To this aim, a recombinant human MBL, rhMBL, and three L-ficolin variants that contained the MBL carbohydrate recognition domain and varying lengths of the collagenous domain, have been produced with potential clinical application against glycosylated enveloped viruses [5]. It was also demonstrated that one of these chimeras, L-FCN/MBL76, had superior effector activity that appeared to be mediated by greater structural flexibility.

Here, the mannose binding activity of both rhMBL, and the chimeric molecule, L-FCN/MBL76 [8] has been studied at atomic level with an NMR based approach to study.

5.1.1 The human mannose binding lectin, MBL

The human mannose binding lectin, MBL, is a soluble, collagen like C-type lectin, belonging to the collectin family, that shows the ability to bind certain N-glycosylated carbohydrates via calcium-dependent CRDs (Carbohydrate Recognition Domain). It is composed by three identical polypeptides that form a helical homotrimer further arranged in octadecameric and larger oligomers with a very complex quaternary structure. The basic structural unit is a homotrimer of three identical polypeptides, each approximately 30 kDa [9] (Figure 5.1).

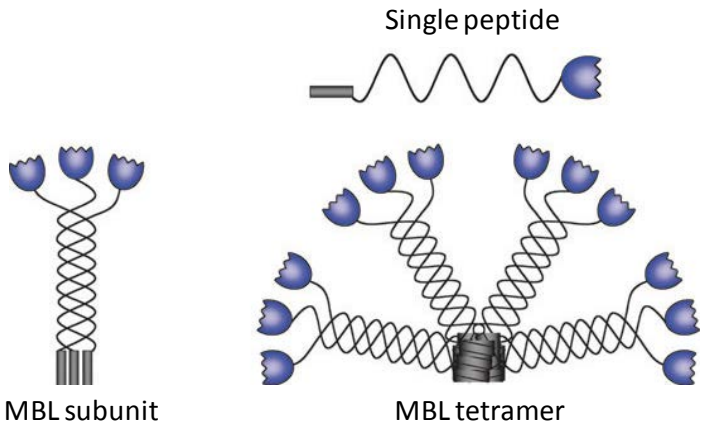


Fig. 5.1: The basic structure subunit of MBL formed by three polypeptide chains. MBL oligomers formed by multiple subunits (Adapted from [1]).

Each chain consists of four domains: a cysteine rich N-terminal collagenous domain that forms stabilizing disulfide bonds, a second collagen-like stalk region, a third neck region that favors an α-helical coiled coil and a fourth C-terminal carbohydrate recognition domain responsible for binding to polysaccharide structures on pathogens. In serum, MBL circulates as oligomers of trimeric subunits ranging from dimers to hexamers. Studies of both plasma-derived MBL and recombinant human MBL (rhMBL) have shown

that larger oligomers bind more avidly to carbohydrate ligands and are more potent activators of complement than smaller oligomers [10]. In humans, the predominant circulating form of MBL is the hexameric complex (eighteen polypeptide chains) [11].

Like all other collectins, MBL shows selective and calcium (Ca^{2+})-dependent binding to terminal sugars like D-mannose, L-fucose and N-acetyl-D-glucosamine (GlcNAc), but not to D-galactose and sialic acid [12,13]. In particular, the presence of several CRDs on a single molecule of MBL enables recognition and high avidity interaction with sugar units. This selectivity relies on the presence of conserved amino acid residues within their CRDs. Furthermore, optimal binding between the CRD complex and a ligand is achieved when the distance between CRDs is 45 Å [14]. The capacity of MBL to recognize conserved sugar structures that decorate diverse microorganisms enables MBL to bind to a wide range of pathogens including bacteria, fungi, parasites and viruses [15-17]. For this reason, it is believed that MBL deficiency contributes to increased susceptibility to infections in humans [18,19]. However, the complex quaternary structure of MBL limits cost effective large scale production. In order to produce structurally simpler molecules, Michelow et al [5] designed three variant recombinant chimeric proteins that comprise fused complementary regions of MBL CRD and the collagenous domain of L-ficolin (L-FCN), another innate immune molecule of the lectin pathway (Figure 5.2).

They took advantage of the fact that both MBL and L-FCN act as opsonins and activate complement but L-FCN forms smaller oligomers [20]. Detailed investigations confirmed that the chimeric proteins retained the binding repertoire of MBL and the ability to activate complement but presented a less complex quaternary structure like that of L-FCN. These fusion proteins were shown to be potent inhibitors of certain glycosylated viruses such as lethal Ebola viruses [5,6] as well as influenza A virus (IAV) [7]. Specifically, L-FCN/MBL76 had superior effector activity

compared with rhMBL that appeared to be mediated by greater structural flexibility.

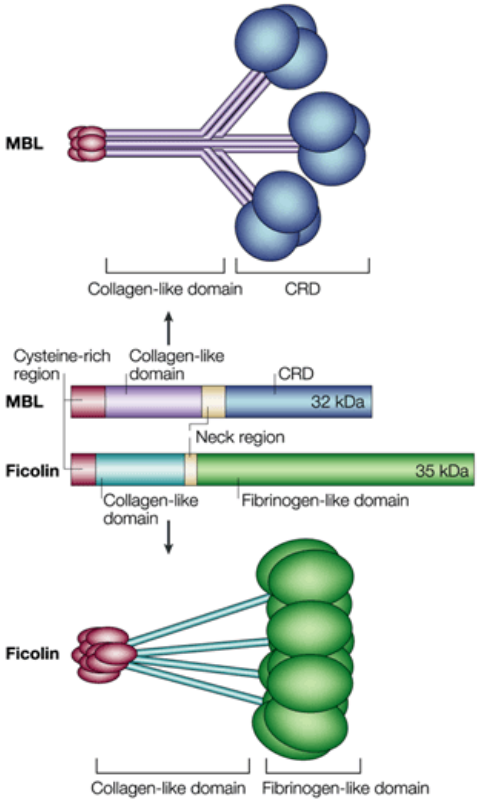


Fig. 5.2: Differences between MBL and ficolin. (Adapted from N*ature Reviews Immunology* 2, 346-35; 2002, doi:10.1038/nri800).

Innate immunity represents humans' first line of defense against pathogens. Certain immune effectors have been conserved throughout evolution dating back to the plant kingdom. This ancient system is exemplified by C-type lectins, which have developed a sophisticated ability to differentiate between infectious agents and self. Lectins do so by recognizing invariant exposed structures adorning pathogens, so-called Pathogen-Associated Molecular Patterns, such as carbohydrates and acetylated

compounds, among others. The precise dynamics of binding between these human pattern recognition receptors and their ligands are poorly understood, however, given the broad therapeutic potential of collectins against certain virulent glycosylated pathogens, it is important to gain a finer molecular characterization of the interactions between lectins and their cognate ligands. To this aim, we used advanced nuclear magnetic resonance spectroscopy to evaluate binding dynamics and affinity between rhMBL in the first instance as well as L-FCN/MBL76 and hexose carbohydrate ligands, which typify the carbohydrates on numerous pathogens. Specifically, we characterized the interaction of molecules at atomic resolution furnishing binding epitopes by Saturation Transfer Difference NMR spectroscopy [21-23]. We used this information to map interacting epitopes on both rhMBL and chimeric L-FCN/MBL76 proteins. In parallel, we employed Transferred NOE NMR spectroscopy to further reveal binding patterns [24-26].

5.2 Binding epitopes of monosaccharide substrates to rhMBL

1D STD NMR experiments on monosaccharides have been performed using *O*-methyl-α-D-mannoside (*a-O*-Me-Man) and N-acetylglucosamine (GlcNAc) (Table 5.1 and 5.2).

Table 5.1 Chemical shifts of ^{1}H and ^{13}C of the α-D-O-methyl mannose.

α-D-O-methyl mannose	1	2	3	4	5	6	CH_3
^{1}H	4.77	3.94	3.77	3.66	3.62	3.90/3.78	3.41
^{13}C	93.52	70.8	71.5	67.8	73.4	62.1	55.5

Table 5.2 Chemical shifts of ^{1}H and ^{13}C of the N-acetylglucosamine.

D-α-GlcNAc	1	2	3	4	5	6	CH_3
^{1}H	5.19	3.88	3.74	3.47	3.83	3.9/3.8	2.03
^{13}C	90.7	54.3	70.7	70.13	71.6	60.7	22.07
D-β-GlcNAcβ	1	2	3	4	5	6	CH_3
^{1}H	4.68	3.66	3.52	3.44	3.44	3.78/3.73	2.03
^{13}C	94.9	56.8	73.9	69.8	76.1	60.61	22.07

To determine the binding epitope of each ligand, relative STD effects have been calculated from the STD amplification factors obtained from the STD spectra. The relative STD effects reflect the relative amount of saturation transferred to the ligand and thus protons with high STD values are assumed in closer contact with the protein.

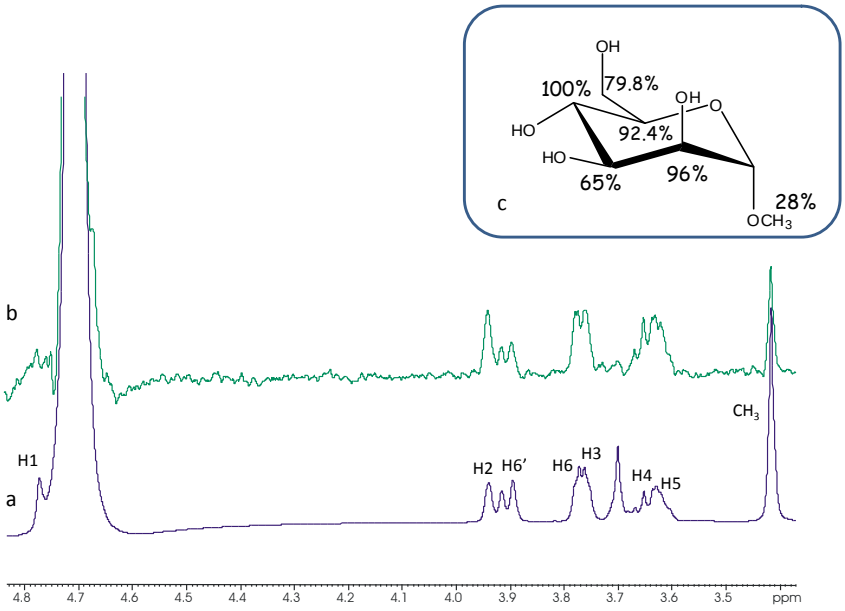


Fig. 5.3: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture rhMBL: α-D-metylmannose 1:50. c. Chemical structure and epitope binding of α-D-metylmannose to rhMBL; protons with highest relative STD and, thus, in closest proximity to the receptor are highlighted in red.

A qualitative and quantitative analysis performed on NMR data obtained from *O*-methyl-α-D-mannose (*a*-*O*-Me-Man)

demonstrated the intimate contact of the ligand with the protein (Figure 5.3a-b).

By comparing STD spectrum with its corresponding reference experiment, it could be clearly inferred that the monosaccharide bound to rhMBL, as showed by several STD enhancements. The degree of enhancement of α-*O*-Me-Man protons in the STD spectrum (Figure 5.3) demonstrated that the entire molecule was bound to the protein, as expected since rhMBL is a mannose binding lectin and the substrate small enough to completely fit in the protein binding site. STD enhancements were greatest for H-4 proton; H-2 and H-5 were also strongly involved into the binding, while the anomeric proton was slightly bound (Figure 5.3c).

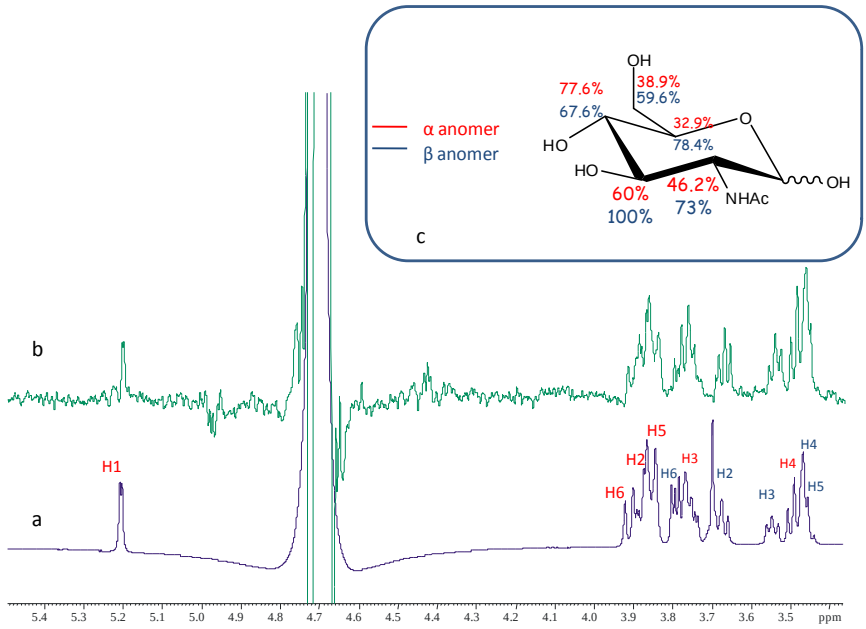


Fig. 5.4: Reference ^{1}H NMR spectrum (a) and 1D STD NMR spectrum (b) of mixture rhMBL: GlcNAc 1:100. c. Chemical structure and epitope binding of GlcNAc to rhMBL; protons with highest relative STD and, thus, in the closest proximity to the receptor are highlighted in red.

rhMBL's capacity of binding other substrates was analyzed by studying the STD spectrum of GlcNAc (Figure 5.4a-b) and the

resulting binding epitope, shown in Figure 5.4c. From STD spectra it could be inferred that rhMBL protein bound both α- and β-anomers, although it exhibited a higher preference for the β-anomer. In detail, epitope mapping revealed that H-2 and H-3 showed the largest STD variation, and thus a strong contribution to the binding indicative of their most intimate contact with the protein surface.

5.3 Binding epitopes of α-(1→2) and α-(1→3) mannose disaccharides to rhMBL and L-FCN/MBL76

Since many viruses display on their cell envelope high-mannose containing glycoproteins, the binding affinity of rhMBL for α-(1→2) (**1**) and α-(1→3) (**2**) mannose disaccharides (Table 5.3 and 5.4) was investigated and then compared with that of the chimeric L-FCN/MBL76 protein, in order to underline possible binding differences. In STD NMR spectra clear signals were observed; a detailed analysis of the spectra provided more insights on the recognition mechanisms of rhMBL and L-FCN/MBL76.

Table 5.3 Chemical shifts of ^{1}H and ^{13}C of the disaccharide **1.** (Only the more abundant reducing anomer has been reported).

1 Ring A	1	2	3	4	5	6
^{1}H	4.92	3.95	3.74	3.51	3.64	3.75/3.60
^{13}C	102.1	69.9	70.1	66.7	73.3	61.0
1 Ring B	1	2	3	4	5	6
^{1}H	5.27	3.83	3.68	3.57	3.82	3.74/3.64
^{13}C	92.4	79.0	72.3	66.9	69.7	60.8

Table 5.4 Chemical shifts of ^{1}H and ^{13}C of the disaccharide **2**.

2-Ring A	1	2	3	4	5	6
^{1}H	5.05	3.98	3.81	3.58	3.68	3.77/3.64
^{13}C	102.2	69.9	70.2	66.7	73.18	60.7
2-Ring B α	1	2	3	4	5	6
^{1}H	4.56	3.99	3.86	3.67	3.71	3.82/3.68
^{13}C	93.8	70.7	77.8	66.1	73.1	60.7
2-Ring B β	1	2	3	4	5	6
1H	4.82	4.01	3.68	3.60	3.32	3.82/3.68
13C	93.36	70.6	80.2	66.1	75.9	60.7

5.3.1 The interaction of the disaccharide **1** with rhMBL. STD NMR and tr-NOESY analysis

STD NMR analysis performed on disaccharide **1** showed that the most prominent STD signals belonged to the terminal ring A (Figure 5.5), which is expected to be strongly involved into the binding with rhMBL.

Qualitatively, it could be assumed that larger STD signals were observed for H-2 and H-3 protons of non-reducing mannose A, indicating smaller distances between those ligand protons and the receptor surface in the bound state. The data obtained by comparing the relative percentages of saturation received by the different ligand moieties showed that signal enhancements belonging to mannose B were also present, although these protons received lower transfer of magnetization from the protein and thus their individual contributions were weaker (see group epitope mapping, Figure 5.5c).

In conclusion, STD-derived data showed, in the case of rhMBL, that the mannose disaccharide was oriented at the Ca^{2+} site in a way that it was mainly the non-reducing unit interacting with the protein. The ability of other MBLs to interact with both sugar moieties, due to the existence of a second binding site and/or of a dual binding mode of mannose oligosaccharide to MBL, has been underlined (cfr 5.3.3 and 5.4).

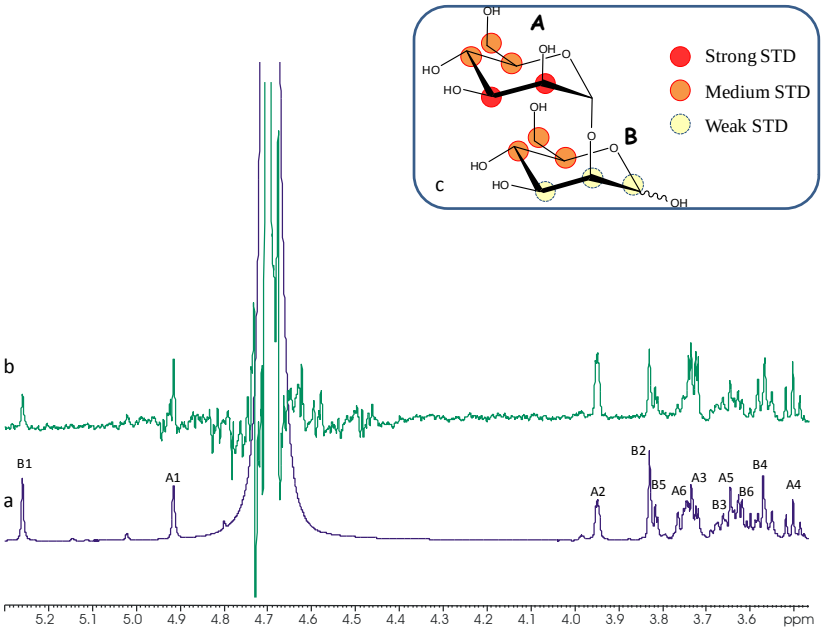


Fig. 5.5: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture rhMBL: α-(1-2)-mannose disaccharide 1:100. c. Chemical structure and epitope binding of α-(1-2)-mannose disaccharide to rhMBL; protons with highest relative STD and, thus, in closest proximity to the receptor are highlighted in red.

In order to quantitatively investigate the interaction with the receptor protein, we acquired STD build up curves by collecting spectra at different saturation times (from 1 sec to 5 sec) [27]. Indeed, since the size of the observed STD signals is dependent not only on its proximity to the receptor but also on its longitudinal relaxation time T_1, to prevent possible misinterpretation, the use of STD build-up curves has been proposed [28,29]. The slope of the STD build up curve at saturation time of 0 (STD_{fit}) is considered to correspond to the STD intensity in the absence of T_1 bias and merely dependent on the proximity of the ligand proton to the protein.

Table 5.6 Experimental STD intensities of **1** bound to rhMBL at different saturation times. STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 – exp(- k_{sat} * t_{sat}))$.

1H	Experimental	STD		STDmax	Ksat	STD (fit)	STD epitopes
	1s	3sec	5sec				(fit)
B1	36,4	30,2	32.3	0,0046	0,3606	0,001659	30,5
A2	90,3	89,5	92,9	0,0107	0,45	0,004815	88,5
B2	56,6	55,8	62,3	0,0075	0,3776	0,002832	52,1
B5	60,3	60,0	59,0	0,0067	0,5105	0,00342	62,9
A6	49,5	43,3	45,8	0,0052	0,4606	0,002395	44,0
A3	100,0	100,0	100,0	0,0114	0,4771	0,005439	100,0
B3	49,3	45,8	52,0	0,0062	0,4149	0,002572	47,3
A5	58,2	52,9	55,5	0,0063	0,4588	0,00289	53,1
B6	54,6	49,9	51,4	0,0058	0,5117	0,002968	54,6
B4	65,5	64,8	66,7	0,0077	0,451	0,003473	63,8
A4	65,0	66,6	69,2	0,0081	0,46	0,003726	68,5

A further advantage of curve fitting, deriving STD intensities close to zero saturation time, is to eliminate artifacts due to differences in capability to accumulate saturation in the free state and minimize the intramolecular spin diffusion within the bound state (cfr. 2.3.2). This method allows a more reliable comparison of binding epitopes from STD NMR between ligands of the same receptor. STD_{fit} is obtained from fitting the saturation time data to the monoexponential equation [30-33]:

$$STD = STD_{max}(1 – exp^{(-k_{sat}t)}),$$

where STD is the STD signal intensity of a given proton at a saturation time t, STD_{max} stands for the asymptotic maximum of the curve, and k_{sat} is the observed saturation rate constant that measures the speed of STD build up. The epitope map of ligand **1** was obtained by normalizing all the values of different protons ligand to the largest STD_{fit}, giving STD $epitopes_{fit}$ (Table 5.6; Figures 5.6).

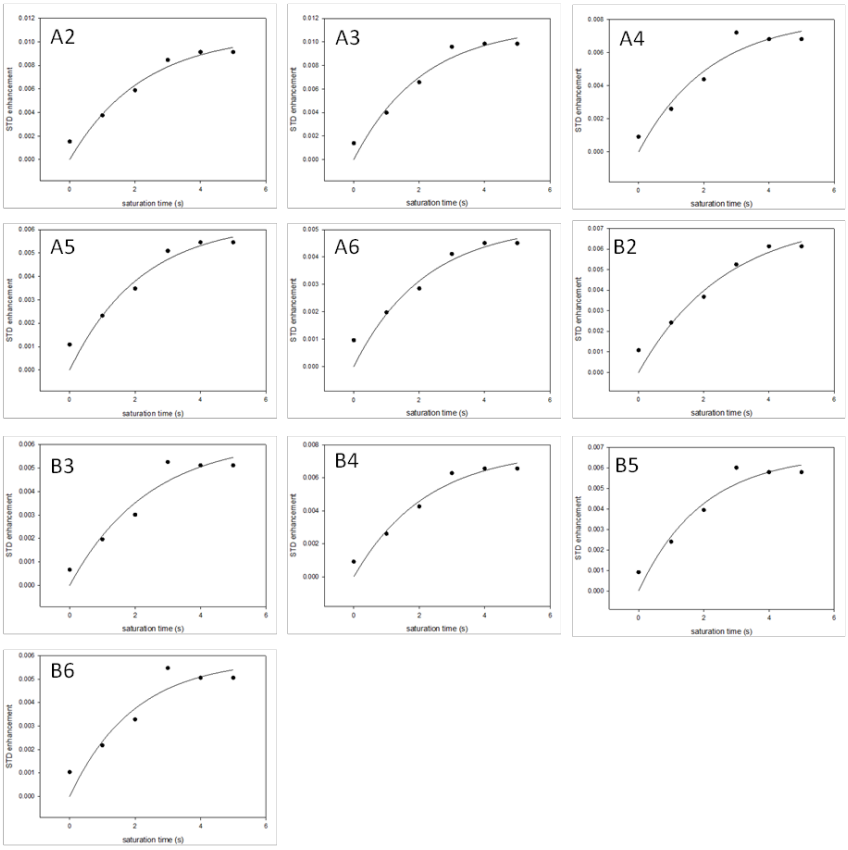


Fig. 5.6: STD build up curves for each proton of the ring A and B of 1 bound to rhMBL.

From the quantitative results, it could be confirmed that protons H-2 and H-3 of ring A were very intimately recognized by the receptor binding-pocket and showed STD epitope$_{fit}$ values close to 100%. It is worth noting that even protons A-4, B-4 and B-5 must be regarded as important for the interaction since they exhibit STD epitope$_{fit}$ values over 60%.

In order to better investigate the binding affinity of the rhMBL to disaccharide **1,** the dissociation constant, K_d, of the α-(1→2) mannose disaccharide to the lectin rhMBL was estimated using single-ligand titration experiments [33]. In detail, the disaccharide **1**

was titrated to a 20uM solution of rhMBL and the resulting STD amplification factors, obtained for each ligand proton, were fitted corresponding to a one-site-binding model.

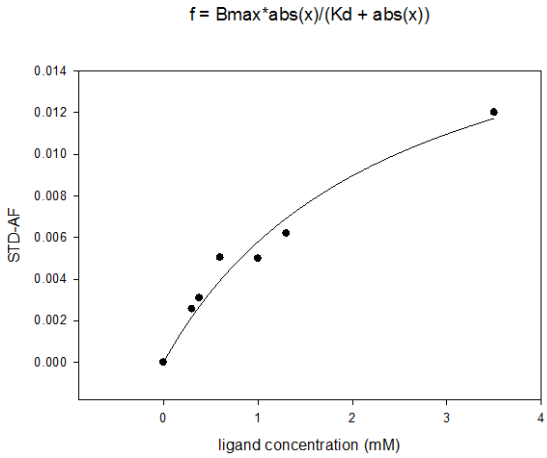


Fig. 5.7: STD NMR titration curve for selected proton (A2) of disaccharide **1** with a close contact to rhMBL.

Affinity constant (kd) estimation was performed by studying the build-up behavior of the STD NMR spectra in conditions of constant protein concentration and increasing ligand concentration. To determine the kd values, at first we converted the measured STD intensity of each proton into the corresponding STD amplification factor (STD-AF), which is defined as the product of the experimental STD effect by the ligand-to-protein excess (a).

$$\text{(a)}\quad \text{STD-AF} = \text{STD\%} * [L] / [P]$$

To construct a binding isotherm from STD NMR titration experiments, we fitted the plotted data points to a one-site binding model (b),

$$\text{(b)}\quad \text{STD-AF} = (\text{STD-AF}_{max} * [L]) / (kd + [L])$$

The kd values of **1** were determined with ligand concentration of 0.3, 0.4, 0.6, 1, 1.3 and 3.5 mM (Figure 5.7). In table 5.7 are reported the kd values estimated for each proton of the ligand **1**. The results showed a K_d value in the mM range, typical for saccharide ligands binding to proteins [34] (Figure 5.7).

Table 5.7 Dissociation constants of individual protons of ring A and B from **1** as determined by STD NMR.

Proton	Kd mM	Proton	kd mM
A1	1,86 ± 0,60	**B1**	0,86 ± 0,31
A2	2,43 ± 0,72	**B2**	2,63 ± 0,89
A3	1,77 ± 0,41	**B3**	1,16 ± 0,54
A4	0,84 ± 0,22	**B4**	1,20 ± 0,32
A5	1,04 ± 0,36	**B5**	0,86 ± 0,31
A6	1,86 ± 0,37	**B6**	nd

The above STD studies were then complemented with the analysis of trNOESY experiments on **1** both free in solution and in the presence of rhMBL, to assess ability of rhMBL to bind to **1** and to analyze disaccharide conformational changes upon binding [35-38].
Thus, trNOESY experiments were carried out at a 20:1 product-to-enzyme molar ratio with different NMR mixing times. As expected, positive NOEs were observed for the free state, and negative for the bound states (Figure 5.8a-b).
Indeed, the appearance of trNOE relies on different tumbling times of free and bound molecules; low molecular weight molecules exhibit positive NOEs, while large ones show negative NOEs. When the ligand binds to receptor protein, the NOEs undergo drastic changes in the sign, demonstrating the binding.

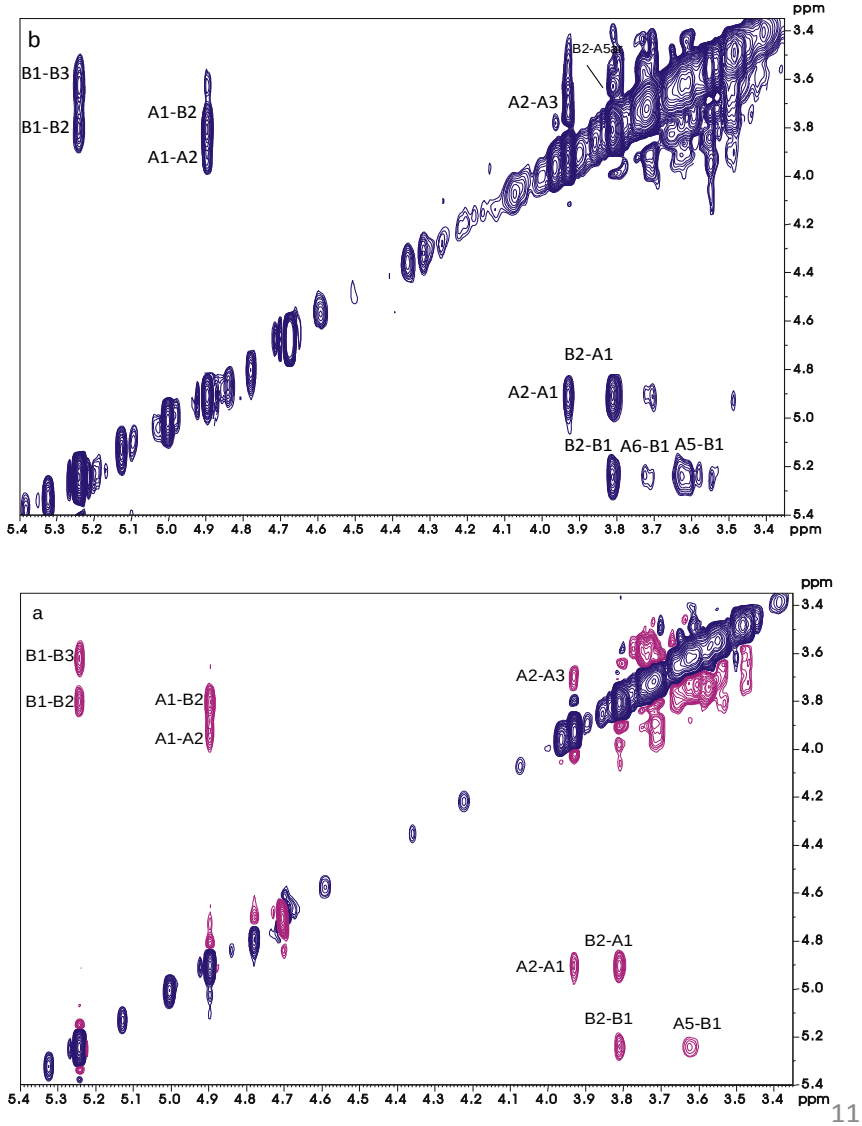


Fig. 5.8: NOESY (a) and tr-NOESY (b) on α-(1-2)-mannose disaccharide and on MBL: α-(1-2)-mannose disaccharide 1 : 20.

In order to obtain more information on the conformation of the ligand **1**, experimental interproton distances were compared with calculated distances values for both energetic minima predicted for α- (1-2)-mannose disaccharide (Fig 5.9).

a

Distance	Experimental Free state	Experimental Bound state	Calculated Φ= -33° Ψ0 57°	Calculated Φ= -50° Ψ= -23°
A1-B2	2.3	2.35	2.4	2.5
A1-B1	nd	nd	2.7	4.3
B1-A5	2.6	2.5	2.7	2.4
B1-A6	nd	3.3	3.4	4.9

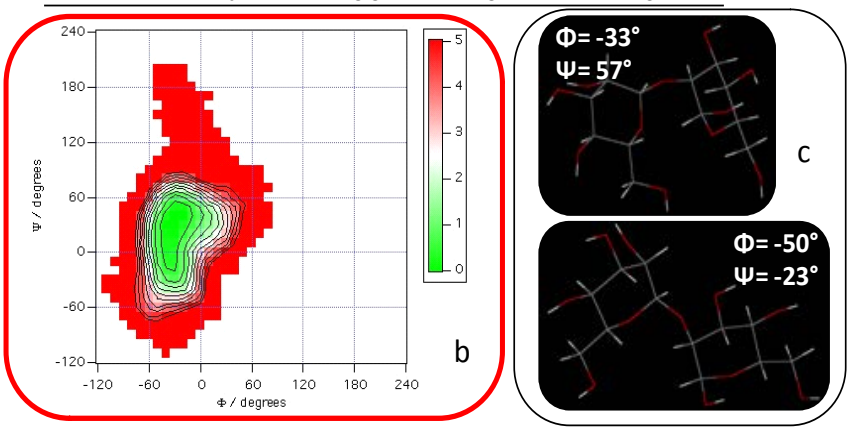


Fig. 5.9: a)Experimental and calculated interproton distances for α- (1-2) mannose disaccharide for both energetic minima (showed in Fig.5.9c).
b) Adiabatic energy maps of the disaccharide **1**.

A detailed analysis of inter-residue cross-peaks in the trNOESY spectrum allowed the deduction that key NOEs in both free and bound states showed few changes. In detail, cross peaks B1-A5 and B2-A5 increased in intensity in the presence of the receptor protein. Furthermore, while in the free state no cross peaks were observed between protons B1 and A6, these NOE contacts were detected in the bound state. Furthermore, NOE signals between A1 and B1 were not observed in neither of NOESY or tr-NOESY.

All collected data suggested that, in both states, there is a conformational equilibrium between the two different minima (showed in Figure 5.9b); however, following recognition and binding, the conformational state characterized from negative values of the inter-glycosidic dihedral angles, φ and ψ, seemed to be more populated (Figure 5.9a-b).

5.3.2 The interaction of the disaccharide **2** with the rhMBL. STD NMR and tr-NOESY analysis

A similar approach was used for the α-(1→3) mannose disaccharide (**2**). STD NMR data (Figures 5.10a-b) showed that the ligand was able to bind rhMBL and both a qualitative (Figure 5.10c) and quantitative (Table 5.8; Figure 5.11) analysis demonstrated that, also in this case, the non reducing mannose, e.g. ring A, was strongly involved in the interaction with the protein.

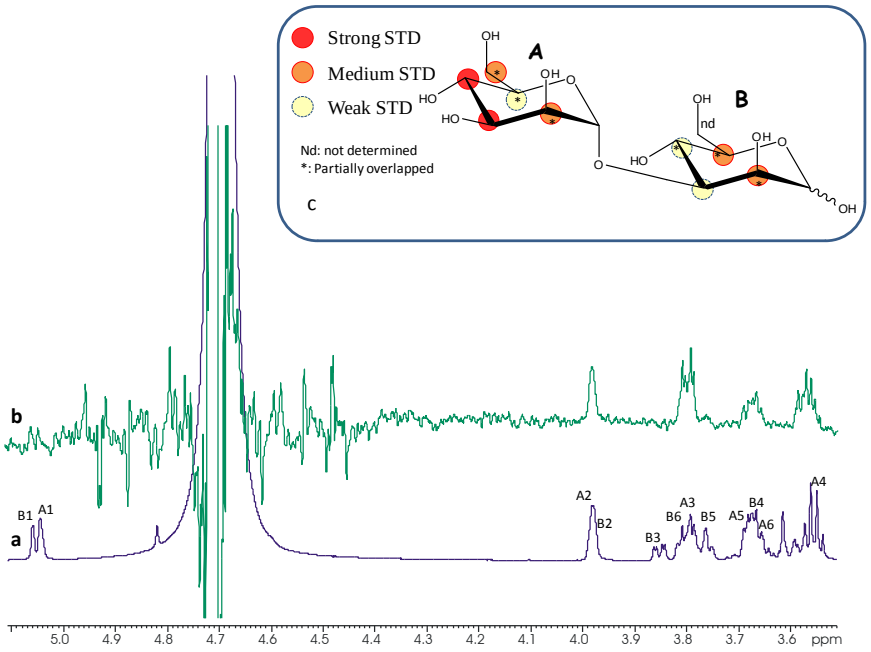


Fig. 5.10: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture rhMBL: α-(1-3)-mannose disaccharide 1:100. c. Chemical structure and epitope binding of α-(1-3)-mannose disaccharide to rhMBL; protons with highest relative STD and, thus, in closest proximity to the receptor are highlighted in red.

Table 5.8 Experimental STD intensities of **2** bound to rhMBL at different saturation times. STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 - exp(- k_{sat} * t_{sat}))$.

1H	Experimental	STD		STDmax	Ksat	STD (fit)	STD epitopes
	1s	3sec	5sec				(fit)
A2+B2	73,9	66,6	67,5	0,0049	0,4976	0,002438	64,6
B3a	59,9	33,02	41,67	0,0025	0,8441	0,00211	55,9
B6	nd	nd	nd	nd	nd	nd	nd
A3	100	100	100	0,0073	0,5174	0,003777	100,0
B5	49,1	35,9	34,5	0,0022	0,8589	0,00189	50,0
B3+A5	44,4	38,1	39	0,0029	0,5284	0,001532	40,6
A6+B4	56,1	49,70	45,9	0,0032	0,6423	0,002055	54,4
B4b	61,6	75,3	78,1	0,0066	0,3586	0,002367	62,7
A4	87,7	77,2	80,32	0,0058	0,5264	0,003053	80,8
B5b	17,8	31,3	21,20	0,0016	0,3799	0,000608	16,1

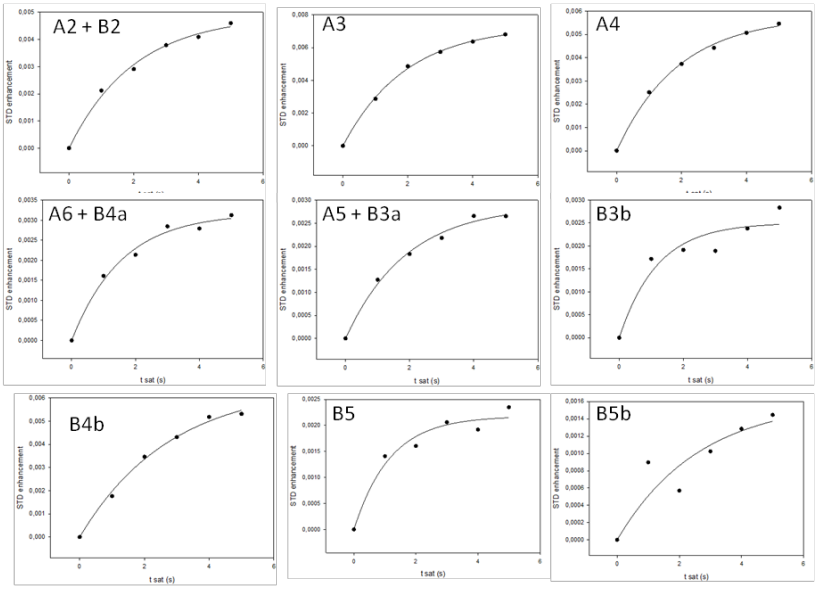


Fig. 5.11: STD build up curves for each proton of the ring A and B of **2** bound to rhMBL.

In particular, in the disaccharide **2** the largest amount of saturation was received by protons H-3 and H-4 of ring A. High STD epitope$_{fit}$ values (over 60%) were observed also for protons A-2, B-2 and B-4.

A comparison between the epitope maps obtained for the two ligands (**1** and **2**) bound to rhMBL revealed that both substrates were recognized by the lectin in a quite similar way, with the non-reducing mannose A always making closest contacts with the

receptor binding pocket (Table 5.6 and 5.8). Indeed, in both cases, the highest STD values were observed for protons A-3, A-2, A-4 and B-4, exhibiting STD epitope$_{fit}$ higher than 60%, suggesting that these moieties played an important role in the binding process. Furthermore, the reducing mannose, for both ligands, showed lower STD contribution in the interaction with rhMBL.

5.3.2 The interaction of the disaccharides **1** and **2** with the L-FCN/MBL76

Afterward, we have also analyzed the relative binding affinities of **1** and **2** to the chimeric molecule L-FCN/MBL76, in order to

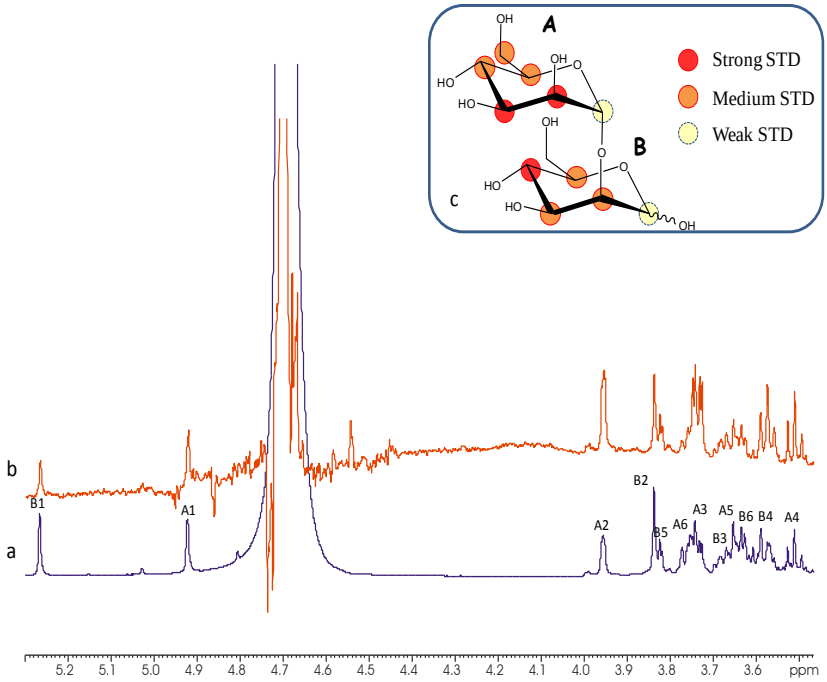


Fig. 5.12: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture L-FCN/MBL76: α-(1-2)-mannose disaccharide 1:100. c. Chemical structure and epitope binding of -(1-2)-mannose disaccharide to L-FCN/MBL76; protons with highest relative STD and, thus, in closest proximity to the receptor are highlighted in red.

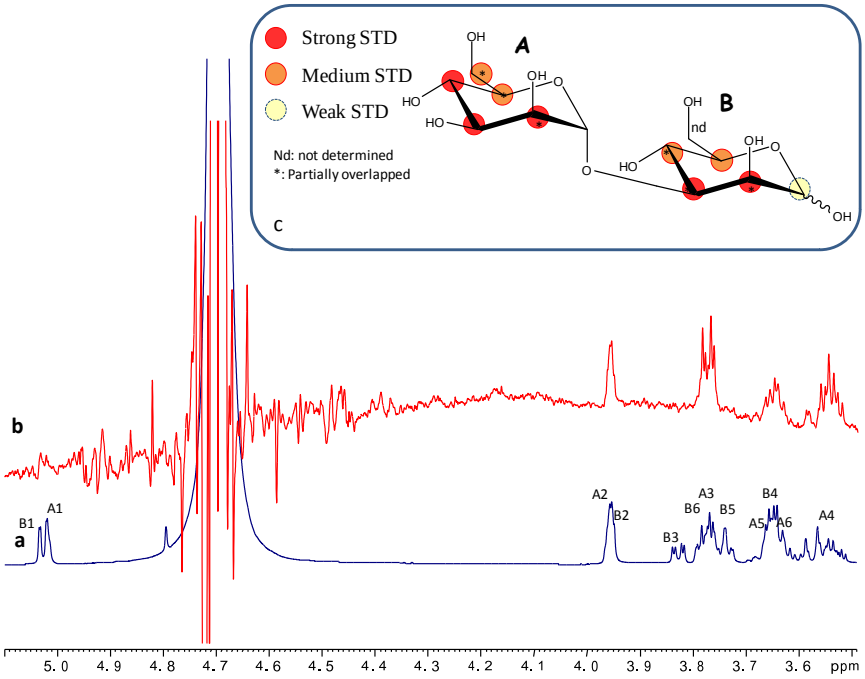


Fig 5.13: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture L-FCN/MBL76 : α-(1-3)-mannose disaccharide 1:100. c. Chemical structure and epitope binding of -(1-3)-mannose disaccharide to L-FCN/MBL76; protons with highest relative STD and, thus, in closest proximity to the receptor are highlighted in red.

investigate the possible binding modes of the protein following recognition of the two mannose disaccharide.
STD spectra are reported in Figures 5.12 and 5.13; to better compare the binding maps of the two ligands, initial slopes were calculated from the STD build-up curves identifying the ligand epitopes, overcoming possible artifacts (see above for a more detailed discussion; Table 5.9-5.10; Figures 5.14, 5.15) [27].

Table 5.9 Experimental STD intensities of **1** bound to L-FCN/MBL76 at different saturation times. STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 – exp(- k_{sat} * t_{sat}))$.

1H	Experimental STD			STDmax	Ksat	STD (fit)	STD epitopes
	1s	3sec	5sec				(fit)
B1	35	38,2	37,5	0,0078	0,4565	0,00356	37,7
A2	100,0	100,0	99,4	0,0161	0,5866	0,009444	100,0
B2	68,5	69,0	68,2	0,0109	0,6056	0,006601	69,9
B5	73,2	69,4	65,8	0,0103	0,686	0,007066	74,8
A6	63,4	54,0	49,7	0,0079	0,8118	0,006413	67,9
A3	95,4	98,6	100,0	0,0165	0,5403	0,008915	94,4
B3	63,3	57,0	54,0	0,0087	0,7212	0,006274	66,4
A5	62,6	55,5	54,2	0,0086	0,6708	0,005769	61,1
B6	56,9	52,4	53,7	0,0083	0,6399	0,005311	56,2
B4	72,7	80,7	85,3	0,0144	0,4635	0,006674	70,7
A4	60,4	65,7	69,0	0,0117	0,4685	0,005481	58,0

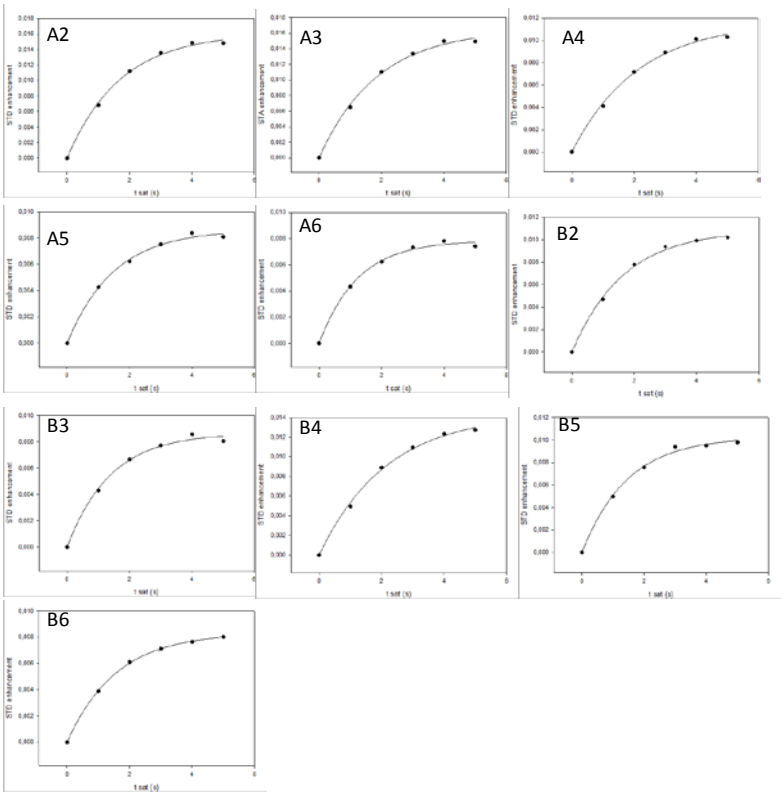


Fig. 5.14: STD build up curves for each proton of the ring A and B of **1** bound to L-FCN/MBL76.

STD values obtained following the interaction between ligands **1** and **2** and the chimeric lectin L-FCN/MBL76, suggested that both disaccharides exhibited a similar recognition mode by L-

FCN/MBL76, specifically interacting with the protein through their non reducing end mannose, ring A (Figure 5.14 and 5.15, Table 5.9 and 5.10). Indeed, in each STD spectrum (Figure 5.12-5.13), the strongest STD effects involved protons H-2 and H-3 of ring A, indicating that they made the closest contacts with the chimer binding site. Furthermore, in both cases, significant STD NMR signals were also observed for protons of residue B, revealing that even this region of the molecule was involved in the binding to the protein.

Table 5.10 Experimental STD intensities of **2** bound to L-FCN/MBL76 at different saturation times.
STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 – exp(- k_{sat} * t_{sat}))$.

1H	Experimental STD			STDmax	Ksat	STD (fit)	STD epitopes
	1s	3sec	5sec				(fit)
A2+B2	96,7	86,4	87,7	0,0072	0,7733	0,005568	92,9
B3a	89,6	75,4	70,2	0,0057	1,0258	0,005847	97,6
B6	nd	nd	nd	nd	nd	nd	nd
A3	100,0	99,0	92,6	0,0078	0,7684	0,005994	100,0
B5	66,2	57,8	54,1	0,0045	0,9265	0,004169	69,6
B3+A5	59,1	61,3	60,3	0,0051	0,6655	0,003394	56,6
A6+B4	58,8	59,1	58,0	0,0051	0,6773	0,003454	57,6
B4b	65,4	63,0	66,4	0,0055	0,6316	0,003474	57,9
A4	93,7	100,0	100,0	0,0086	0,6196	0,005329	88,9
B5b	43,5	38,4	39,7	0,0032	0,7622	0,002439	40,7

In addition, slight differences in the binding modes between **1** and **2** disaccharides were observed; in detail, for ligand **1** bound to L-FCN/MBL76, protons B-4 and B-5 received high transfer of magnetization compared to rhMBL, while for disaccharide **2**, the absolute STD values were high for protons B-2 and B-3.

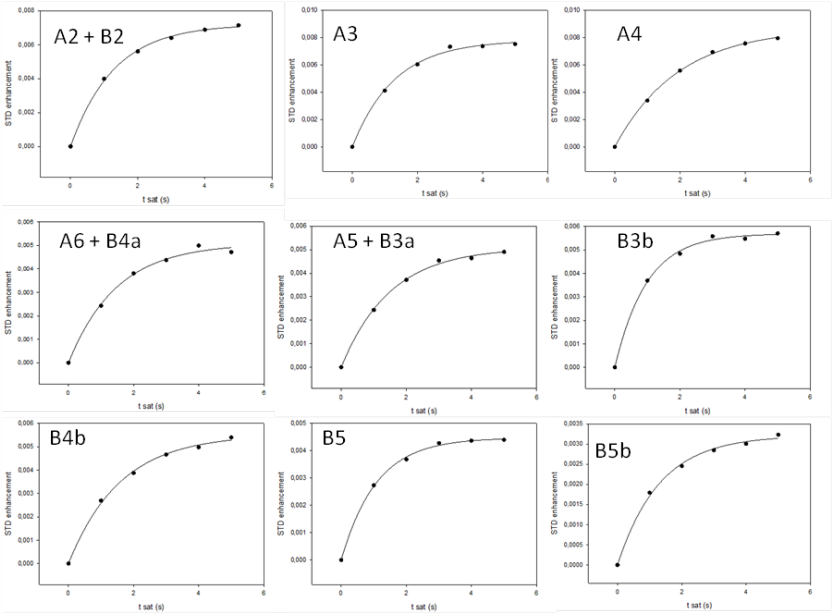


Fig. 5.15: STD build up curves for each proton of the ring A and B of **2** bound to L-FCN/MBL76.

Comparing the interaction of each disaccharide, **1** and **2,** with each receptor protein rhMBL and L-FCN/MBL76 (Figure 5.16 and 5.17), a similar pattern of STD relative intensities was observed for each disaccharide following the interaction with rhMBL as well as with L-FCN/MBL76. From this study it was clear that, for both proteins, ligands from the non reducing end showed the most relevant STD effects revealing a strong contribution of that moiety to the recognition protein. Nevertheless, for both disaccharides, protons of the reducing mannose B exhibited larger STD effects with the chimeric lectin, indicating a higher contribution to the binding compared to rhMBL and thus their stronger interaction with L-FCN/MBL76.

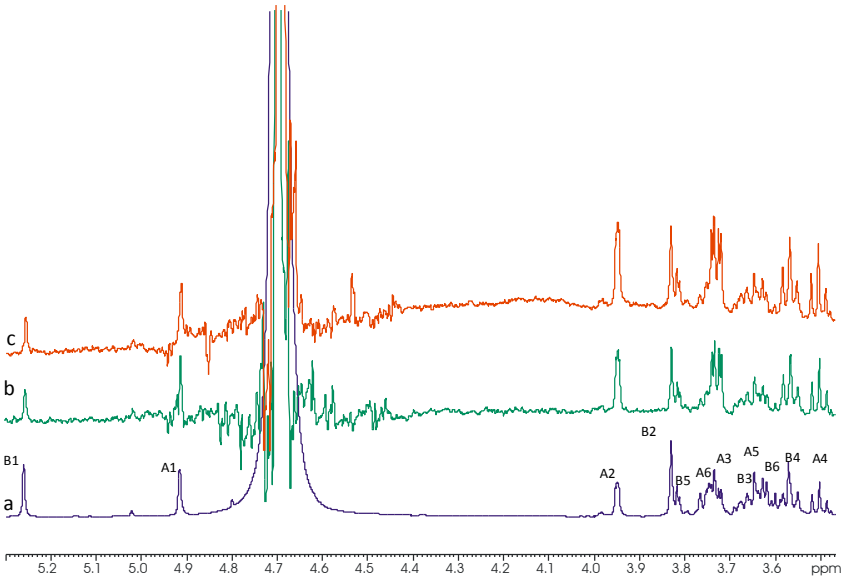


Fig. 5.16: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectra of mixture rhMBL: α-(1-2)-mannose disaccharide 1:100 (b) and L-FCN/MBL76 : α-(1-2)-mannose disaccharide 1:100 (c).

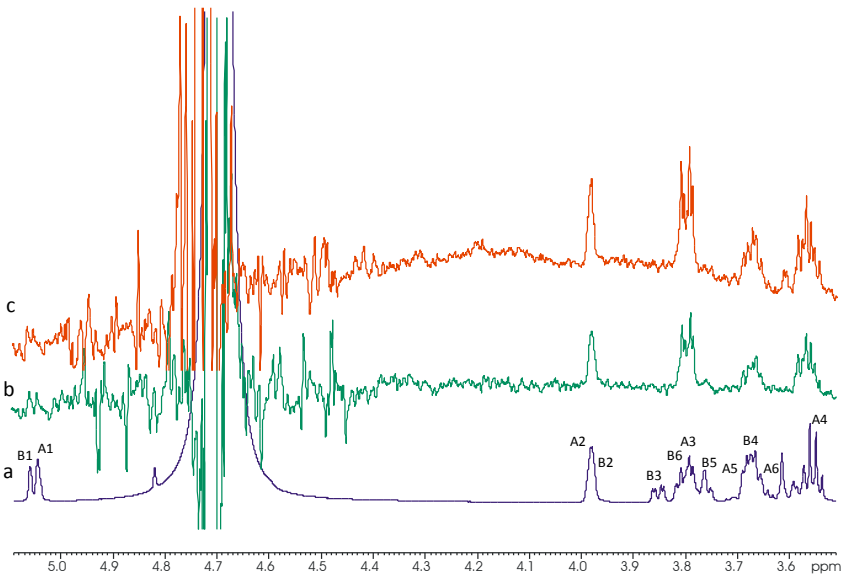


Fig. 5.17: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectra of mixture rhMBL: α-(1-3)-mannose disaccharide 1:100 (b) and L-FCN/MBL76 : α-(1-3)-mannose disaccharide 1:100 (c).

5.4 Conclusion

Herein, we have described an NMR based approach to analyze and compare, at atomic level, interactions between two immune effector molecules, rhMBL and a chimeric derivative, L-FCN/MBL76, and carbohydrate ligands (monosaccharides and mannosyl disaccharides). Using high resolution STD NMR and trNOE techniques, we have demonstrated in detail how rhMBL recognizes and binds hexose sugars and acetylated residues.

Furthermore, we have analyzed the binding activity of rhMBL and L-FCN/MBL76 with two different disaccharides, α-(1→3) and α-(1→2) mannose disaccharides, in order to investigate which carbohydrate moieties were primarily involved in the recognition and interaction processes. Data indicated that both rhMBL and L-FCN/MBL76 were able to bind α-(1→3) and α-(1→2) mannose disaccharides and, although there were slight differences in the binding modes, STD values showed that the non reducing ends of both disaccharides were always strongly involved in the interaction with both rhMBL and L-FCN/MBL76 proteins.

This study represents a first step toward the full description of the interaction between rhMBL and the fusion protein L-FCN/MBL76, and various carbohydrate moieties. The NMR experiments defined the binding epitope of sugar residues bound to both proteins at atomic resolution. In addition, it is worth noting that the NMR data for the chimeric molecule showed that L-FCN/MBL76 exhibited good functional activity. This finding is in agreement with a previous study where it was shown that, compared with rhMBL, the chimeric molecule had superior effector function and binding capacity to lethal N-glycosylated viruses and this is mediated by a greater structural flexibility [17].

In literature there are several examples of such studies of mannose oligosaccharides with proteins. For example, diverse results were obtained when the interaction between α-(1→2) mannose disaccharide and DC-SIGN was investigated via NMR

spectroscopy in which, according to the orientation of the ligand in the binding pocket of the protein, both the non reducing and reducing residues of the disaccharide were involved in the interaction, allowing to deem a dual binding mode of such mannose disaccharide [39-40]. In case of rhMBL, NMR data showed both α-(1→2) and α-(1→3) mannose disaccharides were oriented at the Ca^{2+} site in a way that the non reducing unit was mainly interacting with the protein. On the other site, in case of FCN/MBL76 protein, despite its strong affinity for the non reducing unit of both α-(1→2) and α-(1→3) disaccharides, a slightly different behaviour could be detected for the reducing ends, since even this region of the molecule seemed to be involved in the binding to the protein, likely due to the presence of multiple binding sites on the chimeric protein. Similarly, rat serum and liver MBP were also investigated in their ability to bind α-(1→2) mannose disaccharides [41,42]. While both α-(1→2) and α-(1→3) showed a unique binding more in rat serum MBL, the nature of the liver MBP binding site, less restrictive, allowed an easier accommodation of the non terminal portion of α-(1→2) mannose disaccharides. The possibility of having multiple binding modes might be explained the ability of an oligomeric like lectin to interact simultaneously with different branches of an oligosaccharide chain.

In summary, our findings indicate that atomic level analysis of the binding dynamics of immune effector molecules may be useful for designing and optimizing novel therapeutic proteins. Further studies will be necessary to obtain other detailed insights of binding modes, but the approach presented here represents a key successful method of great potential for pharmaceutical research, useful for designing and optimizing novel therapeutic proteins.

References

[1] Ip, W.K.; Takahashi, K.; Ezekowitz, R.A.; Stuart. L.M. *Immunol Rev*., **2009**. 230(1), 9–21.
[2] Shi, L.; Takahashi, K.; Dundee, J.; Shahroor-Karni, S.; Thiel, S.; Jensenius, J.C.; Gad, F.; Hamblin, M.R.; Sastry, K.N.; Ezekowitz, R.A. *J. Exp. Med.*, **2004.** 199, 1379–1390.
[3] Møller-Kristensen, M.; Hamblin, M.R.; Thiel, S.; Jensenius, J.C.; Takahashi, K. *J. Invest.Dermatol.*, **2007**. 127, 1524-1531.
[4] Kase, T.; Suzuki, Y.; Kawai, T.; Sakamoto, T.; Ohtani, K.; Eda, S.; Maeda, A.; Okuno, Y.; Kurimura, T.; Wakamiya, N. *Immunology*, **1999**. 97, 385–392.
[5] Michelow, I.C.; Dong, M.; Mungall, B.A.; Yantosca, L.M.; Lear, C.; Ji, X.; Karpel, M.; Rootes, C.L.; Brudner, M.; Houen, G.; Eisen, D.P.; Kinane, T.B.; Takahashi, K.; Stahl, G.L.; Olinger, G.G.; Spear, G.T.; Ezekowitz, R.A.; Schmidt, E.V. *J Biol Chem.*, **2010**. 285:24729–39.
[6] Michelow, I.C.; Lear, C.; Scully, C.; Prugar, L.I.; Longley, C.B.; Yantosca, L.M.; Ji, X.; Karpel, M.; Brudner, M.; Takahashi, K.; Spear, G.T.; Ezekowitz, R.A.; Schmidt, E.V.; Olinger, G.G. *The Journal of Infectious Diseases*, **2011**. 203:175–179.
[7] Chang, W.C.; Hartshorn, K.L.; White, M.R.; Moyo, P.; Michelow, I.C.; Koziel, H.; Kinane, B.T.; Schmidt, E.V.; Fujita, T.; Takahashi, K. *Biochem Pharmacol*., **2011**. 81(3), 388-95.
[8] Marchetti, R.; Lanzetta, R.; Michelow, C.J.; Molinaro, A. and Silipo, A. *Eur. J. Org. Chem*., **2012.** 27, 5275–5281.
[9] Presanis, J.S.; Kojima, M.; Sim, R.B. *Biochem Soc Trans*., **2003**. 31(Pt 4), 748-52.
[10] Vorup-Jensen, T.; Sørensen, E. S.; Jensen, U. B.; Schwaeble, W.; Kawasaki, T.; Ma, Y.; Uemura, K.; Wakamiya, N.; Suzuki, Y.; Jensen, T. G.; Takahashi, K.; Ezekowitz, R. A.; Thiel, S.; and Jensenius, J. C. *Int. Immunopharmacol.*, **2001**. 1, 677–687.
[11] Teillet, F.; Dublet, B.; Andrieu, J.P.; Gaboriaud, C.; Arlaud, G.J.; Thielens, N.M. *J Immunol*., **2005.** 174(5), 2870-7.
[12] Weis, W.I.; Drickamer, K.; Hendrickson, W.A. *Nature* **1992**. 360:127–134.
[13] Drickamer, K. *Nature* **1992**. 360:183–186.

[14] Sheriff, S.; Chang, C. Y.; and Ezekowitz, R. A. *Nat. Struct. Biol*. **1994**. 1, 789–794.
[15] Takahashi, K.; Ip, W.E.; Michelow, I.C.; Ezekowitz, R.A. *Curr. Opin. Immunol*. **2006**. 18, 16–23.
[16] Ezekowitz, R.A. *Curr. Biol.* **1991**. 1, 60–62.
[17] Ji, X.; Olinger, G.G.; Aris, S.; Chen, Y.; Gewurz, H.; Spear, G.T. *J. Gen. Virol*. **2005**. 86, 2535–2542.
[18] Roy, S. et al. *THE LANCET,* **2002**. 359, 1569-1573.
[19] Babovic-Vuksanovic, D.; Snow, K.; Ten, R. M. *Annals of Allergy, Asthma & Immunology,* **1999**. *82,134-138,141-143.*
[20] Hummelshoj, T.; Thielens, N. M.; Madsen, H. O.; Arlaud, G. J.; Sim, R. B.; and Garred, P. *Mol. Immunol*. **2007**. 44, 401–411.
[21] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed. Engl*. **2003**. 42, 864.
[22] Mayer, B.; Meyer, B. *J. Am. Chem. Soc*. **2001**, 123, 6108.
[23] Squeglia, F.; Marchetti, R.; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *J Am Chem Soc.* **2011**. 133(51), 20676-9.
[24] Mari, S.; Serrano-Gomez, D.; Canada F.J.; Corbi, A.L.; Jimenez-Barbero, J. *Angew. Chem. Int. Ed. Engl*. **2004**. 44, 296.
[25] Angulo, M.J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Krishna, N.R.; Peters, H.; Palcic, M.M.; Peters, T. *J. Am. Chem. Soc*., **2006**. 128, 13529.
[26] Yuan, Y.; Bleile, D.W.; Wen, X.; Sanders, D.A.; Itoh, K.; Liu, H.W.; Pinto, B:M: *J. Am. Chem. Soc.* **2008**. 130(10), 3157-68.
[27] Mayer, M.; and James, L.T. *J. Am. Chem. Soc*., **2004**. 126, 4453-4460.
[28] Jayalakshmi, V.; Krishna, N.R. *J. Magn. Reson*, **2002**. 155, 106-118
[29] Yan, J.L.; Kline, A.D.: Mo, H.P.; Shapiro, M.J.; Zartler, E.R.; *J. Magn. Reson* , **2003**. 163, 270276.
[30] Szcepina, G. M.; Bleile, W.D.; and Pinto, B.M. *Chem. Eur. J*., **2011**. 17, 11446-11455.
[31] Yuan, Y.; Bleile, D.W.; Wen, X.; Sanders, D.A.R; Itoh, K.; Liu, H.; Pinto, B.M.. *J. Am. Chem. Soc*.**2008**. 130, 3157.
[32] Angulo, J.; and Nieto, P. *Eur Biophys J*, **2011**. 40, 1357-1369.
[33] Angulo, J.; Enríquez-Navas, P.M.; and Nieto, P. *Chem. Eur. J*., **2010**. *16*, 7803-7812.

[34] Wilhelm, D.; Behnken, N.H.; Meyer, B. *ChemBioChem*., **2012**. *13*(4), 524-7.
[35] The Nuclear Overhauser Effect in Structural and Conformational Analysis (Eds.: D. Neuhaus, M. P. Williamson), Wiley-VCH, New York, **2000**.
[36] NMR Spectroscopy of Glycoconjugates (Eds.: J. Jimenez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**
[37] Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chem. Eur. J*. **2011**. 17(5), 1547-60.
[38] Johnson, M.A.; Pinto, B.M. *Carbohydr Res*., **2004**. 339(5):907-28.
[39] Angulo , J.; Díaz , I.; Reina, J.J.; Tabarani, G.; Fieschi, F.; Rojo, J.; Nieto P.M. *ChemBioChem,* **2008**. *9*, 2225-2227.
[40] Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chemistry*, **2011**. *17*(5), 1547-60.
[41] Ng , K.K.; Kolatkar, A.R.; Park-Snyder, S.; Feinberg , H.; Clark D.A.; Drickamer, K.; Weis, W.I. *J Biol Chem.* **2002** 277(18), 16088-95.
[42] Hilge, M.; Gloor, S.M.; Rypniewski, W.; Sauer, O.; Heightman, T.D.; Zimmermann, W.; Winterhalter, K.; Piontek, K. *Structure*. **1998**. *6*(11), 1433-44.

Chapter 6

mAb5D8 – *B. anthina* LPS interaction

6.1 Introduction

In this study the interaction between the O-chain from the lipopolysaccharide from *Burkholderia anthina* and the LPS specific monoclonal antibody has been performed via NMR spectroscopy. In particular, the 5D8-bound epitope at the saccharide entity has been unraveled using a combination of saturation transfer difference NMR and transferred NOESY experiments performed on the whole 5D8/polysaccharide complex. In addition, to further dissect the fine details of the molecular recognition events, further experiments with simpler carbohydrate ligands have also been carried out using the *ad hoc* synthesized trisaccharide and hexasaccharide O-antigen repeating units. Using this multidisciplinary approach of chemical synthesis, NMR spectroscopy and molecular dynamics simulations, the determination of the binding epitope and the contribution to the binding of the sugar units composing the O-chain have then been determined.

6.1.1 Bcc infection in cystic fibrosis disease

Bacteria belonging to the genus *Burkholderia* are ubiquitous Gram-negative microorganisms that are able to adapt, survive and persist in diverse environmental conditions, such as water, soil, plants and insects. Several *Burkholderia* species can be used for biotechnological and agricultural applications since they can act as biopesticides and function in the biodegradation of organic compounds [1]; however, the genus *Burkholderia* also comprises strains that are able to cause severe infections in animals, plants and humans. In particular, many species are human opportunistic

pathogens responsible for lung infections in immunocompromised patients, such as those with cystic fibrosis (CF). This is the most common fatal genetic disease among Caucasian population; it is induced by a point mutation in a gene cluster on the long arm of chromosome 7 [2]. This gene encodes a protein (CFTR: cystic fibrosis transmembrane regulator) responsible of transmembrane conductance. Typically, mutations in CFTR affect the regulation of different ion channels, causing a failure in innate defence against pathogens thus producing chronic bacterial infections. Indeed, in a healthy tissue, the hydration is guaranteed by the capacity of normal epithelium surface to absorb and secrete salts with water moving osmotically in response to salts gradient. On the other hand, in a CF tissue, impaired CFTR provokes an unregulated Na^+ adsorption and a low secretion of Cl^-; the consequent dehydration causes a collapse of mucus layer on airway surface that generates airflow obstruction and creates an ideal habitat for a wide range of microorganisms causing inflammation and chronic infection (Figure 6.1).

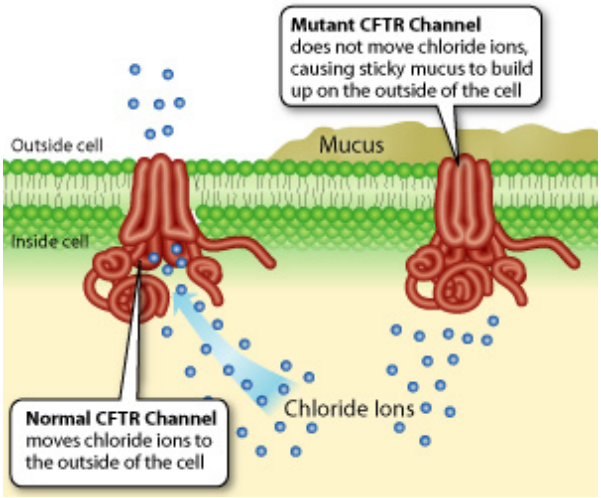


Fig. 6.1: Schematic representation of normal and mutant CFTR channel.

Among CF pathogens, the dominant ones are *Pseudomonas aeruginosa* and many phenotypically closely related bacteria, grouped into the *Burkholderia cepacia* complex (Bcc), which are divided in various genetically different classes, known as genomovars [3].

Burkholderia anthina, belonging to the genomovar VIII, is an uncommon CF pathogen that produces a number of potential virulence factors pivotal in the mechanism of bacterial invasion and virulence, previously defined for various Bcc isolates [4,5]. Among these determinants, one of the most important is the lipopolysaccharide, which plays a key role in the establishment of severe respiratory infections in CF patients as well as in the resistance of Bcc to many antibiotics [6].

Lipopolysaccharides are amphiphilic molecules anchored to the outer membrane of Gram-negative bacteria fundamental for their growth and survival (cfr 1.1.3). Bacterial released LPSs are potent activator of the innate immune response. They are composed of three distinct domains: the lipophilic region named lipid A, a conserved hydrophilic core heteropolysaccharide and a highly variable O-polysaccharide chain, which is the antigenic determinant of the LPS. Bcc LPSs may lack the O-antigen and possess a lipooligosaccharide, but several Bcc isolates from CF patients are characterized by a complete LPS and seem to be more resistant to the killing effects of human serum [7,8].

The structures of several O-antigens from bacteria belonging to the Bcc have been elucidated but less is known, at molecular level, about the mechanisms of Gram-negative infection in cystic fibrosis [9], which is pivotal for the potential development of vaccines to prevent Bcc infections, especially in immunocompromised patients [10-14].

Within this frame, we used NMR and molecular recognition techniques to investigate the interaction between the O-polysaccharide chain of the Gram negative bacterium *Burkholderia*

anthina, isolated from a CF patient from United Kingdom, and a monoclonal antibody (mAb), designated as 5D8, specific for *B. cepacia* LPS. Previous studies [15] have demonstrated that this 5D8 mAb, produced with the aim to study the epitope variability among different Bcc species, was able to efficiently bind to both *B. cepacia* and *B. anthina*, but not to several other Bcc species. Additionally, western blotting analysis showed a stronger reaction of the antibody with *B. anthina* strains than with the *B. cepacia* analogues.

Nowadays the knowledge at the highest possible resolution of the structural details of key molecular recognition events is of paramount importance for the design of new molecules that can prevent disease processes, as the Bcc-mediated infection just described. On this basis, and in order to gain information at the molecular level into the aforementioned interaction, we have used advanced NMR methods, STD NMR and tr-NOE experiments, to map the interacting epitope and to gain key conformational information on the bound state, which combined with MD simulations have permitted to provide key structural information [16,17]. Furthermore, in order to deeply unravel the contribution of the different saccharide moieties to the binding, we have synthesized the two oligosaccharides which are structural motifs of the O-antigen of *B. anthina* and analyzed their binding features to the LPS-specific 5D8 mAb.

6.2 NMR studies of LPS from *B. anthina* bound to mAb5D8

The binding of the O-polysaccharide chain of the LPS from *Burkholderia anthina* to the LPS-specific 5D8 mAb was studied by a combination of STD NMR and TR-NOESY experiments. The structural and affinity details of interaction between various antibodies and antigens from many bacteria such as *Salmonella*, *Shigella flexneri* and *Vibrio cholera*, have been previously investigated by using titration microcalorimetry studies [18], X-ray

crystallography analysis [19], surface plasmon resonance [20], as well as fluorescence spectroscopy [21]. However, in the last years, NMR techniques have been often employed in the study of antibody-antigen recognition mechanisms [22-,26], allowing to explore the fine details of these binding events at atomic resolution.

6.2.1 The interaction of the O-polysaccharide chain with the 5D8 mAb. STD-NMR analysis

As first step in the analysis of the interaction between 5D8 and the O-chain of *B. anthina*, the polysaccharide fragment of the ligand, which consists of the following repeating unit: [2)-α-L-Rha-(1→2)-α-D-Gal(1→3)-α-L-Rha], was characterized using standard 2D NMR experiments (Table 6.1, Figure 6.1) [27].

Table 6.1 Chemical shifts of ^{1}H and ^{13}C of the O-polysaccharide chain from *B. anthina*.

	H1 C1	H2 C2	H3 C3	H4 C4	H5 C5	H6 C6
C	5.11	3.86	4.02	3.96	4.14	3.67
2-α-Gal	94.1	74.3	69.3	69.5	70.7	60.7
B	5.15	4.01	3.82	3.43	3.69	1.24
2-α-Rha	100.3	77.9	69.8	71.8	69.1	16.9
A	4.93	4.21	3.81	3.53	3.70	1.24
3-α-Rha	102	66.2	74.4	70.2	69.3	16.9

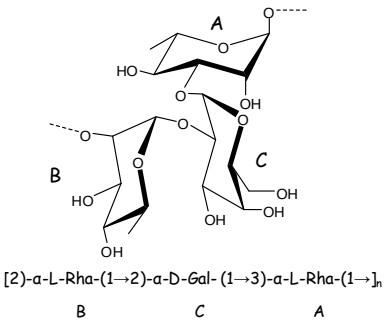


Fig. 6.2: Chemical structure of the repeating unit of O-polysaccharide chain from LPS of *Burkholderia anthina*.

Then, the interaction of the O-polysaccharide with 5D8 was investigated using STD-NMR spectroscopy. This method permits to map the interacting epitope, i. e. the portion of ligand that is in closer contact with the receptor [28-31]. First, different experimental conditions were employed to optimise the STD NMR experiments, which eventually lead to the employment of a sample containing mAb 5D8 (30 uM) in the presence of a 25 molar excess of the O-polysaccharide ligand (cfr. Section IV). The resulting spectrum is reported in Figure 6.3. By comparing each STD spectrum (top in Figure 6.3) with its corresponding reference experiment (bottom in Figure 6.3), the existence of saturation transferred from the antibody to the ligand protons could be deduced. Even though the severe signal overlapping impaired a detailed proton-specific analysis, the qualitative study of the observed STD effects showed that each sugar residue forming the repeating unit was present in the STD spectrum. Therefore, every sugar of the repeating trisaccharide entity is contributing to the binding to the antibody.

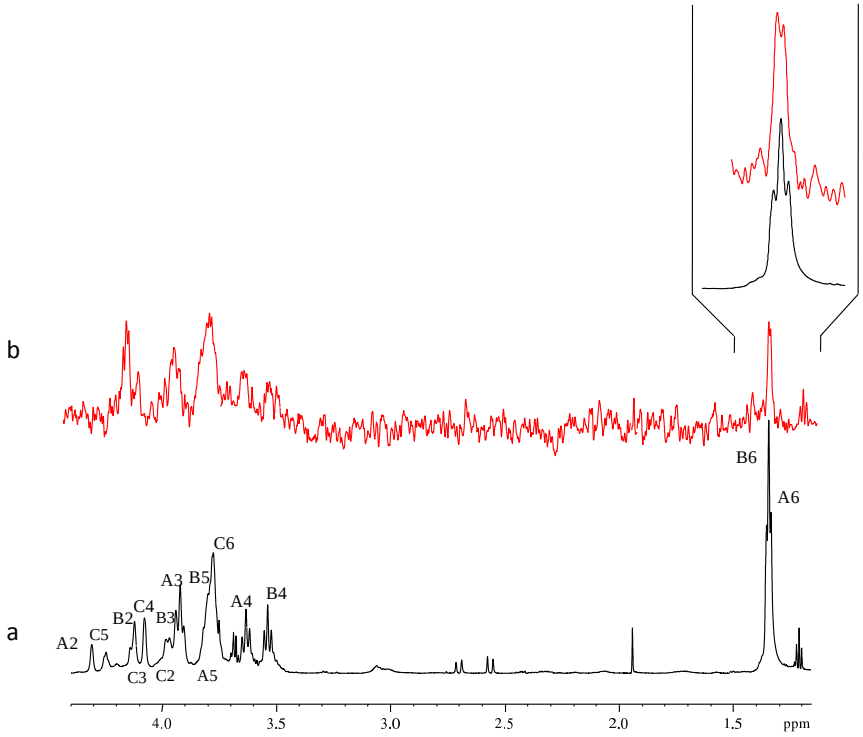


Fig. 6.3: Reference 1H NMR spectrum (**a**) and 1D STD NMR spectrum (**b**) of mixture **mAb 5D8** : **O-chain** from *B. anthina* 1:25.

Although a detailed quantification of the STD signals was not possible, it could inferred that the galactose residue gave rise to the highest STD enhancements, particularly through H-3, H-4 and H-6, likely closer to the binding site of the protein. Both rhamnose units were also involved in the interaction with the antibody, although their protons received less saturation from the mAb. In fact, lower STD effects were observed for H-3 and H-5 of both rhamnose units. However, both methyl groups gave rise to significant STD signals, although different in terms of intensity, suggesting that one of the Rha moieties is oriented closer to the mAb 5D8 than the other one.

6.2.2 The interaction of the O-polysaccharide chain with the 5D8 mAb. Tr-NOESY and MM/MD analysis

In order to study the bioactive conformation of the polysaccharide bound to the antibody, tr-NOESY experiments on a mixture of the 5D8:O-antigen (at a 1:20 molar ratio, Figure 6.4) were performed and compared to the corresponding NOESY spectra [31,32] of the O-antigen in the free state. The visual comparison of the different spectra showed no significant changes in terms of intensity of the cross peaks, thus indicating that the polysaccharide did not undergo significant conformational change upon recognition by the protein. The key inter-residual proton-proton distances for the key protons with conformational information were extracted from the cross peak intensities.

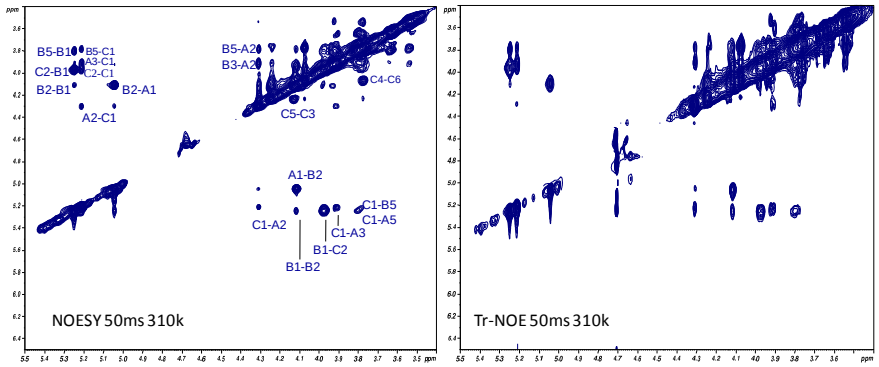


Fig. 6.4: NOESY (left) and tr-NOESY (right) on **polysaccharide chain** and on **mAb5D8: polysaccharide chain** 1 : 20.

Then, the investigation of the conformational features and the available conformational space for the glycosidic angles of the O-chain region was performed through MD simulations using the AMBER suite and the GLYCAM06 force field. First, the three basic constituent disaccharides (Rhaα 1-2Rha, Rhaα1-2Gal and Galα1-3Rha) were constructed for analyzing the energetically accessible conformational regions by employing systematic

molecular mechanic calculations using the standard protocol (see Section IV). The corresponding adiabatic energy maps for the glycosidic torsions φ (H1-C1-O-CX') and ψ (C1-O-CX'-HX') are reported in Figure 6.5.

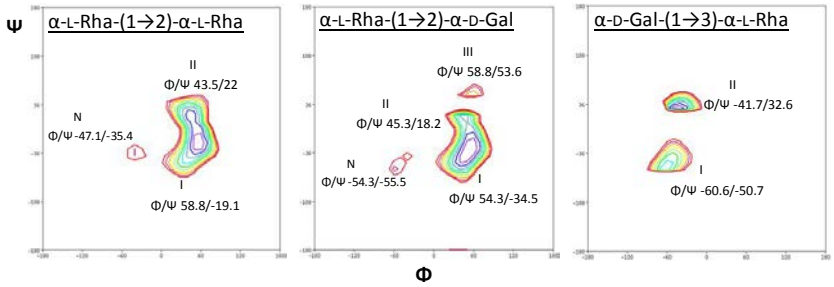


Fig. 6.5: Adiabatic energy maps of the three disaccharide connected by a glycosidic linkage in the repeting unit of O-chain from *B. anthina*. CH_2OH group was set in the gt conformation that is the most stable conformation for a galacto configured sugar unit. Minor local minima in non-exo anomeric region are indicated with letter N.

The inspection of the maps permitted to assess that, for each disaccharide, the global minimum value was in accordance with the *exo*-anomeric effect, although minor local minima populating the non-*exo* anomeric region were observed for both α-(1-2) linkages. Additionally, the calculations suggested the existence of greater flexibility around ψ angle than for φ torsion. Then, once the optimal values for φ and ψ dihedral angles had been estimated for each disaccharide, an octasaccharide was constructed and subjected to MD simulations in explicit water with AMBER, following the protocol described in the experimental section. The corresponding φ/ψ plots are reported in figures 6.6-6.7 and permitted to further support that the MD-based φ/ψ torsion values matched those predicted by the systematic analysis. The *exo* anomeric conformation was always observed for φ. For the α-(1-2) linkage, two major regions were identified with φ values ca. +50º and and ψ of -30 or 50º. degrees. The possible conformations for the α-(1-3)

linkage were characterized by values located around -60º for φ and 30º or -50º degrees for ψ.

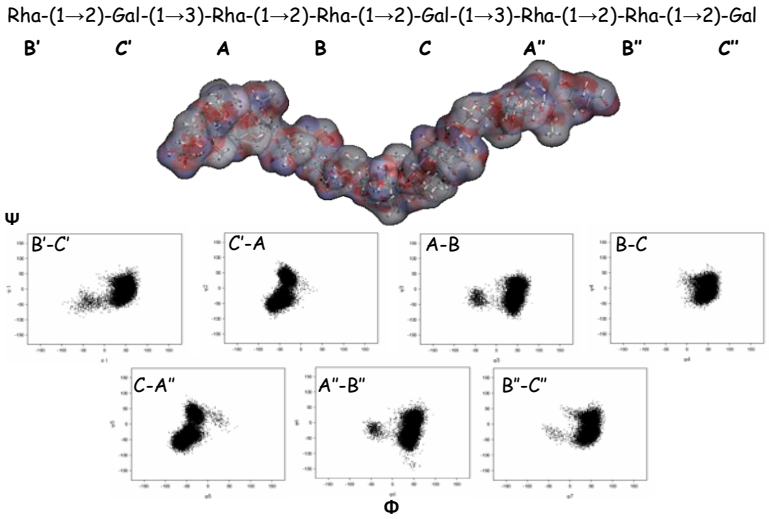


Fig. 6.6: Scatter plots of φ verus ψ values of octasaccharide.

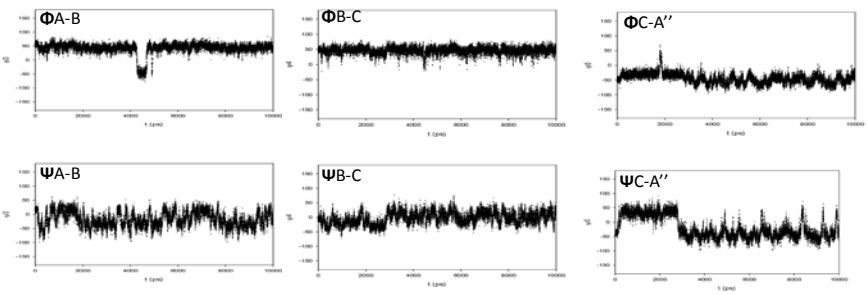


Fig. 6.7: Scatter plots of φ and ψ versus time (ps).

Then, to support the conclusions of the calculations, the *intra*- and *inter*-residual proton-proton distances between the sugar protons, especially between those around the glycosidic linkages, were extracted from the results of the MD simulations and compared to those estimated from the NOESY/trNOESY NMR experimental results (Table 6.2). A satisfactory agreement between the calculated and the experimentally observed distances was observed.

Similar results were obtained subjecting to MM/MD calculations (4 ns) a tetraeicosaccharide containing eight repeating units (data not shown).

Table 6.2 Experimental (NOESY, estimated error is 5%) and calculated (ensemble average from the MD simulation) *inter*-proton distances for the polysaccharide in the free state.

Distance	Experimental	Calculated
A1-B2	2.1	2.4
A3-C1	2.7	2.6
C2-B1	2.3	2.25
A2-C1	2.7	2.8
C2-C1	2.3	2.4
B5-B1	3.7	3.7
C1-B5	2.6	2.5
C1-A2	2.7	2.8

6.3 NMR studies of tri/hexasaccharide bound to mAb5D8

6.3.1 Synthesis of the tri and hexasaccharide

This work was in collaboration with the research group of prof. C. Vogel of University of Rostock, Germany. They synthesized the tri- and hexasaccharide, **7** and **11**, containing the repeating unit of the O-chain from *B. anthina* (Scheme 6.1), to be used as ligands of mb5D8, in order to access to residue-specific binding information and thus to accurately map the recognized polysaccharide epitopes. Monosaccharide **1**[34] was employed as glycosyl acceptor for the preparation of disaccharide **3** by coupling with glycosyl donor **2**. The *p*-methoxybenzyl group in 2-O-position of disaccharide **3** was then selectively removed. The resulting glycosyl acceptor **4** was then coupled with phenyl 2-*O*-acetyl-3,4-*O*-benzyl-1-thio-α-l-rhamnopyranoside **5** [35] to give trisaccharide **6**. Indeed, trisaccharide **6** is the key intermediate in our overall synthetic

strategy for the preparation of higher oligosaccharides (see table 6.3 for NMR assignments).

Scheme 6.1. Synthesis of the key intermediate **6** and the corresponding deprotected trisaccharide **7** containing the repeating unit of the O-chain from *B. anthina*. *Reagents and conditions:* (i) NIS, AgOTf, CH_2Cl_2, 20 min, -10–20 °C; (ii) CAN, $CH_3CN–H_2O$, 3 h, 20 °C; (iii) NIS, AgOTf, CH_2Cl_2, 1 h, -20–20 °C; (iv) NaOMe, MeOH, 24 h, 20 °C; (v) $Pd(OH)_2/C$, MeOH, H_2 atmosphere, 48 h, 20 °C.

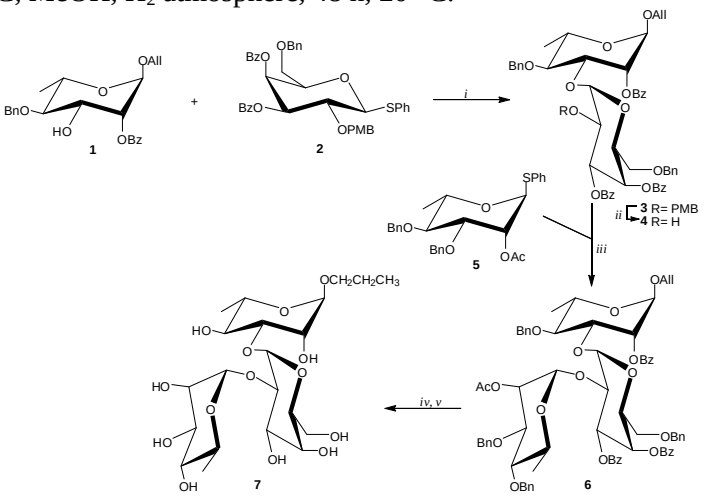
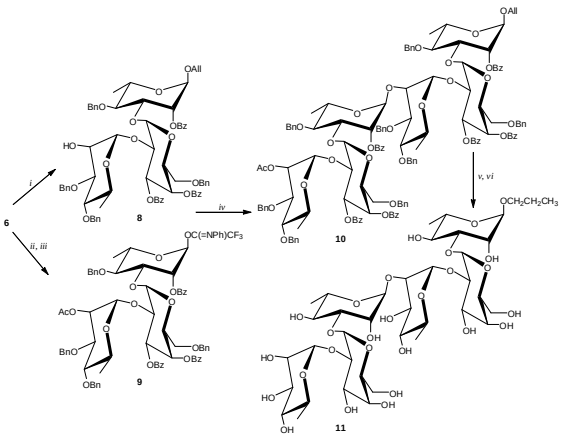


Table 6.3 Chemical shifts of ^1H and ^{13}C of the trisaccharide αRha-(1→2)-αGal-(1→3)-αRha-O-$CH_2CH_2CH_3$

	$^1H1$	$^1H2$	$^1H3$	$^1H4$	$^1H5$	$^1H6$	-O-CH_2	CH_2	CH_3
	$^{13}C1$	$^{13}C2$	$^{13}C3$	$^{13}C4$	$^{13}C5$	$^{13}C6$			
C	4.89	3.7	3.83	3.8	3.98	3.48			
Gal	94.8	74.7	70.0	69.4	70.8	61.0			
B	4.79	3.85	3.7	3.25	3.56	1.09			
Rha	102.1	70.0	68.9	71.8	70.1	16.8			
A	4.59	3.86	3.63	3.35	3.53	1.1	3.42/3.27	1.39	0.69
Rha	99.5	66.9	75.1	70.4	69.2	16.6	69.8/69.7	21.8	10.1

This key compound (**6**) was transformed either into glycosyl acceptor **8** by selective deacetylation or into glycosyl donor **9** by deallylation and the consecutive introduction of a leaving group (Scheme 6.2). Glycosylation of compounds **8** and **9** leads to hexasaccharide **10**. In the last step of the synthesis, the protecting groups of trisaccharide **6** and hexasaccharide **10** were removed to obtain the corresponding deprotected propyl oligosaccharides **7** and **11**.

Scheme 6.2. Synthesis of the hexasaccharide **11** starting from the key intermediate **6***:* (i) meth. HCl, 18 h, 20 °C; (ii) $PdCl_2$, MeOH–CH_2Cl_2, 16 h, 20 °C; (iii) $ClC(=NPh)CF_3$, $CsCO_3$, acetone, 3 h, 20 °C; (iv) TMSOTf, CH_2Cl_2, 30 min, -10–20 °C; (v) NaOMe, MeOH, 43 h, 20 °C; (vi) $Pd(OH)_2/C$, MeOH–H_2O, H_2 atmosphere, 72 h, 20 °C.



These molecules were employed for the detailed binding studies with the 5D8 antibody of Bcc.

6.3.2 The conformation of the constituent trisaccharide bound to mAb5D8. STD and TR-NOESY analysis

Thus, the interaction between 5D8 mAb with the synthetic trisaccharide was analyzed using STD NMR. The corresponding

STD spectrum showed that the trisaccharide ligand was recognized by the mAb5D8 antibody (Figure 6.8) and allowed deducing the ligand epitope. The inspection of the STD spectrum showed that all sugar residues were involved in the interaction with the antibody, although changes in the relative areas of individual signals were evidenced. In detail, the most prominent STD signals belonged to the galactose residue. Therefore, this semiquantitative analysis permits to assess that the Gal moiety is in closer contact to 5D8 than the two rhamnose residues. Fittingly, these results are in good agreement with the data collected on the polysaccharide chain of *B. anthina*, as described above. Additionally, only one methyl group, which belongs to rhamnose unit **B** showed a clear interaction with 5D8.

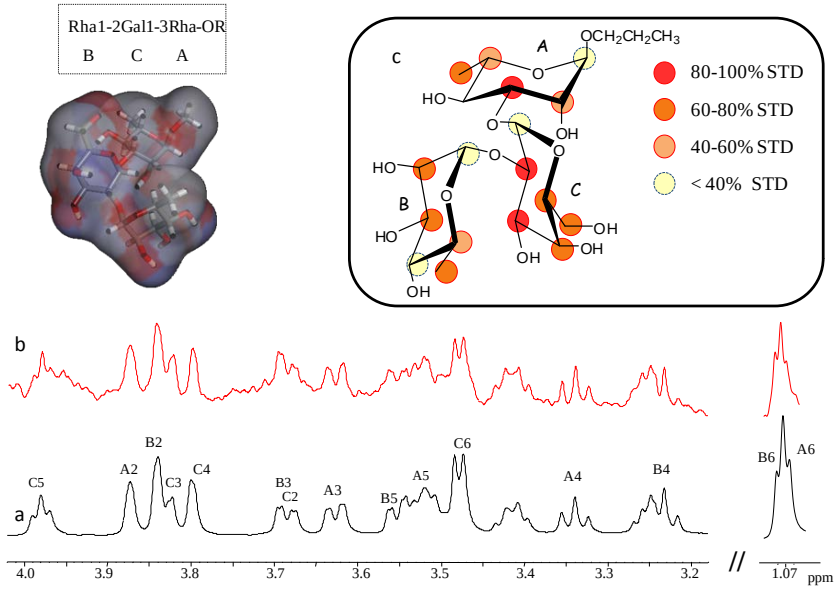


Fig. 6.8: Reference ^{1}H NMR spectrum (**a**) and STD 1D NMR spectrum (**b**) of mixture **mAb 5D8: trisaccharide 1: 50**. c. Chemical structure and epitope binding of trisaccharide to mAb 5D8; the protons with highest relative STD and, thus, the closest proximity to the receptor are highlighted in red.

Table 6.4. Experimental STD intensities of the **trisaccharide** bound to antibody mAb5D8 at different saturation times. STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 – exp(- k_{sat} * t_{sat}))$.

^{1}H	Experimental STD			STDmax	Ksat	STD (fit)	STD epitopes (fit)
	1s	2sec	5sec				
C3	100	100	82.9	0.4252	0.3152	0.1425	100
A2	60.1	72	65.6	0.4792	0.1726	0.0827	58
B2	22.9	57.6	58.3	0.6740	0.0930	0.0627	44
C6	78.1	75.5	62.4	0.3492	0.2853	0.1000	70
C5	31.2	60	56.2	0.6718	0.0870	0.0584	40.9
B3+C2	63.1	85	100	0.9887	0.1136	0.1123	78.8
A3	54.1	66.5	78.4	1.0851	0.0773	0.0838	58.8
B5	20	52	62.5	0.7280	0.1145	0.0834	58.5
A5	25	57.4	58.4	0.9669	0.0662	0.0640	44.9
C4	60	54.8	47.8	0.2408	0.3383	0.0814	57.1

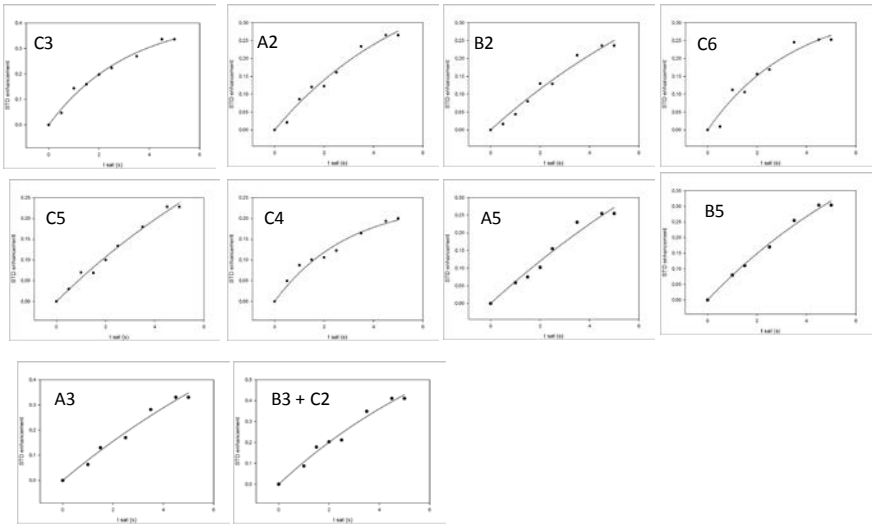


Fig. 6.9: STD build up curves for each proton of trisaccharide bound to mAb 5D8.

In order to refine these qualitative results, STD build up curves were derived (Table 6.4, Figure 6.9). [36-38] The spectra acquired at different saturation times were fitted to a monoexponential

equation: STD= STDmax(1 – exp(-ksat t)), where STD is the STD intensity of a given proton at a saturation time t, STDmax is the asymptotic maximum of the curve, and ksat stands for the observed saturation rate constant.

Then, the epitope of the trisaccharide was properly defined by normalizing all the values of the protons to the highest STDfit. The quantitative analysis confirmed galactose residue to be closer to the antibody. Indeed, the highest STD effect belonged to C3 and the protons C6 and C2 showed values above 70%. For the Rha residues, A2, A3 and B3, B5 exhibited STD effects around 60%.

These results were then complemented with tr-NOESY experiments. These experiments may permit deducing the bound conformation of the ligand, as well as providing information on the potential conformational changes in the trisaccharide upon binding (Figure 6.10).

By comparing the NOESY spectra performed on the ligand in the free state with the transferred NOE experiments obtained on the trisaccharide in the presence of the antibody (using a 1:10 protein/ligand molar ratio), binding of trisaccharide to mAb 5D8 was further demonstrated, since the NOE cross peaks changed from positive (for the free ligand) to negative (for the ligand in the presence of the mAb). This fact, which is correlated with a change in the effective rotational correlation time of the ligand, is a clear indication of binding (Figure 6.10) [39]. The detailed comparison of the NOESY spectra of the trisaccharide in the free and bound state showed few changes in the relative intensities of the cross peaks, strongly suggesting that no drastic changes in the major conformation of ligand indeed occurred upon binding.

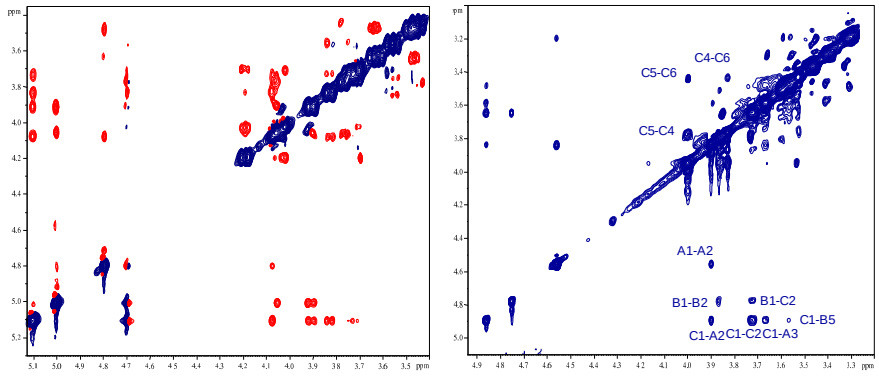


Fig. 6.10: NOESY (left, at 400MHz) and tr-NOESY (right, at 600MHz) on **trisaccharide** and on **IgG5D8: trisaccharide** 1 : 10.

Table 6.5 Experimental key interproton distances for the **trisaccharide** in the free and bound states. Estimated errors are below 5%. The corresponding key distances for the two major minima with either positive or negative ψ values are also given.

Distance	Experimental Free state	Experimental Bound state	Calculated	Calculated
			$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/50°	$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/-30°
B1-C2	2.4	2.7	2.7	2.2
			$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ = -47°/32°	$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ =-60°/-50°
C1-A3	2.4	2.6	2.3	2.9

Nevertheless, the experimental data, especially the B1-C2 distance (Table 6.5), allowed deducing that in the bound state, for the α-Rha-(1→2)-α-Gal linkage, the minimum characterized by positive values of both inter-glycosidic dihedral angles, φ (60°) and ψ (50º) is now the most populated geometry of the trisaccharide upon binding. Therefore, although some motion still remains, the antibody favors one of the existing conformations in the free state.

6.3.3 The conformation of the constituent hexasaccharide bound to mAb5D8. STD and TR-NOESY

Further details on the fine description of the interaction were collected by employing the corresponding hexasaccharide

[α-Rha-(1→2)-α-Gal-(1→3)-α-Rha-(1→2)-α-Rha-(1→2)-α-Gal-(1→3)-α-Rha],containing the repeating unit of O-chain from LPS of *Burkholderia anthina.*

The quantitative analysis of the STD NMR spectra (Figure 6.11), based on the STD build up curves (Figure 6.12), perfectly fitted with the results described for the poly- and the trisaccharide.

The STD epitope fit values (Table 6.6) underlined the key role of the galactose residue in the binding event, since highest STD effects were observed for H3, H5 and H6 of the Gal moiety (above 80%). According to the STD data, the two residues of rhamnose also contributed to the molecular recognition process, although in a minor extent.

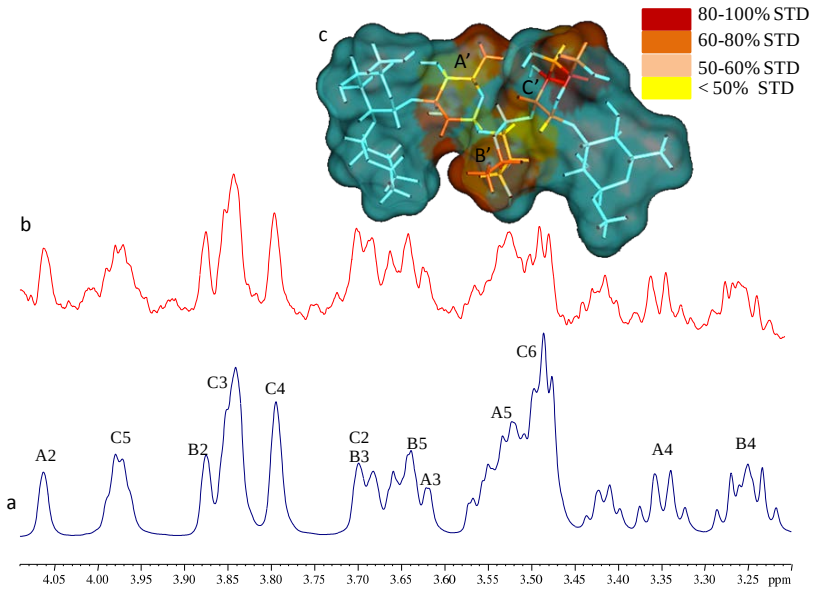


Fig. 6.11: Reference ^{1}H NMR spectrum (**a**) and STD 1D NMR spectrum (**b**) of mixture **mAb 5D8: hexasaccharide 1: 50**. c. STD-derived epitope mapping on the molecular envelope of hexasaccharide with color coding from the highest(red) to lowest (yellow) observed STD effect.

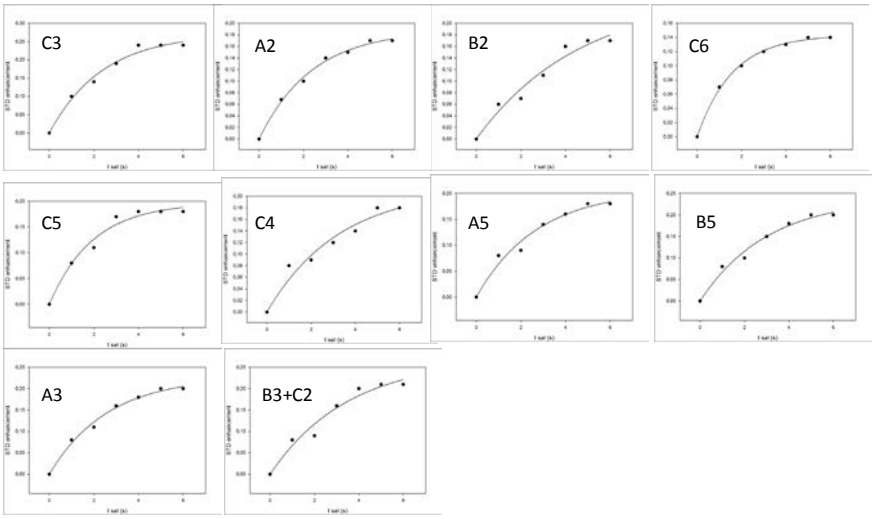


Fig. 6.12: STD build up curves for each proton of hexasaccharide bound to mAb 5D8.

Table 6.6 Experimental STD intensities of the **hexasaccharide** bound to antibody mAb5D8 at different saturation times. STDmax were calculated by fitting the data to a monoexponential equation: $STD(t_{sat}) = STD_{max} * (1 - \exp(- k_{sat} * t_{sat}))$.

^1H	Experimental STD 1sec	2sec	5sec	STDmax	Ksat	STD (fit)	STD epitopes (fit)
C'3	100	100	100	0,2718	0,4177	0,1135	100
A'2	69	83	70.8	0.1882	0.4198	0.0790	69.6
B'2	60	80	70.8	0,2482	0,2146	0,0532	46.8
C'6	85	91	58.3	0,1435	0,6235	0,0894	78.8
C'5	80	91.6	75	0,1968	0,5145	0,1012	89.1
B'3+C'2	67	75	87.5	0,2698	0,2854	0,0770	67.8
A'3	80.5	90	83.3	0.2278	0.3837	0.0874	77
B'5	81	83	83.3	0,2393	0,3291	0.0787	69.3
A'5	77	75	75	0,2064	0,3681	0,0759	66.8
C'4	80	89	75	0,2141	0,3050	0,0653	57.5

NOESY experiments were then used to deduce the bound conformation of the hexasaccharide when bound to mAb 5D8 antibody. Generally speaking, the spectrum obtained in the presence of the antibody was similar to that in the free state,

although some key cross peaks displayed noticeable changes in their relative intensities (Figure 6.13).

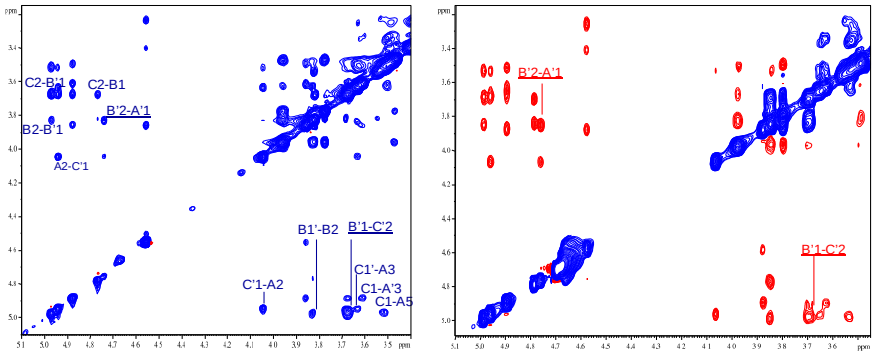


Fig. 6.13: NOESY (left) and tr-ROESY (right) on **hexasaccharide** and on **IgG5D8: hexasaccharide** 1 : 10.

Table 6.7 Experimental key interproton distances for the hexasaccharide in the free and bound states. Estimated errors are below 5%. The corresponding key distances for the different glycosidic torsion angles with either positive or negative Ψ values are also given.

[α-Rha-(1→2)-α-Gal-(1→3)-α-Rha-(1→2)-α-Rha-(1→2)-α-Gal-(1→3)-α-Rha-OR]

B C A' B' C' A

Distance	Experimental Free state	Experimental Bound state	Calculated	Calculated
			$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/50°	$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/-30°
B1-C2	2.5	2.6	2.7	2.2
			$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ =-47°/32°	$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ =-60°/-50°
C1-A'3	2.7	2.6	2.4	2.5
			$\Phi/\Psi_{\text{Rha-(1→2)-Rha}}$ = 45°/20°	$\Phi/\Psi_{\text{Rha-(1→2)-Rha}}$ = 60°/-20°
A'1-B'2	2.9	2.4	2.7	2.4
			$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/50°	$\Phi/\Psi_{\text{Rha-(1→2)-Gal}}$ = 60°/-30°
B'1-C'2	2.4	2.7	2.7	2.2
			$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ =-47°/32°	$\Phi/\Psi_{\text{Gal-(1→3)-Rha}}$ =-60°/-50°
C'1-A3	2.6	2.55	2.5	2.6

In detail (Table 6.7), a reduction of intensity of the α-Rha-(1→2)-α-Gal cross peak and an increase of the α-Rha-(1→2)-α-Rha NOE contacts were detectable in the bound with respected to the free state. From the obtained data, it can be safely concluded that the interaction with the antibody induced a conformational selection process, with a preference for geometries with Φ/Ψ torsion angle values 60°/50° for the α-Rha-(1→2)-α-Rha glycosidic linkage and 60º/-20º for the α-Rha-(1→2)-α-Gal one (minima III and I in Figure 6.5, respectively).

6.4 Conclusion

A multidisciplinary approach combining synthetic organic chemistry with NMR experiments and molecular mechanics and dynamics simulations has been carried out to explore the molecular recognition process between the polysaccharide chain of the LPS from *Burkholderia anthina* and the 5D8 monoclonal antibody. A chemical "reductionistic" approach has been employed by also investigating the interaction of the the small tri- and hexa-saccharide repeating units of the polysaccharide with the antibody.
The STD NMR data have permitted to demonstrate that the entire trisaccharide repeating unit is involved in the interaction with the antibody. However, although the Rha units are close to the antibody surface, the galactose residue appears to afford the best intermolecular contacts, since it provided the highest STD effects in the polysaccharide and in its smaller fragments. These results suggest that the galactose residue was the O-chain sugar unit playing a key role in the mechanism of interaction.
Furthermore, NOESY experiments performed on free and bound state assisted with molecular modeling tools have permitted to unravel the bound conformation of the tri and hexa-saccharide. The recognition event takes place with a conformational selection process, in which the major conformation existing in solution is indeed bound by the antibody. This fact is frequently found in

saccharide recognition by proteins [39], since it is a proper manner to minimize the entropic cost of the binding process. Nevertheless, some cases also exist in which the protein recognizes a minor conformation of the saccharide, [40], especially when dealing with glycomimetics [41].

Thus, epitope information at atomic resolution and key clues for the interaction of LPS with a key pathogen have been deduced from this multidisciplinary approach. This study may thus open new avenues for conceiving novel glycoconjugate-based immunogens founded on the essential carbohydrate skeleton, as well as the initial step for designing new higher-affinity ligands as vaccine candidates.

References

[1] De Soyza, A.; Silipo, A.; Lanzetta, R.; Govan, J.R.; Molinaro, A. *Innate Immun*., **2008**. 14(3):127-44.
[2] Hart, C. A.; and Winstanley, C. *Br. Med Bull* **2002**. *61*, 81-96.
[3] Silipo, A.; Molinaro, A.; Ieranò, T.; De Soyza, A.; Sturiale, L.; Garozzo, D.; Aldridge, C.; Corris, P.A.; Khan, C.M.; Lanzetta, R.; Parrilli, M. *Chemistry*., **2007**.13(12):3501-11.
[4] Carvalho, A.P.; Ventura, G.M.; Pereira, C.B.; Leão, R.S.; Folescu, T.W.; Higa, L.; Teixeira, L.M.; Plotkowski, M.C.; Merquior, V.L.; Albano, R.M.; Marques, E.A. *APMIS*., **2007.** 115(4), 311-8.
[5] Sousa, S.A.; Ramos, C.G.; and Leitão, J.H. *International Journal of Microbiology,* **2011.**
[6] Bamford, S.; Ryley, H.; Jackson, S.K. *Cell Microbiol*., **2007.** 9(2), 532-43.
[7] Zlosnik, J.E.A.; Gunaratnam, C L.; Speert, D.P. *Front Cell Infect Microbiol*., **2012.** 2:67.
[8] Butler, S.L.; Nelson, J.W.; Poxton, I.R.; Govan, J. R. *FEMS Immunol. Med. Microbiol*. **1994**. 8, 285–292.
[9] Arlene, D.; Vinion-Dubiel,.A.D.; Goldberg, J.B. *Journal of endotoxin research*, **2003**, 9 (4).
[10] Ieranò, T.; Silipo, A.; Cescutti, P.; Leone, M.R.; Rizzo, R.; Lanzetta, R.; Parrilli, M.; Molinaro, A. *Chemistry*. **2009** 15(29), 7156-66.
[11] Isshiki, Y.; Zahringer, U.; Kawahara, K. *Carbohydrate Research*, **2003**. 338, 2956-2666.
[12] Leone, S.; Molinaro, A.; Gerber, I.B.; Dubery, I.A.; Lanzetta, R.; Parrilli, M. *Carbohydr Res*., **2006**. 29;341(18), 2954-8.
[13] Cimino, P.; Corsaro, M.M.; De Castro, C.; Evidente, A.; Marciano, C.E.; Parrilli, M.; Sfalanga, A.; Surico, G. *Carbohydr Res*., **1998**. 307(3-4), 333-41.
[14] Vinion-Dubiel, A.D.; Goldberg, J.B. *J. Endotoxin research*, **2003**. 9, 201-213.
[15] AuCoin, D.P.; Crump, R.B.; Thorkildson, P.; Nuti, D.E.; LiPuma, J.J. and Kozel, T. R. *J.of Medical Microbiology* , **2010**. 59, 41-47.
[16] Mayer, M. and Meyer,B. *J. Am. Chem. Soc*., **2001**. 123, 6108-6117.
[17] Meyer B. and Peters, T. *Angew. Chem. Int. Ed. Engl.*, **2003**. 42, 864-890.

[18] Sigurskjold, B.W.; Bundle, D.R.J. *Biol. Chem*., **1992**. 267, 8371–8376.
[19] Vyas, N.K.; Vyas, M.N.; Chervenak, M.C.; Johnson, M.A.; Pinto, B.M.; Bundle, D.R.; Quiocho, F.A. *Biochemistry*, **2002**. 41, 13575–13586.
[20] Mu¨ller-Loennies, S.; Brade, L.; MacKenzie, C. R.; Di Padova, F. E.; Brade, H. *J. Biol. Chem*., **2003**. 278, 25618–25627.
[21] Liao, X.; Poirot, E.; Chang, A.H.C.; Zhang, X.; Zhang, J.; Nato, F.; Fournier, J.-M.; Kova´cˇ, P.; Glaudemans, C. P. J. *Carbohydr. Res.*
[22] Nitz, M.; Ling, C.-C.; Otter, A.; Cutler, J. E.; Bundle, D. R. *J. Biol. Chem*., **2002.** 277, 3440–3446.
[23] Tsvetkov, Y.E.; Burg-Roderfeld, M.; Loers, G.; Ardá, A.; Sukhova, E.V.; Khatuntseva, E.A.; Grachev, A.A.; Chizhov, A.O.; Siebert, H.C.; Schachner, M.; Jiménez-Barbero, J.; Nifantiev, N.E. *J Am Chem Soc*., **2012**. 134(1), 426-35.
[24] Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chemistry*. **2011**, 17(5), 1547-60.
[25] Houliston, R.S.; Jacobs, B.C.; Tio-Gillen, A.P.; Verschuuren, J.J.; Khieu, N.H.; Gilbert, M.; Jarrell, H.C. *Biochemistry*., **2009.** 48(2), 220-2.
[26] Oberli, M.A.; Tamborrini, M.; Tsai, Y.-H.; Werz, D.B.; Horlacher, T.; Adibekian, A.; Gauss, D.; Moller, H.M.; Pluschke, G.; Seeberger, P.H. *J. Am. Chem. Soc*., **2010.** 132, 10239–10241.
[27] Carillo, S.; Silipo, A.; Perino, V.; Lanzetta, R.; Parrilli, M.; Molinaro, A. *Carbohydate research*, **2009**. 344, 1687-1700.
[28] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A. *Chem. Eur. J.*, **2012**. doi: 10.1002/chem.201200488.
[29] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*., **2012**. 22(10), 1387-98.
[30] Squeglia, F.; Marchetti, R.; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *J Am Chem Soc*., **2011**. 28;133(51), 20676-9.
[31] Marchetti, R.; Lanzetta, R.; Michelow, I.C.; Molinaro, A.; Silipo, A. *Eur. J. Org. Chem*., **2012.** 27, 5275–5281.
[32] Angulo, J.; Nieto, P. *Eur Biophys J.*, **2011**. 40, 1357-1369.

[33] NMR Spectroscopy of Glycoconjugates (Eds.: J. Jimenez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**
[34] Pinto, B.M.; Morissette, D.G. *J. Chem. Soc. Perkin Trans. 1* **1987.** 9-14.
[35] Lau, R.; Schuele, G.; Schwaneberg, U.; Ziegler, T. *Liebigs Annalen,* **1995**. 10, 1745–1754.
[36] Jayalakshmi, V.; Krishna, N.V. *J. Magn. Reson*, **2002**. 155, 106-118.
[37] Yan, J.L.; Kline, A.D.; Mo, H.P.; Shapiro, M.J.; Zartler, E.R. *J. Magn. Reson*., **2003**. 163, 270276.
[38] Szcepina, G.M.; Bleile, W.D.; Pinto, B.M. *Chem. Eur. J*., **2011**. 17, 11446-11455.
[39] Roldos, V.; Cañada, F:J:; Jimenez-Barbero, J. *ChemBioChem,* **2011**. 12, 990-1005.
[40] Gabius, H.J.; André, S.; Jimenez- Barbero, J.; Romero, A.; Solís, D. *Trends Biochem Sci*., **2011.** 36(6), 298–313.
[41] Martin-Santamaria, S.; Gabius, H.J.;Jimenez- Barbero, J. *Pure and Applied Chemistry*, **2011**.

Chapter 7

LysM – chitooligosaccharides interaction

7.1 Introduction

The innate immunity constitutes the first line of defence against attempted microbial invasion and is a well-described phenomenon in vertebrates and insects while is less studied in plant system [1]. Like mammals, plants have acquired the ability to recognize, as non-self, invariant microbial associated molecular patterns, characteristic of microbial organisms but completely absent in potential host plants [2]. MAMP molecules, such as flagellin, EF-Tu, peptidoglycan, LPS, chitin and β-glucan, have been isolated from the cell walls and secreted materials of both bacteria and fungi, and their biological activities to induce defense responses in various plant species have been well documented.

On the other hand, plant receptor proteins carrying extracellular lysine motifs (LysM) serve as biochemical modules implicated in the recognition of these molecules, facilitating the establishment of plant resistance to bacterial and fungal pathogens.

In this study, the interaction between LysM domains and fungal chitooligosaccharides has been deeply investigated by using a multidisciplinary approach combining NMR spectroscopy with protein mutagenesis and docking analysis.

7.2 LysM proteins as key mediators in plant immune response

Sensing microbial surface patterns via host-encoded pattern recognition receptors is a prerequisite for the activation of antimicrobial defenses in multicellular organisms [3].

The plant immune system acts through two major types of plasma membrane PRRs, that are distinguishable, by their ectodomain structures, in leucine-rich repeat receptor proteins (LRR-RP) and leucine-rich repeat receptor kinases (LRR- RK), which preferentially bind proteinaceous ligands.
More recently, plant receptor proteins and receptor kinases carrying extracellular lysin motifs have been shown to act as biochemical modules for the direct as well as indirect recognition of microbial glycans, commonly components of surface glycoconjugates [4-6].
LysM domains are ubiquitous and ancient proteins, found in many living organisms, containing approximately 40 amino acids and characterized by a βααβ secondary structure (Figure 7.1 [7]) [8].

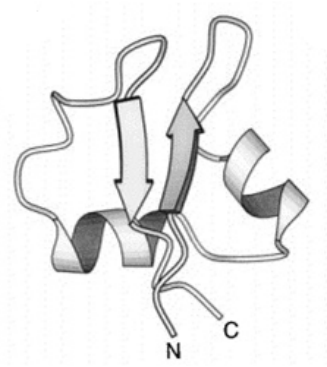


Fig. 7.1: Secondary structure of LysM motif. Two helices pack onto the same side of a two-stranded anti-parallel β sheet.

In general, LysM motifs mediate both plant immune response to microbial infections and plant symbiosis; actually, they mediate recognition of GlNAc backbone, present in peptidoglycan, lipochitooligosaccharides and fungus derived chitin [2] (Figure 7.2).

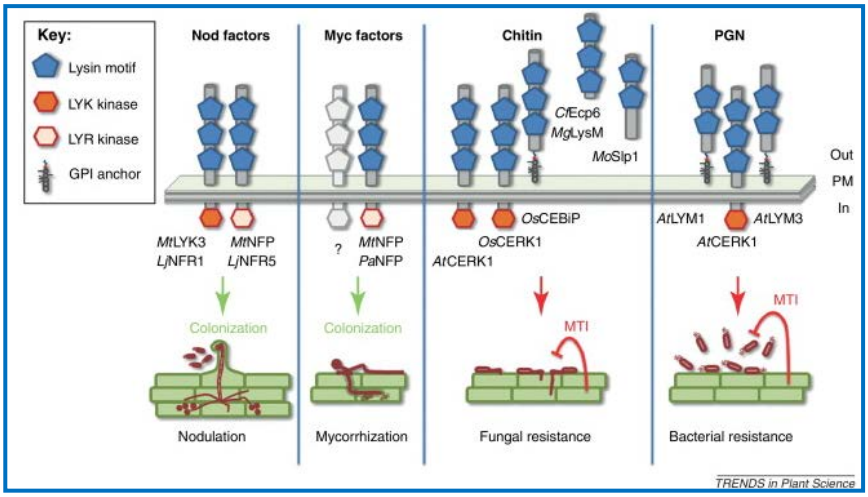


Fig. 7.2: Plant lysin motif (LysM) receptors mediating the establishment of symbiosis (Nodulation and Mycorrhizination) and activation of plant immunity (Fungal and Bacterial resistance).

7.2.1 Plant LysM proteins mediating symbiosis

Important forms of symbiosis between soil-borne microbes and host plants include the association of leguminous with Rhizobia and mycorrhizal fungi.

The legume–Rhizobium symbiosis is a sequential process that starts with rhizobial attachment to plant root hair tips through the interaction between plant LysM domains and bacterial NOD factors (Figure 7.3B). The bacteria transform the atmospheric nitrogen (N_2) into NH_4^+, which can then be used by the plant; and in turn, they benefit from sugars coming from the photosynthesis.

LysM proteins are crucial elements also for the establishment of legume-mycorrhizum relationship that represents the most widely spread plant symbiotic interaction. This type of symbiosis consists in root colonization by different types of fungi [9] (Figure 7.3A). The plant, thanks to the photosynthetic process, provides the fungus with carbohydrates and in return it improves plant water and nutrient uptake capacity.

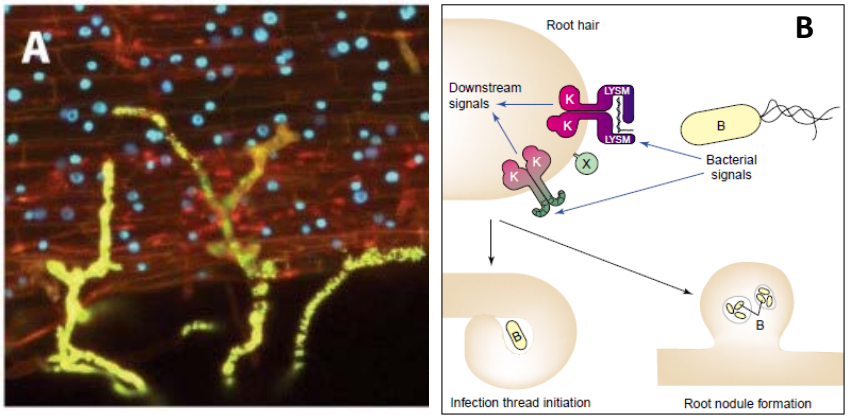


Fig. 7.3: A) Establishment of the mycorrhizal symbiosis producing swollen structures called hyphopodia (in yellow). B) The recognition of NOD factor (lipochitooligosaccharides) from plant LysM domains leads to the development of nodules where the biological process of nitrogen fixation initiates. (Adapted from Gough et al, 2011 and Spaink, 2004).

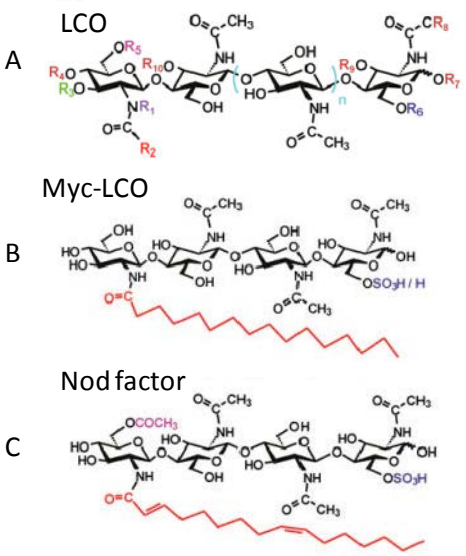


Fig. 7.4: A) Generic symbiotic LCO (lipo-chitooligosaccharide) showing sites of chemical substitutions. B) Structure of sulphated and nonsulphated Myc-LCO of Glomus intradices. C) Structure of major Nod factor of *Sinorhizobium meliloti*. (Adapted from Gough and Cullimore, 2011).

For the initiation of both symbiotic processes, a signal exchange between the two partners is needed in order to identify each other as a “friend” and not as an “enemy”. These symbiotic signals consist in molecule with remarkably similar primary structure (Figure 7.4), all containing N-acetylglucosamine residues.

7.2.2 LysM modulating plant immunity response to pathogens

Plant LysM domain proteins are implicated in the activation of plant immune response triggered by PAMPs, such as bacterial peptidoglycan and fungal chitin oligosaccharides.
Recent studies demonstrated that the carbohydrate backbone of PGN is important for its immunogenic activity and that it is recognized, in a receptor-mediated manner [3], by two LysM receptor proteins, known as LYM1 and LYM3, together with AtCERK1 (Chitin Elicitor Receptor Kinase 1 of *Arabidopsis thaliana*). The plasma membrane-tethered proteins, lym1 and lym3, are able to physically bind peptidoglycan but lack any intracellular domain, whereas, AtCERK1 is a transmembrane protein with kinase activity, required for the initiation of an intracellular signaling cascade.
The Arabidopsis LYM1 LYM3/CERK1 perception system seems similar to the OsCEBiP (Chitin Elicitor Binding Protein of *Oryza sativa*) OsCERK1(Chitin Elicitor Receptor Kinase 1 of *O. sativa*) complex that mediates chitin perception and immunity to fungal infection in rice [4,10,11]. Structurally, OsCEBiP is very similar to lym1 and lym3; it is a protein that contains extracellular LysM domains but lacks a recognizable intracellular signaling domain. Knock-down experiments of specific genes showed that both these LysM proteins are required for chitin elicitor signaling in rice, whereas CEBiP type molecules are not involved in chitin signaling in Arabidopsis, indicating the difference between the chitin receptor systems in these model plants [11]. In the case of rice chitin receptor, yeast two-hybrid analyses suggested that CEBiP and

OsCERK1 have a potential to form a receptor complex as a hetero- or homo-oligomer through the interaction of their ectodomains. Co-immunoprecipitation studies further indicated that CEBiP and OsCERK1 form a heterooligomeric receptor complex ligand dependently [10].
Although both CEBiP and OsCERK1 have lysin motifs, in their ectodomains, CEBiP should be the major component for the perception of chitin oligosaccharides in rice, because the knock-down of *CEBiP* almost abolished the ability of the plasma membrane to bind a radio-labeled chitin oligosaccharide [4] and also OsCERK1 was shown not to bind chitin [11].

The detailed analysis of ligand recognition by CEBiP-like molecules and succeeding activation of the receptor kinase is critically important to understand the molecular mechanisms leading to the activation of downstream defense responses triggered by these MAMPs. Such information would also contribute for the design of novel receptor molecules suitable for future biotechnological application.
STD-NMR analysis of chitin oligosaccharides in the presence of the ectodomain of CEBiP resulted in the mapping of the epitopes that interact with that rice receptor. In addition, NMR studies supported docking analysis, site directed mutagenesis and biochemical experiments on various deletion mutants as well as chimeric proteins for the ectodomain of CEBiP, in identifying critical amino acid residues for ligand recognition.

7.3 Biochemical studies on the ectodomain of CEBiP

As described previously, CEBiP has two lysin motifs (LysMs) in the region of Y^{85}-P^{131} (LysM1) and Y^{149} to P^{192} (LysM2) in the ectodomain (Figure 7.5) [4].

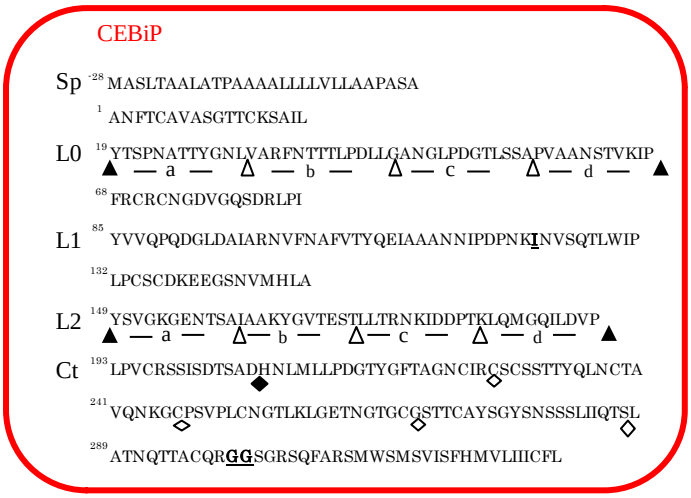


Fig 7.5: Amino acid sequence of CEBiP. Sp, signal peptide; L0, LysM0; L1, LysM1; L2, LysM2; Ct, C terminal region. The LysM0 and LysM2 divided into four parts, from a to d, and the C terminal region also divided into four parts from ◆ to ◇.

To evaluate the contribution of CEBiP LysMs for chitin binding, deletion mutants were constructed and expressed in tabacco BY-2 cells, by the group of professor Shibuya, of University of Tokyo, Japan (Figure 7.6). Then, the binding of the mutant proteins to chitin oligosaccharides were examined by using affinity labeling with biotinylated *N*-acetylchitooctaose (GN8-Bio) [11].

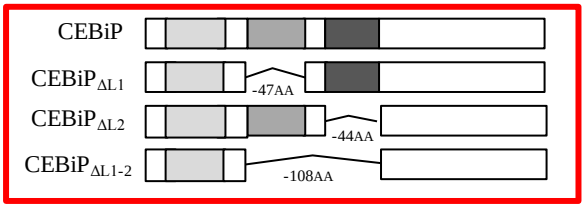


Fig 7.6: The deletion of each LysM and LysM region of CEBiP.

The obtained results indicated that the deletion of LysM2 affected the expression of mutant proteins; whereas, affinity labeling

experiments showed the importance of LysM1 for the chitin binding activity of CEBiP (data not shown).
To further characterize the contribution of LysM1 and 2 for chitin binding, their interchangeability in the ectodomain of CEBiP has been examined. Chimeric constructs, containing LysM2 replaced with the LysM1 or LysM1 replaced with LysM2, have been expressed and analyzed. The results showed again the importance of both lysM2 and lysM1 for the expression and the chitin binding respectively (data not shown).
The function of each LysM in CEBiP by replacing them with the corresponding regions of an *Arabidopsis* CEBiP homolog, which does not bind chitin, has also been analyzed. All collected data underlined that the ectodomain of CEBiP played a major role for the perception of chitin oligosaccharides. Indeed, the replacement of only LysM1 with the corresponding region of an Arabidopsis CEBiP homolog which does not bind chitin (LYM1) resulted in the complete loss of chitin binding, whereas, similar substitution with the region from a chitin-binding homolog (LYM2) maintained the ability to bind chitooligosaccharides.
In addition to the analysis of LysM domains, the possible contribution for chitin binding of a stretched C terminus region in the CEBiP ectodomain, which possesses 136 amino acid residues with a high cysteine content, has also been evaluated. All deletion mutants completely retained the chitin binding activity indicating that the C-terminal region did not affect chitin binding activity significantly.

7.4 STD NMR epitope mapping of chitin oligosaccharides

STD NMR spectroscopy [12-15] was used to map interacting epitopes of chitin oligosaccharides and to reveal the molecular mechanism of interaction with the LysM domains of CEBiP. In detail, we acquired a series of STD NMR spectra using, as ligands, chitin fragments of different length, starting from tri- to octa- saccharides,

with the aim to determine which was the best oligosaccharide length in the interaction with the LysM domains. As it was shown that LysM1 is important for chitin binding and also LysM2 contributes for stable expression, we decided to use the LysM1-LysM2 region of CEBiP for STD NMR studies. We recombinantly produced LysM1-LysM2 domains (LysM1-2, residues 113-221) as fused to a carrier protein (Trx-tag), since Saturation Transfer Difference (STD) NMR is highly successful for high molecular weight systems.

Prior to NMR experiments, we analysed the conformational state of LysM1-2 once Trx had been cleaved using far-UV circular dicroism (CD) experiments (Figure 7.7). Spectra are typical of proteins with a high content of β-sheets, with a minimum at 204 nm; consistent with the typical LysM fold, CD signal at 222 nm also evidences the existence of a α-helix (Figure 7.7).

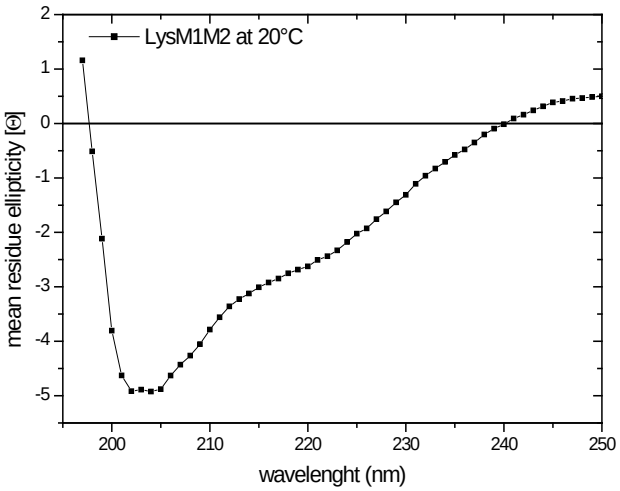


Fig. 7.7: CD spectrum of LysM1-LysM2 (residues 113-221) in 50 mM TrisHCl, 150 mM NaCl, and 5 mM DTT.

Since thioredoxin has been used as a stabilizing partner of CEBiP LysM (LysM1-LysM2), we optimized the STD conditions to

perform Saturation Transfer Double Difference (STDD) NMR experiments useful to overcome strong background signals and to subtract possible artifacts due to unspecific interactions between ligands and the stabilizing partner [16-18]. The STDD NMR requires the acquisition of an additional STD spectrum of the receptor in absence of ligands, that is then subtracted from the STD spectrum acquired in the presence of ligands. In our case, STDD spectra have been obtained performing a subtraction between the normal STD NMR spectrum and those performed in the presence of thioredoxin alone.

7.4.1 The interaction of LysM domains with short length chitin oligosaccharides

Although a qualitative analysis of the STDD spectra using chitin tri-, tetra-, and hexa-saccharides as ligands demonstrated their binding with the LysM domains (Figure 7.8), a deeper look to the spectra also showed that these molecules weakly bound to LysM motifs. In detail, low STD effects were observed on tri- and tetra-saccharides (Figure 7.8a,b), while higher signals were obtained using hexasaccharidic substrate (Figure 7.8c).
The corresponding STD-derived binding epitopes obtained on the hexasaccharide has been reported in Figure 7.9; although the analysis was almost qualitative, it could be inferred that prominent STD signals belonged to H6 protons, although also the remaining part of the sugar units significantly contributed to the binding.

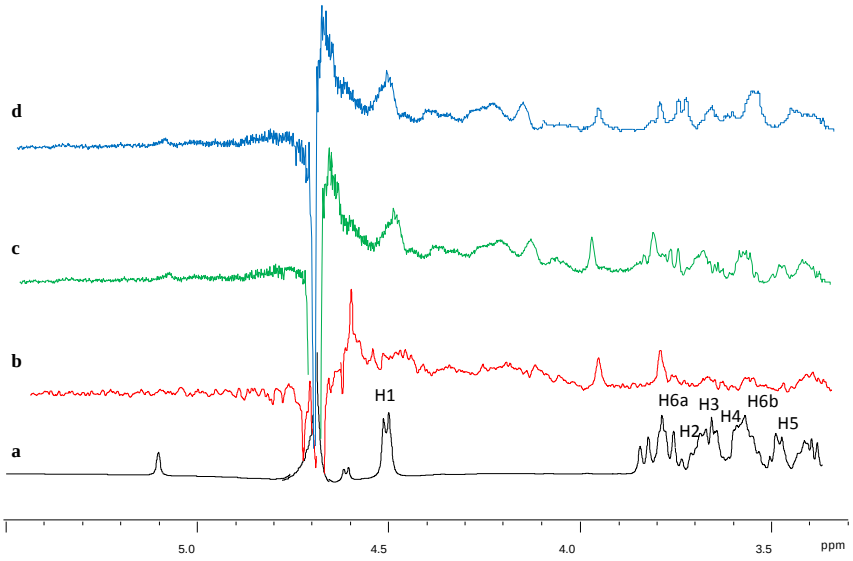


Fig. 7.8: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectra of mixture LysM1-2-trx : trisaccharide (b), tetrasaccharide (c), hexasaccharide (d)1:100.

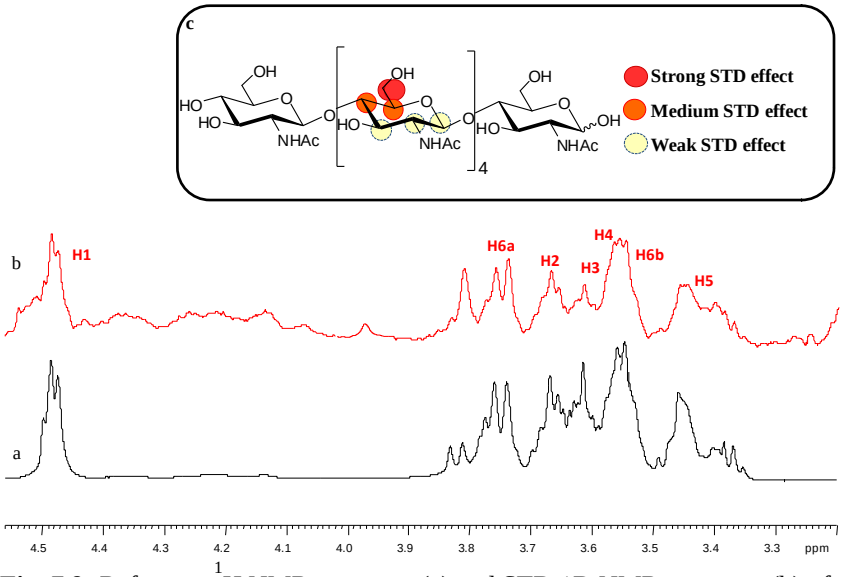


Fig. 7.9: Reference ^{1}H NMR spectrum (a) and STD 1D NMR spectrum (b) of mixture LysM1-2-trx : hexasaccharide 1:100.
c. Chemical structure and epitope binding of chitohexasaccharide to LysM1-2-trx.

7.4.2 The interaction of LysM domains with hepta- and octa-chitin

oligosaccharides. STD NMR and tr-NOESY analysis

Based on the above findings, we focused our attention on studying chitooligosaccharides with higher molecular weight, that is hepta- and octasaccharides. The STDD NMR spectrum of the heptasaccharide bound to the protein has been reported in Figure 7.10.

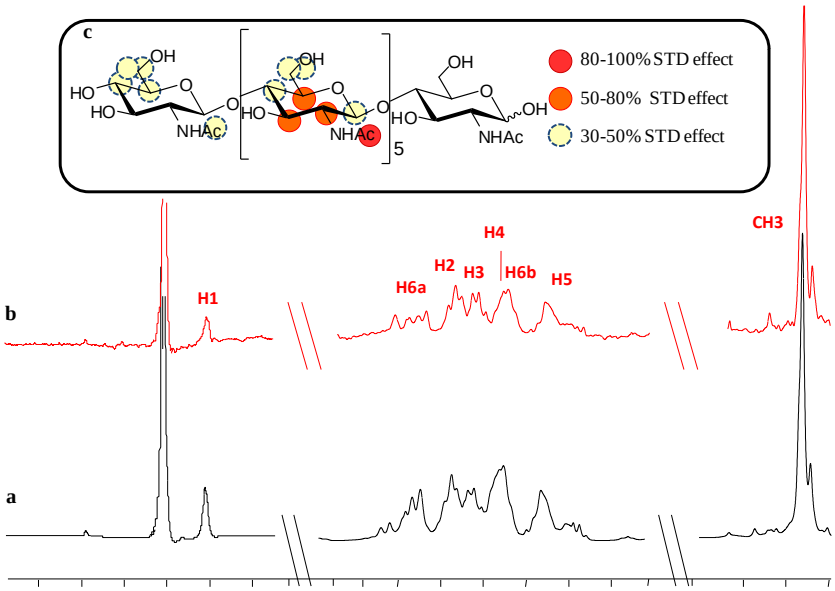


Fig. 7.10: Reference ^{1}H NMR spectrum (a) and STDD NMR spectrum (b) of mixture LysM1-2-trx : heptasaccharide 1:100.c. Chemical structure and epitope binding of chitoheptasaccharide to LysM1-2-trx.

In order to precisely map ligand epitopes in close contact with the protein, we acquired STD build up curves by collecting spectra at different saturation times [19-22]. This method permits to avoid artifacts in the epitope definition, which are consequences of differences in ability to accumulate saturation in the free state.
In detail, STDD NMR spectra at different saturation times have been acquired and, for each proton that gave rise to STD effects, the corresponding STD build up curve has been constructed. It has

been performed by fitting experimental data with the monoexponential equation: STD= STDmax(1 – exp(-ksat t)) where STD stands for the STD signal intensity of a given proton at a saturation time t, STDmax is the asymptotic maximum of the curve that corresponds to the maximal STD intensity obtainable when long saturation time are used, and ksat represents the observed saturation rate constant. Once constructed STD build up curves, it is possible to process resulting values of STDmax and Ksat in order to determine the slope of each curve at zero saturation time (Table 7.1). The values of resulting STD fit are very informative because they are affected only by the vicinity of ligand protons to the binding site of the protein and reflect the real STD effect of each proton independently of T1 bias, preventing any misinterpretation of STD enhancements. In order to obtain the epitope map and to easier compare the results, all the values have been normalized to the highest STD fit giving the final percentages of STD, namely STD epitope fit.

Table 7.1 Experimental STD intensities of **GlcNAc7** bound to LysM at different saturation times. STDmax and Ksat were calculated by fitting the data to a monoexponential equation:

$$STD(t_{sat}) = STD_{max} * (1 – exp(- k_{sat} * t_{sat})).$$

Protons of internal GlcNAc	Experimental STD%				STDmax	Ksat	STD (fit)	STD epitopes (fit)%
	1sec	2sec	3sec	4sec				
H1	30	25	40.8	57	0.2905	0.2287	0.066	32.9 %
H2	43.5	35	61.2	73	0.6243	0.1702	0.106	52.8 %
H3	51.6	47.5	73.5	80.9	0.5924	0.2258	0.133	66.3 %
H4	36.6	30	46.9	58.7	0.5259	0.1574	0.082	40.9 %
H5	48.3	40	57.1	63.5	0.4962	0.2002	0.099	49.3 %
H6a	34	27.5	40.8	46	0.3524	0.2057	0.072	35.9 %
H6b	35	27.5	51	58.7	0.5732	0.1387	0.079	39.3 %
CH3	100	100	100	100	0.6315	0.3176	0.2006	100 %

Thus, the binding epitope of the hepta-saccharide has been reported in Figure 7.10c. The most prominent STD effect belonged to the acetyl group, thus contributing primarily to the binding process; additionally, different STD enhancements were observed for all the protons of internal residues of GlcNAc.
From these results it was inferred that the acetyl group was in close contact and pivotal in the interaction with the protein and, according to the qualitative analysis, also protons H2, H3 and H5 were strongly involved into the binding with LysM domain.
These STD studies were then complemented with the analysis of trNOESY experiments both free in solution and in the presence of LysM domains to assess oligosaccharide conformational changes upon binding [21, 23-25]. Thus, trNOESY experiments were carried out at a 7:1 molar ratio with different NMR mixing times; negative NOEs were observed for both the free and bound states (data not shown), and detailed analysis of *inter*-residue cross-peaks in the trNOESY spectrum allowed the deduction of the bioactive conformation of the oligosaccharide product when bound to the enzyme

Table 7.2 Experimental and calculated interproton distances for chito-heptasaccharide

Distance	Experimental Free state	Experimental Bound state	Calculated Φ= 56,6° Ψ= 6.9°
H1-H'4	2.4	2.35	2.4
H1-H'5	3.8	3.79	3.9

Remarkably, key NOE distances obtained in both the free and bound states showed only minor changes, thus suggesting that the

bioactive conformation of chitin heptasaccharide resembled the most populated conformational state of the free ligand (Table 7.2). The NOE data were used to obtain a conformational model of the oligosaccharide by molecular mechanics (MM) and molecular dynamics (MD) simulations. A molecular snapshot of a representative conformer is depicted in Figure 7.11, with the STD-derived epitope mapping on the molecular envelope of heptasaccharide with color coding from the highest (in red) to lowest (in yellow) observed STD effect.

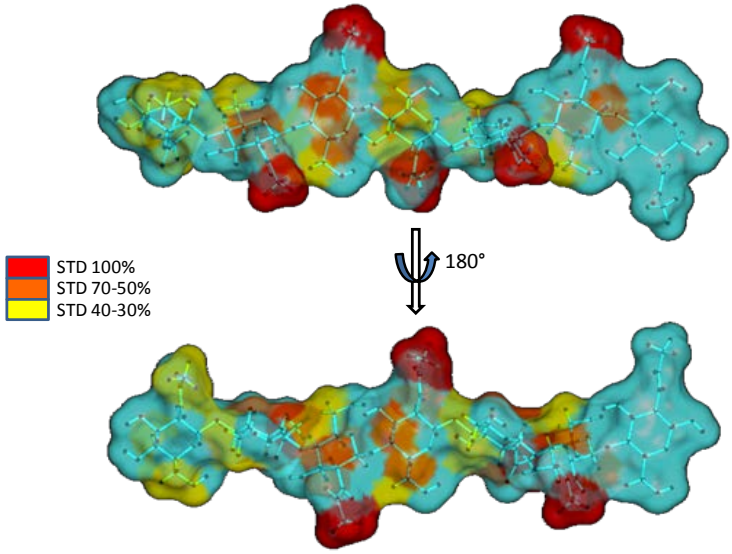


Fig. 7.11: STD-derived epitope mapping on the molecular envelope of heptasaccharide with color coding from the highest (red) to lowest (yellow) observed STD effect.

By using the same strategy, the interaction of CEBiP with the chitooctasaccharide has been evaluated and the corresponding epitope map has been obtained using a qualitative analysis because of scarce solubility of the substrate, (Figure 7.12 and 7.13). The corresponding results were in good agreement with collected data on the heptasaccharide; the acetyl group gave rise to the highest STD effect, indicating that this moiety was fundamental in the

binding process with LysM1-2 of CEBiP and once again protons H2, H3 and H5 also significantly contributed to the interaction. Furthermore, the non reducing ends showed in all cases the lowest interaction with the LysM1/2 domains.

Therefore, the binding epitope as derived by NMR analysis provided clear support that CEBiP, with its LysM domains, recognized and bound chitin oligosaccharides and showed a strong preference for longer chain substrates, especially hepta- and octa-saccharides.

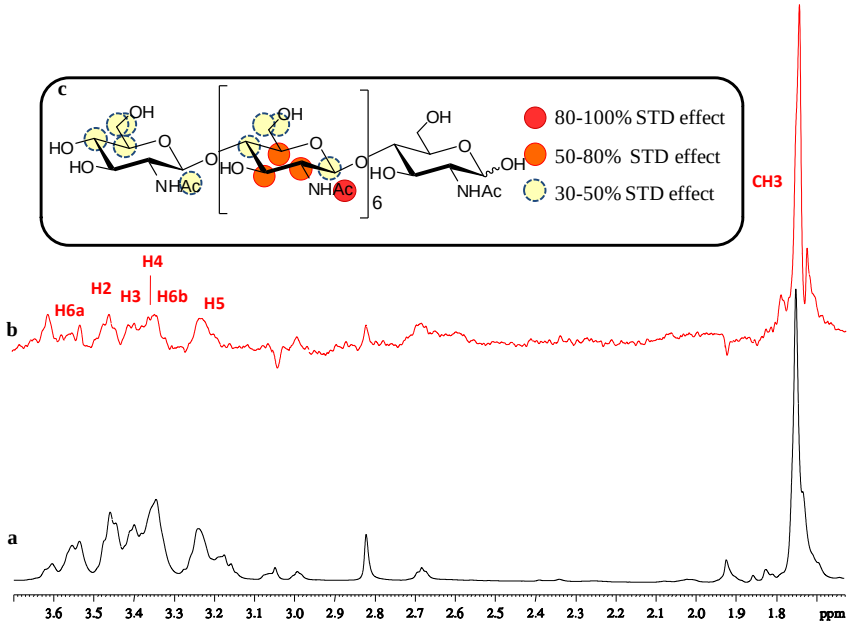


Fig. 7.12: Reference ^{1}H NMR spectrum (a) and STDD NMR spectrum (b) of mixture LysM1-2-trx : octasaccharide 1:100. c. Chemical structure and epitope binding of chitoheptasaccharide to LysM1-2-trx.

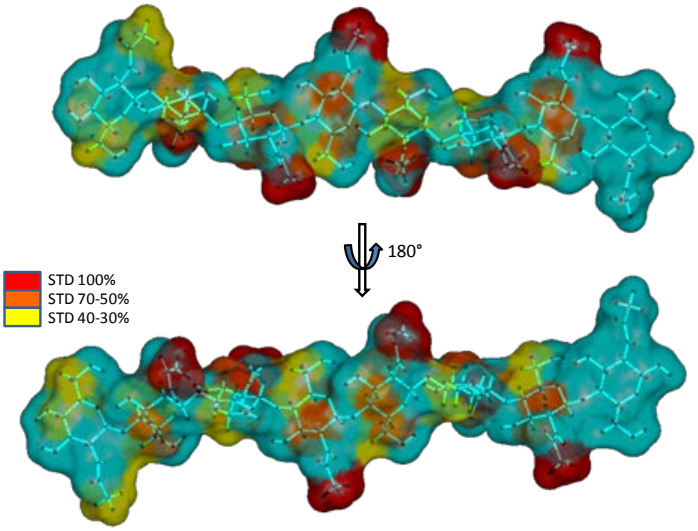


Fig. 7.13: STD-derived epitope mapping on the molecular envelope of chitooctasaccharide with color coding from the highest (red) to lowest (yellow) observed STD effect. .

7.5 Docking studies and model of binding

The data so obtained enabled the rational docking of oligosaccharides into binding site of CEBiP. To identify the CEBiP binding pocket, the group of prof. Berisio, CNR of Naples, determined the homology model structure of the protein. The entire extracellular region of CEBiP was built using the program MODELLER and the structure of Chitin elicitor receptor kinase 1 (CERK1) from *Arabidopsis thaliana* as a template.

The combination of STD-NMR epitope mapping data with knowledge of the bound conformation of ligands, which may be obtained by trNOESY experiments, is a powerful method to build up models of protein-ligand interaction.

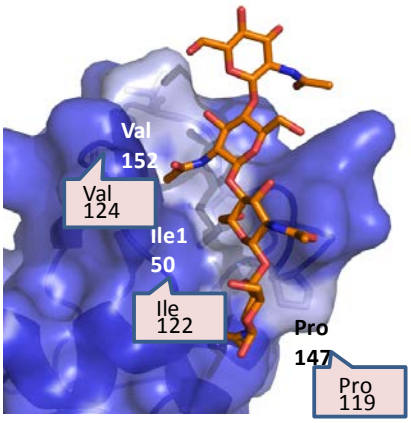


Fig. 7.14: Amino acid residues involved in the interaction between CEBiP lysM1 and chito-oligosaccharides.

As shown in Figure 7.14, four GlcNAc moieties can be accommodated in the binding cleft of LysM1 domain. Matching modeling studies with STD NMR data, showing that reducing ends of oligosaccharides barely contribute to the binding, it can be suggested that the smallest oligosaccharide which saturates the four binding subsites on LysM1 is $(GlcNAc)_6$.

In the interaction model of the hexa-saccharide, main interactions involve the three sugars from GlcNAc-2 to GlcNAc-4; weaker interactions characterize GlcNAc-5 whereas GlcNAc-1 and GlcNAc-6 do not interact with LysM1. Consistent with STD NMR results, the acetyl moieties of $(GlcNAc)_6$ provide a crucial contribution to binding. Both GlcNAc-2 and GlcNAc-4 sugar moieties are involved in hydrophobic interactions (with Leu24 and Val18, respectively) and in hydrogen bonds with backbone nitrogen atoms (of Leu 24 and Gln21, respectively) (Figure 7.14). Apart from hydrogen bonds mediated by the acetyl moieties, $(GlcNAc)_6$ binding is mainly stabilized by hydrophobic interactions involving sugar ring carbons, a finding which well agrees with our STD NMR results. The described interaction model also explains

the low STD effects observed upon binding of tri- and tetra-saccharides, compared to hexasaccharide or larger substrates.
Furthermore, since STD-NMR epitope mapping of $(GlcNAc)_8$ showed that the internal six GlcNAc residues strongly interacted with the LysM1-2 polypeptide, we hypothesized that two LysM1-2 molecules simultaneously interacted with a single oligosaccharide.

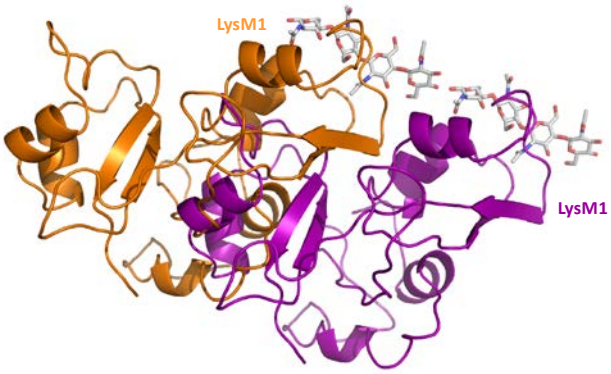


Fig. 7.15: Adjacent binding: side-by-side modeling is not compatible with CEBiP structure model.

It has been confirmed, by light scattering experiments, that the protein dimerised in the presence of an excess of $(GlcNAc)_8$ (data not shown). Modeling studies based on such assumption rejected a model in which two CEBiP molecules bind through interactions with acetyl moieties $(GlcNAc)_8$ from the same side of the oligosaccharide because of the steric hindrance between these CEBiP molecules (Figure 7.15). On the other hand, a model in which two CEBiP molecules bind from the opposite sides of $(GlcNAc)_8$ was acceptable without such a steric problem (Figure 7.16).

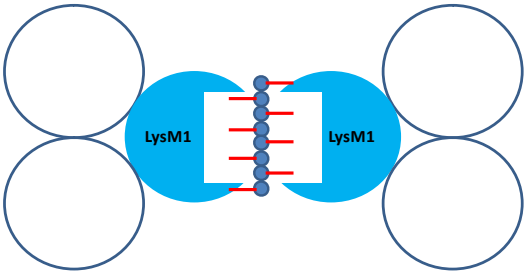


Fig.7.16: CEBiP dimerization induced by (GlcNAc)8

7.6 Binding of LysM12/TRX I150A to chitoheptasaccharide

Modeling of binding of $(NAG)_n$ to the LysM1 domain of CEBiP identified possible residues involved in the interaction (Figure 7.14). The predicted binding site is lined by hydrophobic residues like Pro147, Ile150, Val152. Among these, Ile150 is located in the central part of the putative binding cleft, a finding which suggested us that this residue may be critical for oligosaccharide binding.

To corroborate this hypothesis, Ile 150 was mutated to Ala and the mutated protein, LysM12/TRX I150A, was analyzed in its interaction with chitooligosaccarides performing STDD NMR experiments under the same conditions used for wild type LysM1-2.

Prompted by previously NMR observations on the affinity of LysM1-2 domains of CEBiP for chitin oligosaccharide, heptasaccharide was used as ligand. NMR results unequivocally showed that the mutant protein LysM12/TRX I150A was not able to bind chitooligosaccharide at all. Indeed, no STD signals were observed (Figure 7.17). The inability of mutant to interact with heptasaccharide definitively identify I150 as a key aminoacid in the recognition process of fungal chitooligosaccharides by CEBiP LysM1-2 repeats.

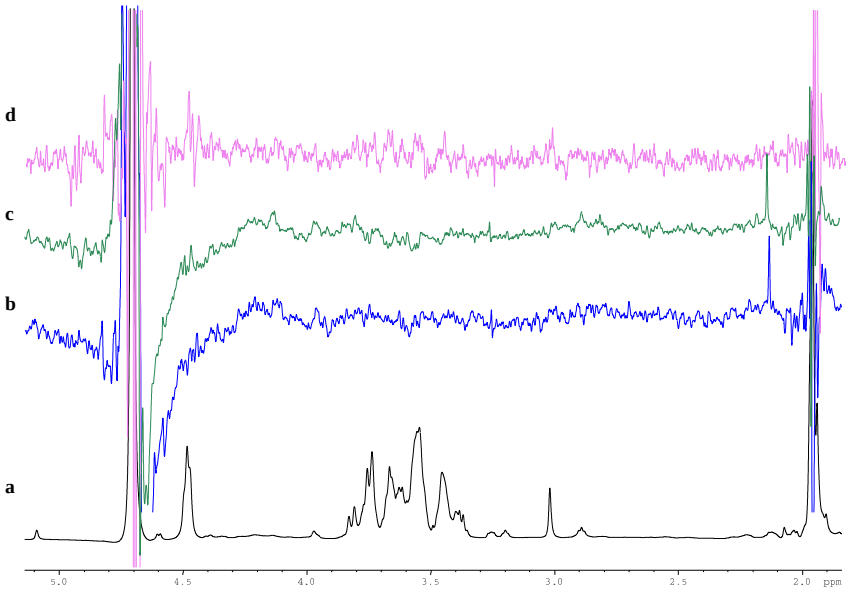


Fig. 7.17: Reference ^{1}H NMR spectrum (a) and STDD NMR spectra of mixture LysM12/TRX I150A : heptasaccharide at saturation time of 2sec (b), 3sec (c), 4s (d).

These results were confirmed by affinity labeling analysis performed on mutant CEBiP proteins containing point mutation for five amino acid residues, L_{121}, Y_{135}, P_{147}, I_{150} and V_{152}, which were suggested to have close contact with chito-oligosaccharides in the docking simulation (data not shown).

7.7 Conclusion

A combinatorial approach based on advanced STD and NOE-based NMR techniques and molecular dynamic allowed us to get new insight into the binding mechanism between the LysM domain in the ectodomain of CEBiP and chitin oligosaccharides.
Biochemical studies, molecular modeling and docking analysis all indicated that LysM1 in the ectodomain of CEBiP plays a major role for the perception of chitin oligosaccharides. This conclusion matched the observation on the chitin binding site in the

ectodomain of Arabidopsis CERK1 receptor kinase, where the chitin oligosaccharide also bound to the central LysM domain [26]

Prompted by the above biological data, we have provided a molecular description of binding of the central LysM domain (LysM1) of CEBiP with chitin by revealing the conformations and epitope patterns of a plethora of chitin oligosaccharides bound to LysM1-2 domains.

STD NMR experimental data, gathered on chitin fragments, clearly showed that short length oligosaccharides weakly bound to LysM1 while a strong binding was detectable for chito-hepta and octasaccharides. Both qualitative and quantitative analysis, based on the construction of STD build up curves, allowed to precisely map ligand epitopes in contact with the protein. These studies demonstrated that acetyl groups of chitin oligosaccharides are crucial moieties for binding. Likely, acetyl groups are key molecular signatures which allow LysM1 to distinguish chitin from glucose containing glycoconjugates. Furthermore, the inner units of the chitin fragments exhibited stronger interaction with LysM motifs whereas both the outer residues seemed to be less involved in the interaction, likely because more exposed to the solvent.

On the other hand, modeling study indicated that several amino acid residues are in close contact with the bound chitin oligosaccharides. Site directed mutagenesis of these amino acids revealed that one of these residues, I_{150}, which is located in the central region of LysM1 binding cleft, plays an important role in ligand recognition.

We used all of this information to produce a three-dimensional model of LysM1 interaction with chitin oligosaccharides through modeling studies. These demonstrated that CEBiP LysM1 domain accommodate chitin oligosaccharides in a hydrophobic cleft which hosts four GlcNAc moieties. Of these, two GlcNAc moieties mainly interact through hydrogen bonding and hydrophobic interactions mediated by their acetyl groups. Since STD NMR data

showed that outer ends of oligosaccharides barely contributed to binding, $(GlcNAc)_6$ was the shortest oligosaccharide to saturate the four sugar binding subsites in LysM1 cleft. This finding well explained the significantly stronger binding of hexasaccharide or larger substrates.

Furthermore, we demonstrated that long substrates induced a dimerization of CEBiP, explaining the strong affinity of that receptor for $GlcNAc_7$ and $GlcNAc_8$. Indeed, such a dimerization required at least five internal GlcNAc moieties and can, therefore, be induced by sugars larger than $(GlcNAc)_{7,}$ since smaller oligosaccharides did not provide enough number of internal GlcNAc moieties required for crosslinking.

Ligand-induced dimerization of receptor is a critical event in various receptor systems [27-29], especially as a trigger of autophosphorylation of receptor kinase as classically reported for receptor tyrosine kinases. Thus, it could be hypothesized that such dimerization plays a key role in the activation of CERK1 to initiate a signaling cascade upon chitooligosaccharides binding.

References

[1] Nurnberger, T.; Brunner, F.; Kemmerling, B.; Piater, L. *Immunological Reviews*, **2004**. 198, 249–266.
[2] Gust, A.A.; Willmann, R.; Desaki, Y.; Grabherr, H. M.; Nürnberger, T. *Trends Plant Sci*, **2012.** 17(8), 495–502.
[3] Willmann et al. *PNAS*, **2011**. 108(49), 19824-9.
[4] Kaku, H.; Nishizawa, J.; Ishii-Minami, N.; Akimoto-Tomiyama, C.; Dohmae, N.; Takio, K.; Minami, E. and Shibuya, N. *PNAS*, **2006**. 103, 11086-11091. doi:10.1073/pnas.0508882103.
[5] Miya, A. et al *PNAS*, **2007**. 104, 19613–19618.
[6] Bart, P.H.J.; Thommaa, B.L.; Nürnberger T.; Matthieu H.A.J.; Joostena, B. *The Plant Cell*, **2011**. 23(1), 4-15.
[7] Spaink, H.P. *TRENDS in Microbiology*, **2004**. 12(5), 201-203.
[8] Ohnuma, T. et al. *J. Biol. Chem*., **2008**. 283, 5178–5187.
[9] Gough, C. and Cullimore, J. *Mol. Plant Microbe Interact*., **2011**. 24, 867–878.
[10] Shimizu, T.; Nakano, T.; Takamizawa, D.; Desaki, Y.; Ishii-Minami, N.; Nishizawa, Y. et al. *Plant J*. **201**0. 64: 204–214.
[11] Shinya, T.; Motoyama, N.; Ikeda, A.; Wada, M.; Kamiya, K.; Hayafune, M.; Kaku, H.; Shibuya, N. Plant Cell Physiol., 2012. 53(10), 1696–1706.
[12] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed. Engl.* **2003**. 42, 864.
[13] Squeglia, F.; Marchetti, R.; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *JACS*., **2011**. 133(51), 20676–20679.
[14] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*., **2012**. 10, 1387-98.
[15] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A. Chem. Eur. J., **2012**. 18(42),13395-404.
[16] Claasen, B.; Axmann, M.; Meinecke, R.; Meyer, B. *J Am Chem Soc*., **2005**. 127(3), 916-9.
[17] Haselhorst, T.; Münster-Kühnel, A.K.; Oschlies, M.; Tiralongo, J.; Gerardy-Schahn, R.; von Itzstein, M. *Biochem Biophys Res Commun*., **2007**. 359(4), 866-70.

[18] Assadi-Porter, F.M.; Tonelli, M.; Maillet, E.; Hallenga, K.; Benard, O.; Max, M.; Markley, J.L. *J Am Chem Soc.*, **2008**. 130(23), 7212-3.
[19] Szcepina, G.M.;. Bleile, W.D.; Pinto, B.M. *Chem. Eur. J.* **2011**. 17, 11446-11455.
[20] Angulo, J.; Díaz, I.; Reina, J.J.; Tabarani, G.; Fieschi, F.; Rojo, J.; Nieto, P.M. *ChemBioChem*, **2008**. 9, 2225-2227.
[21] Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chemistry*, **2011**. 17(5), 1547-60.
[22] Mayer, M.; James, L.T. *J. Am. Chem. Soc*,. **2004**. 126, 4453-4460.
[23] The Nuclear Overhauser Effect in Structural and Conformational Analysis (Eds.: D. Neuhaus, M. P. Williamson), Wiley-VCH, New York, **2000**.
[24] NMR Spectroscopy of Glycoconjugates (Eds.: J. Jimenez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**.
[25] Johnson, M.A.; Pinto, B.M. *Carbohydr Res*. **2004**. 339(5), 907-28.
[26] Liu, Y. et al. *Science* **2012**. 338, 1327-1330.
[27] Greenlund A.C.; Farrar, M.A.; Viviano, B.L. and Schreiber, R.D. *EMBO J.*, **1994**. 13(7), 1591–1600.
[28] Graus-Porta1, D.; Beerli1, R.R.; Daly1, J.M.; and Nancy, he, E.H. *EMBO Journal*, **1997**. 16, 1647-1655 doi:10.1093/emboj/16.7.1647.
[29] Plotnikov1, A.N.; Schlessinger1, J.; Hubbard1, S.R.; Mohammadi1, M. *Cell*, **1999**. 98(5), 641–650.

Chapter 8

*Cj*Xyl31A – xyloglucan interaction

8.1 Introduction

The enzymatic degradation of cellulose and matrix glycans of the plant cell wall into simple sugars provides energy for organisms from all kingdoms of life, from microbes to higher animals [1]. Understanding of the mechanisms of microbial plant cell wall degradation is of great interest since it is related to the possibility of improving methods of biomass conversion into fuels and feedstocks. The study of the interaction of glycoside hydrolases with their substrates is fundamental not only to the biomass resource utilization, but also to diverse applications in medicine, food and feed production.

Recent molecular modeling of the α-xylosidase *Cj*Xyl31A from the soil saprophyte *Cellvibrio japonicus*, in harness with protein crystallography and enzyme kinetic analysis, has suggested that an appended PA14 protein domain, unique among glycoside hydrolase family 31 members, may confer specificity for large oligosaccharide fragments of the ubiquitous plant polysaccharide xyloglucan [1]. In this work [2], we have used advanced NMR spectroscopy methods to gain further insight into the recognition and degradation of large xyloglucan oligosaccharides by *Cj*Xyl31A.

8.1.1 Glycoside hydrolase *Cj*Xyl31A

Cellvibrio japonicus is a Gram-negative, soil bacterium, which is able to utilize diverse plant polysaccharides as nutrient source, thanks to glycoside hydrolases (GHs) encoded by its genome [3]. These enzymes are essential for carbohydrate metabolism and are

thus ubiquitous among species in the tree of life. GHs are central to manifold processes in Nature, including anti-bacterial defense strategies, pathogenic mechanisms, homeostatic cellular functions, and, not least, breakdown of plant biomass in the global carbon cycle [4,5]. In recent years, GH enzymology has come into special focus in the context of renewable biofuel production from plant cell walls via saccharification and fermentation [6-9]. A growing appreciation of the importance of xyloglucans (cfr 1.1.2; Figure 8.1) in plant cell wall architecture is likewise fueling an interest in the specific enzymes responsible for their modification and/or degradation [9-11].

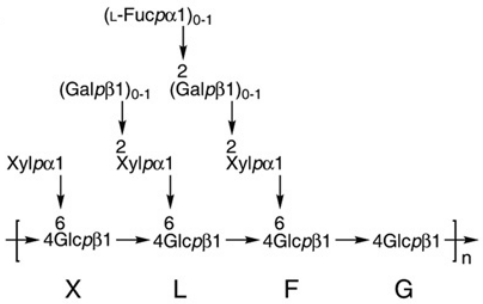


Fig. 8.1: General structure of primary cell wall xyloglucans showing variable branching. Xyloglucans are composed by $\beta(1\rightarrow4)$-D-glucan backbone substituted with $\alpha(1\rightarrow6)$-D-xylopyranosyl moieties at regular intervals. The basic repeating motifs may be further decorated with galactopyranose, arabinofuranose, fucopyranose and *O*-acetyl moieties in a tissue and species-dependent manner. (Adapted from [1]).

Several enzymatic activities are required to degrade xyloglucan into its constituent monosaccharides. While the endogenous enzymes involved in the xiloglucans (XGs) mobilization in plants are so far known [12,13], those responsible for XG degradation by micro-organisms, such as *C. japonicus*, have been studied more recently [1,14]. In detail, the structural and enzymatic characterization of the unique α-xylosidase, *Cj*Xyl31A – the only

member of *C. japonicus* glycoside hydrolase family 31 (GH31, see www.cazy.org and ref. [15]) – has been performed (Figure 8.2) [1].

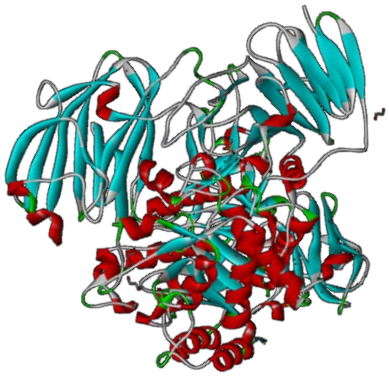


Fig. 8.2: Global structures of CjXyl31A.

This *exo*-acting enzyme is responsible for the cleavage of α-(1→6)-linked xylose residue, from the non-reducing terminus of xyloglucan oligosaccharides, which result from extracellular endo-xyloglucanase activity (Figure 8.3) [1].

Kinetic analyses have demonstrated that *Cj*Xyl31A shows a strong substrate preference for the heptasaccharide XXXG over shorter congeners (cfr Figure 1.6 for oligosaccharide nomenclature).

This preference is likely to arise from the presence of an appended PA14 domain found only in a limited subset of GH31 members, including *Cj*Xyl31A. Indeed, a preliminary substrate-docking model has highlighted that this PA14 domain endows the enzyme with an extended substrate binding platform, which may create productive interactions with larger xyloglucan oligosaccharides based on the XXXG motif [1]. Unfortunately, extensive experimentation failed to produce a requisite *Cj*Xyl31A:oligosaccharide complex *in crystallo* to provide firm experimental support for this model.

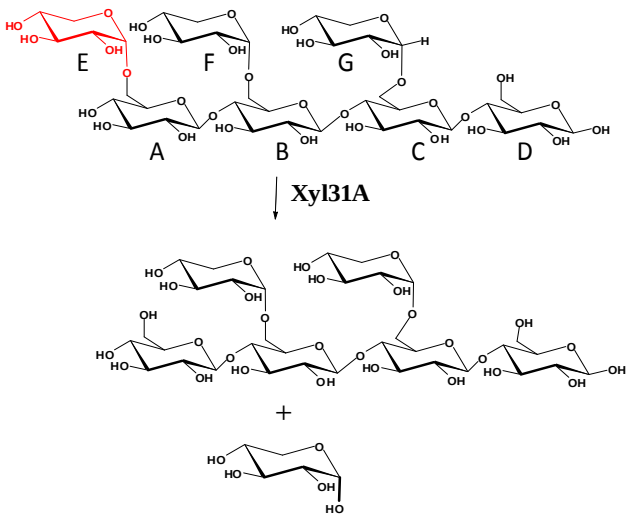


Fig. 8.3: Action of *Cellvibrio japonicus* α-xylosidase 31A (*Cj*Xyl31A) on xyloglucan heptasaccharide XXXG.

In the present study, a combination of NMR techniques, including Saturation Transfer Difference and transfer Nuclear Overhauser effect spectroscopy, was used to reveal extensive interactions between *Cj*Xyl31A active-site variants and xyloglucan hexa- and heptasaccharides [2]. This analysis provided an important rationalization of previous kinetic data on *Cj*Xyl31A and unique new insight into the role of the PA14 domain, which was not otherwise obtainable by protein crystallography.

8.2 The stereochemistry of hydrolysis of the XXXG heptasaccharide by *Cj*Xyl31A

The majority of GH31 members utilizes the canonical double-displacement mechanism of glycoside hydrolysis [16,17]. This mechanism proceeds via a covalent glycosyl-enzyme intermediate and results in the overall retention of the substrate anomeric stereochemistry in the product. Consistent with its membership in GH31, kinetic analysis of product formation by NMR indicated that

*Cj*Xyl31A is also a retaining glycosidase. Incubation of *Cj*Xyl31A with the heptasaccharide XXXG rapidly generated α-D-xylopyranose and the hexasaccharide GXXG, due to cleavage of the terminal non-reducing-end α-(1→6)Xyl*p* moiety. Immediately following addition of the enzyme, the α-anomeric proton signal of xylopyranose predominated, then decreased due to mutarotation to a ca. 35:65 ratio of α- and β-anomers at equilibrium [18] (Figure 8.4).

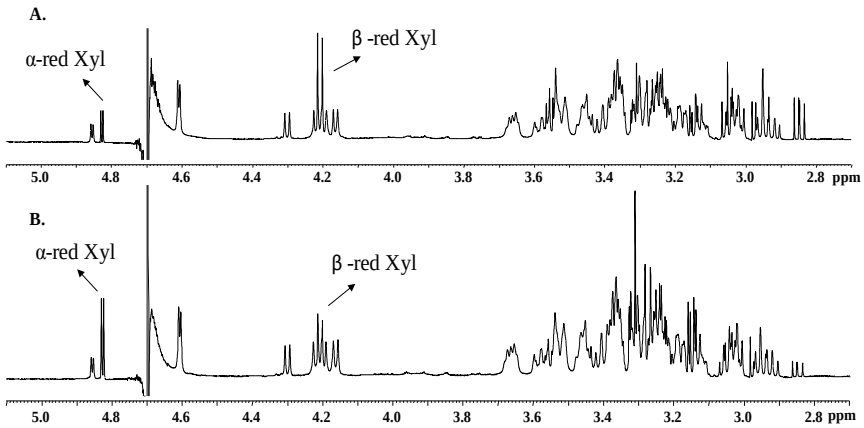


Fig. 8.4: Time-course ^{1}H NMR analysis of the hydrolysis of XXXG by *Cj*Xyl31A. **A**. Spectrum immediately following addition of the enzyme to the substrate (1:45 final ratio), which slows the evolution of the signal corresponding to H-1 of α-D-xylopyranoside and disappearance of the anomeric signal of the terminal α-D-Xylp-(1→6) of XXXG (residue E in fig. 8.3). **B**. Spectrum after 12 h, at which time a ca. 35:65 mixture of α- and β-anomers is observed, due to equilibration by mutarotation.

8.3 Epitope mapping of xyloglucan bound to *Cj*Xyl31A

Saturation Transfer Difference NMR spectroscopy [19,20] was used to map interacting epitopes on both the natural substrate and product of the enzyme, xyloglucan heptasaccharide XXXG and hexasaccharide GXXG, respectively. In parallel, Transferred NOE NMR spectroscopy, together with molecular mechanics (MM) and molecular dynamics (MD) simulations, was employed to reveal the

bound versus solution conformations of these oligosaccharides [21-23].

8.3.1 The interaction of *Cj*Xyl31A with its product, GXXG hexasaccharide. STD NMR and tr-NOESy analysis

CjXyl31A demonstrates an exquisite regioselectivity for the terminal non-reducing-end α-(1→6)-xylose of xyloglucan oliogosaccharides. Even after extended incubation, the complex heptasaccharide and tetradecasaccharide substrates XXXG and XXXGXXXG yield only GXXG and GXXGXXXG, respectively [1]. As a first step toward understanding the interaction of *Cj*Xyl31A with large oligosaccharides, we studied the product complex of the enzyme with GXXG, in the absence of the by-product xylose, by a combined NMR spectroscopy and molecular modeling approach.

The one-dimensional ^{1}H-NMR spectrum of GXXG is shown in Figure 8.5.

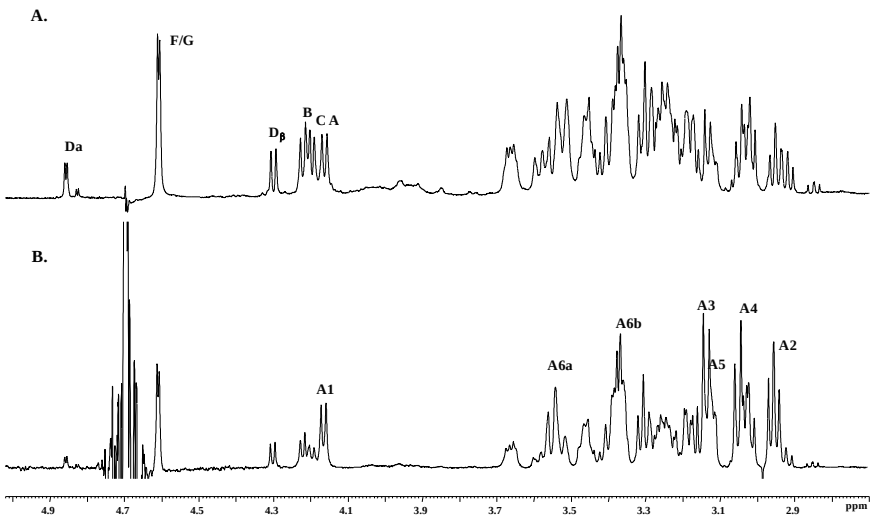


Fig. 8.5: One-dimensional proton NMR spectral analysis of the binding of GXXG to wild-type *Cj*Xyl31A. **A.** ^{1}H NMR spectrum of GXXG in presence of *Cj*Xyl31A. Anomeric signals from each of the monosaccharide residues A-F are indicated. **B.** STD spectrum. An on-resonance irradiation frequency at 0 ppm and

a saturation time of 2 seconds was used, with a 1:45 enzyme-ligand ratio. STD intensities are given in Table 8.2; the most prominent STD intensities belonging to individual protons on residue A have been indicated.

A combination of homo- and hetero-nuclear 2-D NMR experiments, including DQF-COSY, TOCSY, T-ROESY, ^{1}H-^{13}C HSQC, and ^{1}H-^{13}C HMBC experiments, was executed in order to assign all the spin systems of the carbohydrate (Table 8.1).

Table 8.1 ^{1}H and ^{13}C NMR chemical shifts (ppm) of the GXXG hexasaccharide.

Residue	nucleus	1	2	3	4	5	6	5,6 coupling
A: t-β-Glc	^{1}H	4.16	2.95	3.14	3.05	3.12	R: 3.55/ S: 3.37	$^{3}J_{5,6R}$< 1 Hz $^{3}J_{5,6S}$=2.5 Hz
	^{13}C	102.3	72.6	74.9	75.3	75.6	60.0	
B: 4,6-β-Glc	^{1}H	4.21	3.02	3.30	3.39	3.46	R: 3.66/ S: 3.52	$^{3}J_{5,6R}$=5.2 Hz $^{3}J_{5,6S}$=1.8 Hz
	^{13}C	102.2	72.3	73.6	78.3	72.9	65.5	
C: 4,6-β-Glc	^{1}H	4.19	3.02	3.30	3.40	3.46	R: 3.66/ S: 3.52	$^{3}J_{5,6R}$=5.2 Hz $^{3}J_{5,6S}$=1.8 Hz
	^{13}C	102.2	72.3	73.6	78.6	72.9	65.5	
D: 4-β-Glc	^{1}H	4.30	2.91	3.26	3.29	3.24	R: 3.59/S: 3.45	$^{3}J_{5,6R}$=10.9Hz $^{3}J_{5,6S}$=ND
	^{13}C	95.2	73.4	73.9	78.6	74.1	59.5	
F: t-α-Xyl	^{1}H	4.61	3.18	3.36	3.25	3.36/3.21		
	^{13}C	98.5	71.1	72.5	68.9	61.1		
G: t-α-Xyl	^{1}H	4.61	3.18	3.36	3.25	3.36/3.21		
	^{13}C	98.5	71.1	72.5	68.9	61.1		

Six signals were identified in the anomeric region of the ^{1}H-NMR spectrum, which corresponded to those of both anomers of the reducing-end Glc*p* residue, the three internal β-(1→4)Glc*p* residues, and the co-incident signals of the pendant α-(1→6)Xyl*p* residues (Figure 8.5). $^{3}J_{HH}$-Coupling constants, obtained from ^{1}H-NMR and DQF-COSY spectra, and *intra*-residue NOE contacts indicated that all sugars were present exclusively in the $^{4}C_{1}$ chair conformation. Furthermore, to obtain information on the

conformation of the C-6 hydroxymethyl groups of Glc residues, three bond coupling constants of the ligand in the bound state (at 1:7 and 1:14 molar ratios of *Cj*Xyl31A to GXXG) were also determined, (Table 8.1) [24-26]. The rotamer distribution in the bound state indicated a 50:50 population of *gg*:*gt* conformers of the two central glucoses (residues **B** and **C**), a *gg* rotamer preference of non-reducing-end Glc (residue **A**), and *gt* rotamer preference for the reducing-end Glc (residue **D**) (Table 8.1).

To further examine the interaction of the oligosaccharide product with the enzyme, the binding of GXXG to *Cj*Xyl31A was investigated using STD-NMR spectroscopy. STD-NMR spectroscopy allows mapping of ligand epitopes in close contact with proteins by magnetization transfer, and has previously been proven to be very informative in the examination of protein-carbohydrate interactions [19-23]. Qualitative comparison of the 1D ^{1}H NMR spectrum of GXXG and the STD spectrum using a saturation time of 2 s revealed changes in the relative areas of individual signals (Figure 8.5). The most prominent STD signal belonged to the non-reducing-end Glc (residue **A**), which occupies the +1 enzyme subsite (nomenclature according to Davies et al. [27]) in the catalytic cycle and is thus expected to form one of the most intimate interactions in the enzyme-product complex. Additionally, the data also clearly indicated that the whole saccharide moiety was involved in the interaction with the protein. However, quantification of the STD signals in the 1D spectrum, to estimate the contribution of the binding of individual monosaccharide units, was hampered by significant signal overlap.

To circumvent this problem, we then used in-depth 2-D experiments, including STD-HSQC (Figure 8.6) and STD-TOCSY (Figure 8.7) to resolve STD signals.

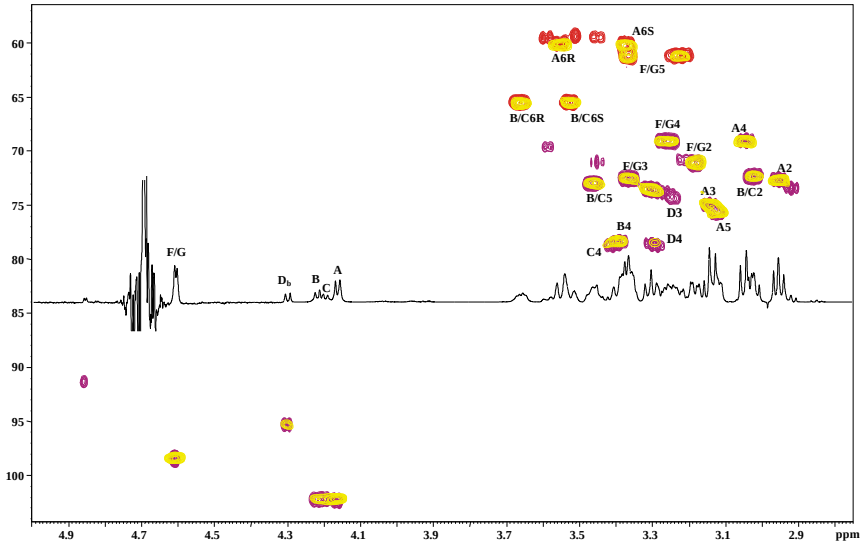


Fig. 8.6: Two-dimensional proton NMR spectral analysis of the binding of GXXG to wild-type *Cj*Xyl31A. HSQC (red and violet) and STD-HSQC (yellow) spectra of GXXG in presence of *Cj*Xyl31A (enzyme-ligand ratio 1:45).

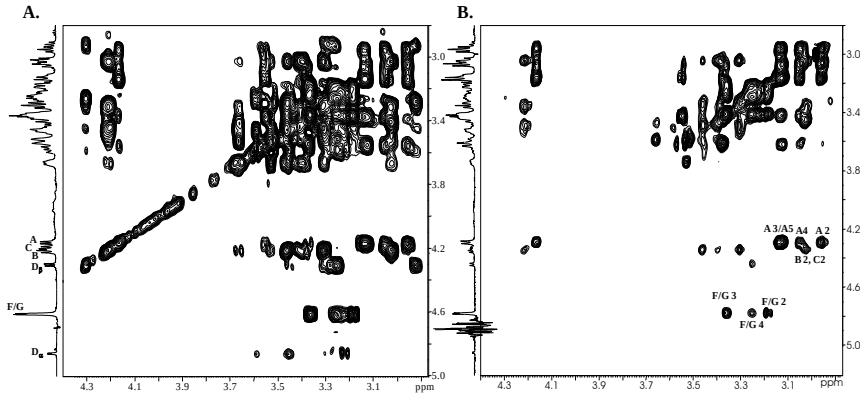


Fig. 8.7: TOCSY (A) and STD-TOCSY (B) spectra of GXXG in presence of Xyl31A at a 1:45 ratio of enzyme to ligand.

Table 8.2 Relative observed STD effects for GXXG bound to wild-type *Cj*Xyl31A.

Residue (cf. Fig. 1)	Relative STD signal [(a)]					
	H-1	H-2	H-3	H-4	H-5	H-6
A: t-β-Glc	55 %	93.4 %	76 %	100 %	100 %	R: 53 %
	25 %	28.5 %	24 %	25.8 %	17.6 %	R: 19.8% S: 13.2 %
B: 4,6-β-Glc						
C: 4,6-β-Glc	18 %	10.8 %	9.8 %	16 %	11.6 %	R: 6.6% S: 5.5 %
D: 4-β-Glc	9.8 %	12 %	-	-	-	10.8 %
	33 %	32 %	-	23 %	ax.: 27% eq.: 28.5%	
F/G:[(b)] t-α-Xyl						

[(a)] Relative observed STD effects were calculated according to the equation: $(I_o - I_{sat})/I_o$ and normalized to the largest STD value set at 100%.
[(b)] NMR resonances of residues F and G co-incident and thus the corresponding STD values are indistinguishable.

STD-TOCSY experiments, in particular, were used to qualitatively and quantitatively estimate the contribution for each proton to the STD spectrum as follows: The row corresponding to an isolated signal containing its spin coupled partners was extracted from both STD-TOCSY (Figure 8.7B) and TOCSY (Figure 8.7A); the relative contribution of a spin coupled proton could then be ascertained by measuring the enhancements relative to a reference signal for which the STD effect is known from 1D spectra [28]. Relative STD effects, indicative of the saturation transferred from *Cj*Xyl31A to GXXG, were calculated from the STD amplification factors (Table 8.2). Due to the co-incidence of the resonances of the pendant xylose residues (residues **C** and **G**), we could not distinguish their individual contributions to the observed STD signal.
The epitope map derived from the 2-D STD NMR data, shown in Figure 8.8, provided clear support that the entire GXXG molecule interacted with *Cj*Xyl31A, with the non-reducing-end Glc (residue **A**) exhibiting the strongest interaction with the protein (as was suggested by the 1-D STD NMR spectrum). These observations are consistent with the specific interaction of the oligosaccharide with the active-site, in which the non-reducing-end Glc would be expected to be intimately associated with the +1 subsite prior to

release of the product. Notably, whereas the majority of *exo*-acting glycosidases have comparatively shallow active-site pockets comprised of one negative and one or possibly two positive subsites [5], the STD data indicate that *Cj*Xyl31A is likely to possess at least four primary positive subsites (accommodating the Glc residues of GXXG) and one or two auxiliary, side-chain binding positive subsites (accommodating one or both of the pendant Xyl residues).

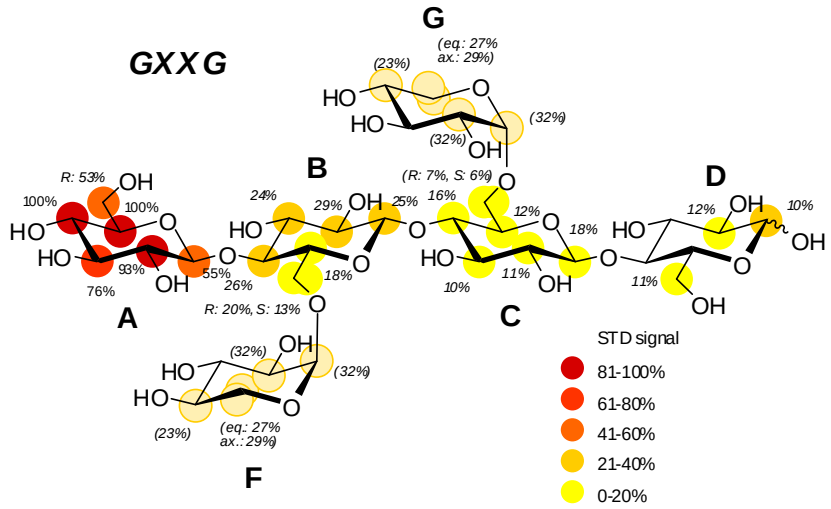


Fig. 8.8: Relative observed STD effects for GXXG bound to *Cj*Xyl31A. STD effects were calculated according to the equation: $(I_o\text{-}I_{sat})/I_o$ and normalized to the largest STD value set at 100%. Relative STD intensities are coded with solid colors according to the scale shown. Ambiguous values, due to signal overlap, are represented in parenthesis and shown with faded color-coding.

These STD studies of the enzyme-product interaction were then complimented with the analysis of trNOESY experiments on GXXG both free in solution and in the presence of *Cj*Xyl31A to assess oligosaccharide conformational changes upon binding [29-32]. Thus, trNOESY experiments were carried out at a 7:1 product-to-enzyme molar ratio with different NMR mixing times. Negative NOEs were observed for both the free and bound states (data not

shown), and detailed analysis of *inter*-residue cross-peaks in the trNOESY spectrum allowed the deduction of the bioactive conformation of the oligosaccharide product when bound to the enzyme, including information on the conformational selection of glycosidic linkages.[24-29]. Remarkably, key NOE distances obtained in both the free and bound states showed only minor changes (Table 8.3), thus suggesting that the bioactive conformation of GXXG resembled the most populated conformational state of the free ligand.

Table 8.3 Experimental (from ROESY experiments) and calculated (from MD and MM calculations) distances for the hexasaccharide GXXG.

	Free	Bound	MD(gg conformer)	MD(gt conformer)
G1/C6$_S$	2.5	2.4	2.36±0.16	2.36±0.17
G1/C6$_R$	2.99	3.0	3.03±0.28	3.05±0.28
F1-B6$_S$	2.5	2.4	2.35±0.17	2.39±0.17
F1-B6$_R$	2.99	3.0	3.11±0.27	3.17±0.25
F1-B5	4.3	4.27	4.34±0.36	4.38±0.33
G1-C5	4.3	4.27	4.37±0.36	4.01±0.39
C1-D4	2.49	2.4	2.62±0.29	2.62±0.25
B1-C4	2.53	2.57	2.56±0.24	2.56±0.27
A1-B4	2.60	2.54	2.52±0.26	2.53±0.23

The NOE data were used to obtain a conformational model of the GXXG oligosaccharide by molecular dynamics (MD) and molecular mechanics (MM) simulations. Since *J* coupling constants evidenced the existence both *gg* and *gt* rotamers of the internal Glc residues **B** and **C** (Table 8.1), two oligosaccharides, which differed in the conformations of the C_5-C_6 bonds of both glucose residues of the [α-D-Xyl-(1→6)]-β-D-Glc-(1→4) moieties, were built. Parallel simulations of these two starting structures were performed under identical conditions (data not shown). Ensemble-average inter-

proton distances for each molecule were extracted from MD simulations and translated into NOE contacts using a full-matrix relaxation approach. The corresponding average distances obtained for the simulation from $\langle r^{-6} \rangle$ values were compared to those derived experimentally, and satisfactory agreement was observed between the calculated and the experimental values (Table 8.3). Molecular snapshots of the most representative conformers are depicted in Figure 8.9. The oligosaccharide exhibited a rigid glucan backbone and notable flexibility of the two α-(1→6) xylosidic linkages, which oscillated around the two possible *gg* and *gt* rotamers; based on the NOE data, similar dynamics are suggested in the enzyme-bound state.

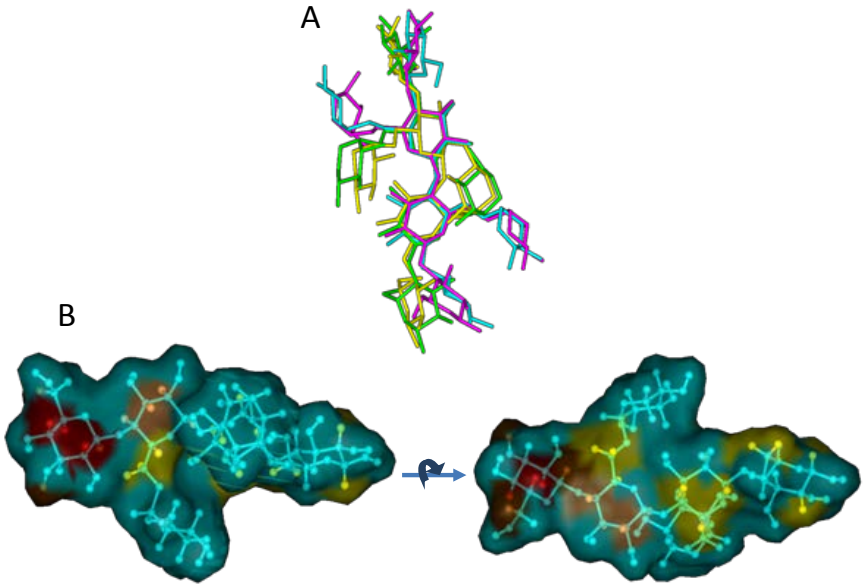


Fig. 8.9: Structures of the oligosaccharide GXXG in the bound conformation with *Cj*Xyl31A. **A.** Dominant conformers of GXXG derived from NMR data and MD/MM simulations, **B**. STD-derived epitope mapping on the molecular envelope of GXXG, with color coding from the highest (red) to lowest (yellow) observed STD effect (cfr. Table 8.2, Figure 8.8).

8.3.2 The interaction of CjXyl31A-D659A with the native substrate, XXXG heptasaccharide. STD NMR and tr-NOESY analysis

As described above, wild-type *Cj*Xyl31A rapidly hydrolyses its natural substrate XXXG under NMR experimental conditions. To allow the analysis of the complete heptasaccharide with the protein, a catalytically inactive variant of *Cj*Xyl31A was produced, by the group of professor Brumer, by mutating the predicted catalytic/acid base residue, D659 [1,33], from aspartic acid (sidechain carboxylic acid) to alanine (sidechain methyl). Subsequently, the same combined 1D and 2D NMR approach employed for the wild-type enzyme and hexasaccharide product, GXXG, was executed to reveal the overall oligosaccharide conformation and the interaction of individual saccharide residues of XXXG with *Cj*Xyl31A-D659A.

Table 8.4 ^{1}H and ^{13}C NMR chemical shifts (ppm) of the XXXG heptasaccharide.

Residue (cf. Fig. 1)	**nucleus**	**1**	**2**	**3**	**4**	**5**	**6**	**5,6 coupling**
A: 6-β-Glc	^{1}H	4.20	2.98	3.15	3.34	3.17	3.59/3.40	$^{3}J_{5,6R}$<1Hz $^{3}J_{H6S,H5}$=2.3 Hz
	^{13}C	102.4	72.4	75.0	73.8	68.9	65.2	
B: 4,6-β-Glc	^{1}H	4.23	3.02	3.31	3.40	3.47	R: 3.66/S: 3.53	$^{3}J_{H6R,H5}$=5.3Hz $^{3}J_{H6S,H5}$=1.7 Hz
	^{13}C	102.2	72.3	73.5	78.7	72.9	65.38	
C: 4,6-β-Glc	^{1}H	4.20	3.02	3.31	3.414	3.47	R: 3.66/ S: 3.53	$^{3}J_{H6R,H5}$=5.3.Hz $^{3}J_{H6S,H5}$=1.7 Hz
	^{13}C	102.2	72.3	73.5	78.7	72.9	65.38	
D: 4-β-Glc	^{1}H	4.30	2.92	3.26	3.29	3.24	R: 3.59/ S: 3.45	$^{3}J_{H6R,H5}$=10.4Hz $^{3}J_{H6R,H5}$=ND
	^{13}C	95.3	73.4	73.9	78.6	74.2	59.4	
E: t-α-Xyl	^{1}H	4.58	3.17	3.36	3.25	3.36/3.20		
	^{13}C	97.8	71.0	72.5	69.0	60.9		
F: t-α-Xyl	^{1}H	4.61	3.19	3.36	3.25	3.35/3.22		
	^{13}C	98.4	70.7	72.5	69.0	61.0		
G: t-α-Xyl	^{1}H	4.61	3.19	3.36	3.25	3.35/3.22		
	^{13}C	98.4	70.7	72.5	69.0	61.0		

A combination of 2-D NMR experiments was harnessed to assign all of the spin systems of the oligosaccharide (Table 8.4). Analysis of $^{3}J_{HH}$ coupling constants from 1D ^{1}H-NMR (Figure 8.10) and DQF-COSY spectra of XXXG, together with analysis of the *intra*-residual NOE contacts, indicated that all sugars were present exclusively in $^{4}C_{1}$ chair conformation, consistent with the observations of GXXG discussed above. Three-bond coupling constants of the ligand in the free and bound state (1:7 molar ratio of *Cj*Xyl31A-D659A to XXXG) were further analyzed to obtain the rotamer distributions in the bound state (Table 8.5). The solution structures obtained here were similar to those previously determined in the seminal work of Picard et al [25]. Again paralleling observations on GXXG, the two central Glc residues (residues **B** and **C**) had an essentially equal distribution of *gg*:*gt* rotamers in the bound state, while the data suggested that the *gg* rotamer of the non-reducing-end Glc (residue **A**) and the *gt* rotamer of the reducing-end Glc residue (residue **D**) were selected upon XXXG binding.

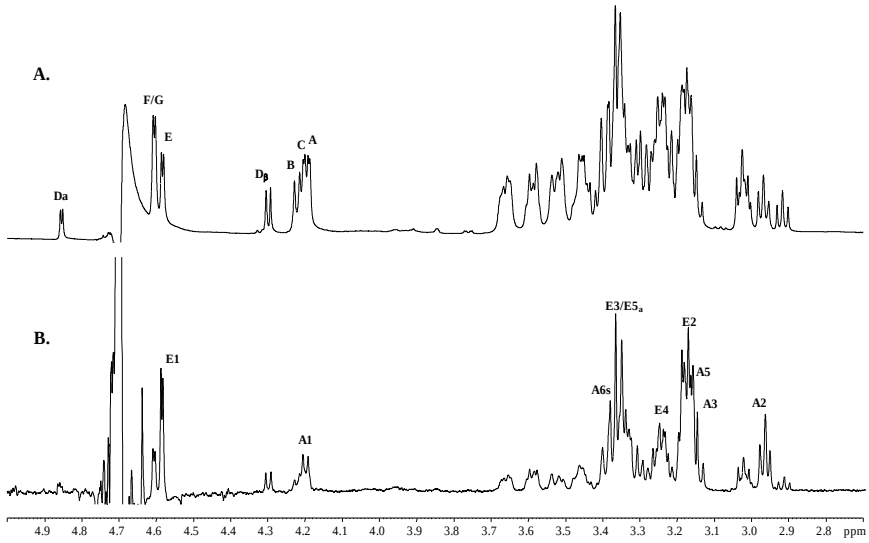


Fig 8.10: One-dimensional proton NMR spectral analysis of the binding of XXXG to the inactive *Cj*Xyl31A-D659A general acid/base residue mutant. **A**. ^{1}H

NMR spectrum of GXXG in presence of *Cj*Xyl31A. Anomeric signals from each of the monosaccharide residues A-F are indicated. **B**. STD spectrum. An on-resonance irradiation frequency at 0 ppm and a saturation time of 2 seconds was used, with a 1:45 enzyme-ligand ratio.

Table 8.5 Relative observed STD effects for XXXG bound to the catalytically inactive *Cj*Xyl31A-D659A general acid—base residue mutant.

Unit	Relative STD signal [a]					
	H-1	H-2	H-3	H-4	H-5	H-6
A: 6-β-Glc	40 %	100 %	75 %	64 %	83 %	R: 30% S: 40 %
B: 4,6-β-Glc	23 %	32 %[a]	53 %	-	27 %	R: 27% S: 34 %
C: 4,6-β-Glc	20 %	32 %	53 %	-	27 %	R: 27% S: 34 %
D: 4-β-Glc	30 %	12 %	-	-	-	R: 35% S: 16 %
E: t-α-Xyl	98 %	67 %	51 %	35 %	ax.: 46% eq.: 10%	
F/G: t-α-Xyl	20 %	-	-	-	eq: 17%	

Using the same strategy as for the *Cj*Xyl31A:GXXG complex, we have employed a battery of 1- and 2-D STD experiments to analyse the binding to *Cj*Xyl31A-D659A to XXXG. Here again, the data showed that the whole oligosaccharide interacted with the catalytically crippled enzyme. As for the *Cj*Xyl31A:GXXG complex, large STD signals for the terminal non-reducing-end Glc residue (**A**) were observed, with the additional observation of large signals for the ultimate, non-reducing-end Xyl residue (**E**) (Table 8.5 and Figure 8.11).

These results are again consistent with the oligosaccharide binding in the expected productive mode, in which the terminal isoprimeverose moiety [Xyl*p*-(1→6)-Glc*p*-(1→4)] spans the catalytic residues, placing the Xyl and Glc units in the -1 and +1 subsites, respectively[1,33]. The STD signals obtained for the remainder of the monosaccharide residues provided further evidence of an extended binding surface of *Cj*Xyl31A, comprised of manifold positive subsites, which harness both the backbone Glc residues and the pendant Xyl branches of XXXG.

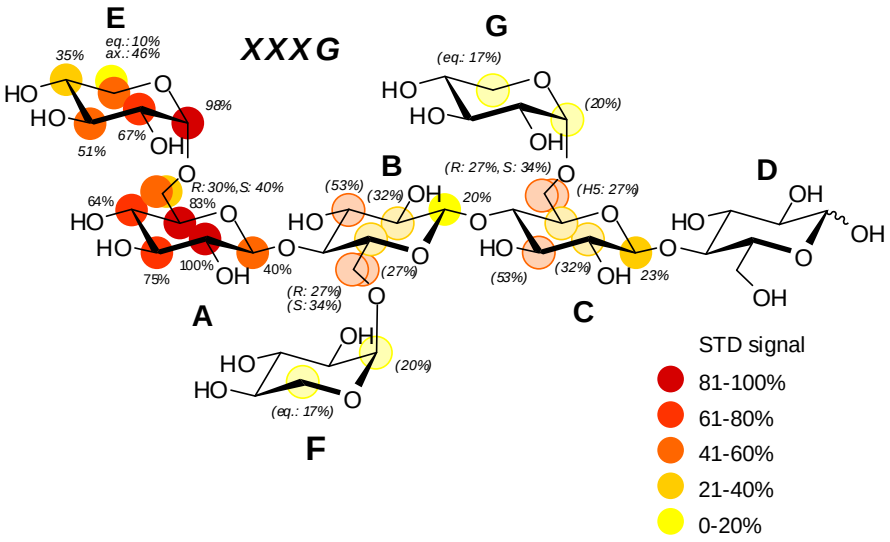


Fig. 8.11: Relative observed STD effects for XXXG:*Cj*Xyl31A-D659A mutant. Interaction pairs were calculated according to the equation: $(I_o-I_{sat})/I_o$ and normalized to the largest STD value set at 100%. Relative STD intensities are coded with solid colors according to the scale shown. Ambiguous values, due to signal overlap, are represented in parenthesis and shown with faded color-coding.

Table 8.6 Experimental (from ROESY experiments) and calculated (from MD and MM calculations) distances for the heptasaccharide XXXG.

	Free	Bound	MD (gg conformer)	MD (gt conformer)
G1/C6$_S$	2.55	2.4	2.36± 0.17	2.39±0.17
G1/C6$_R$	2.9	3.0	3.13± 0.25	3.14±0.26
F1-B6$_S$	2.55	2.4	2.38±0.16	2.40±0.17
F1-B6$_R$	2.9	3.0	3.03± 0.24	3.09±0.27
E1/A6$_S$	2.48	2.4	2.40±0.17	2.40±0.18
E1/A6$_R$	3.1	2.9	2.96±0.25	3.12±0.26
C1-D4	2.4	2.4	2.67±0.28	2.57±0.25
B1-C4	2.54	2.62	2.67±0.26	2.69±0.29
A1-B4	2.55	2.53	2.46±0.26	2.57±0.26

Mirroring the GXXG:*Cj*Xyl31A system, tr-NOESY analysis of the XXXG:*Cj*Xyl31A-D659A complex indicated only minor conformational changes upon substrate binding (Table 8.6). NMR data were likewise used to obtain a conformational model of XXXG oligosaccharide by MM and MD approaches, using two starting structures built considering the equal distribution of *gg* and *gt* rotamers of the internal (**B** and **C**) Glc residues (data not shown). Average distances obtained from $\langle r^{-6} \rangle$ values in the simulation were in good agreement with the experimentally determined values (Table 8.6); a representative conformer, onto which the observed STD enhancements have been mapped, is shown in Figure 8.12.

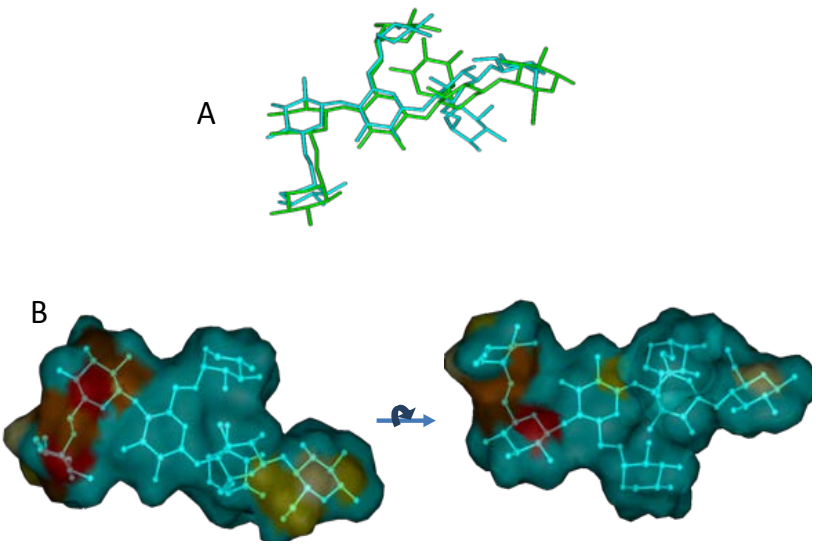


Fig 8.12: Structures of oligosaccharides XXXG in the bound conformation with *Cj*Xyl31A-D659A. **A**. Dominant conformer of XXXG derived from NMR data and MD/MM simulations. **B**. STD-derived epitope mapping on the molecular envelope of XXXG, with color coding from the highest (red) to lowest (yellow) observed STD effect.

8.4 Docking analysis of XXXG bound to *Cj*Xyl31A

NMR data enabled the rational docking of the substrate into the enzyme active-site, thereby revealing, for the first time, a holistic picture of the interaction of *Cj*Xyl31A and its appended PA14 domain with a complex xyloglucan oligosaccharide.

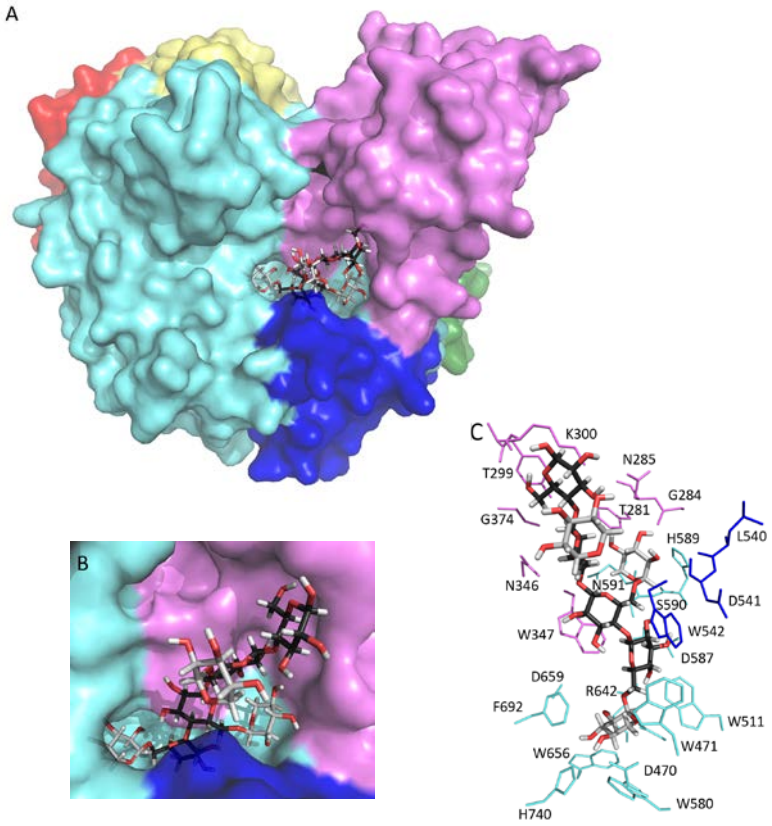


Fig. 8.13: The XXXG conformer docked into the active-site of *Cj*Xyl31A based on NMR data with the catalytic domain colored cyan, the PA14 domain violet and a loop insert ("Ins1") in the catalytic domain blue. **A**. Global view **B**. Expansion of the active-site region in wall-eyed stereo. **C**. Specific interacting active-site residues in wall-eyed stereo. The glucan backbone is depicted in black sticks with the appended Xyl residues in light gray sticks. The figure was prepared using PyMOL.

In detail, the region of XXXG comprising the glucose backbone and the terminal α-D-Xyl-(1→6)]-α-D-Glc glycosidic linkage was observed to be relatively rigid. This allowed us to manually dock the oligosaccharide onto the 3-D structure of a covalently trapped β-5-fluoroxylosyl-*Cj*Xyl31A enzyme intermediate [1].

The resulting complex clearly implicates the involvement of the whole saccharide in binding the enzyme through an extended active site pocket formed, in large part, by the PA14 domain (Figure 8.13). The PA14 domain plays a key role in the interaction with the linear glucan backbone, comprising a shelf consisting of residues Trp^{347}, Asn^{346}, Gly^{374}, Tyr^{299} and Lys^{300} (Figure 8.13c). Residue **B** of the XXXG ligand makes additional contacts with Trp^{511} from the catalytic domain and Trp^{542} from a loop insertion, previously denoted "Ins1", in the $(\beta/\alpha)_8$ barrel. Trp^{542} also coordinates residue **C** together with Trp^{347}. The appended Xyl residue **F** lies in a pocket created by residues Tyr^{281}, Gly^{284} and Asn^{285} from the PA14 domain; residues 589-591/His^{589}, Ser^{590}, and Asn^{591} from the catalytic domain; and Leu^{540} and Asp^{541} from Ins1. The 2-hydroxyl moiety of residue **F** faces the solvent, which would allow additional substitutions on this xylosyl unit, such as Gal*p* or Ara*f*, which are commonly found on diverse xyloglucans. Likewise, the other internal xylosyl unit (residue **G**) is not coordinated by the protein, so extension of this sidechain are also unlikely to significantly affect oligosaccharide binding. These observations do not only describe the specialized topology of the enzyme with the PA14 domain, but also explain the good activity of *Cj*Xyl31A on XLLG where residues **F** and **G** are further substituted at O2.

8.5 Conclusion

In this study, NMR, crystallographic, and molecular modeling data strongly implicate the role of the PA14 domain in the recognition of the complex substrates by *Cj*Xyl31A. In detail, a combination of great variety of advanced NMR spectroscopic techniques have

revealed the conformations and epitope patterns of large xyloglucan oligosaccharides – a substrate and a product – bound to the main alpha-xylosidase of the soil saprophyte *Cellvibrio japonicus*, *Cj*Xyl31A.

NMR analysis demonstrated that, in contrast with most *exo*-glycosidases,[5] *Cj*Xyl31A engulfs its substrates (typified by the highly branched xyloglucan heptasaccharide XXXG) in a deep active-site pocket with an extended oligosaccharide binding platform. This peculiar active-site architecture arises from the presence of an accessory PA14 domain in *Cj*Xyl31A, which has not previously been described in GH31 enzymes, including the structural archetype, *E. coli* α-xylosidase YicI [1].

The observed oligosaccharide epitope patterns mirror kinetic data on *Cj*Xyl31A, which indicate that recognition of the terminal tetrasaccharide motif of XXXG (underlined) is of primary importance to catalysis. Notably, the present analysis extends this earlier kinetics work to reveal subtle interactions along the entire oligosaccharide backbone, including the distal, reducing-end glucose residue. Taken together, the data imply that the active-site of *Cj*Xyl31A consists of up to four primary enzyme subsites with ramped affinity for backbone glucosyl units, in addition to secondary subsites that actively bind branching xylosyl residues. Given the previous lack of success in obtaining crystal complexes of *Cj*Xyl31A with xyloglucan oligosaccharides, [1] the present study provides an illustrative example of the power of magnetization-transfer NMR to provide ligand binding information which is otherwise unattainable by other methods [34]. As such, the complex epitope map represents one of the largest yet observed in an oligosaccharide-protein complex, which in the case of canonical lectin-glycan interactions are typically restricted to 1-3 contiguous monosaccharides.

Here, we hypothesize that the presence of this domain in GH31 members may serve as a molecular fingerprint to identify individual enzymes with a penchant for the degradation of large

xyloglucan fragments, thereby guiding fundamental biochemical analyses that will set the stage for the development of improved applications in plant biomass utilization.

References

[1] Larsbrink *et al. Biochem. J.*, **2011**. 436, 567–580.
[2] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A. *Chemistry*. **2012**. 18(42),13395-404.
[3] Deboy, R.T.; Mongodin, E.F.; Fouts, D.E.; Tailford, L.E.; Khouri, H.; Emerson, J.B.; Mohamoud, Y.; Watkins, K.; Henrissat, B.; Gilbert, H.J.; Nelson, K.E. *J. Bacteriol.*, **2008.** *190*, 5455-5463.
[4] Henrissat, B.; Callebaut, I.; Fabrega, S.; Lehn, P.; Mornon, J.P.; Davies, G. *Proc. Natl. Acad. Sci. U.S.A.*, **1995.** *92*(15), 7090–7094.
[5] Henrissat, B.; Davies, G. *Structure*, **1995**. *3*(9), 853–859.
[6] Gilbert, H.J.; Stalbrand, H.; Brumer, H. *Curr. Opin. Plant Biol.*, **2008.** *11*, 338-348.
[7] Himmel, M.E.; Ding, S.Y.; Johnson, D.K.; Adney, W.S.; Nimlos, M.R.; Brady, J.W.; Foust, T.D. *Science* **2007**. *315*, 804-807.
[8] van der Maarel, M.; van der Veen, B.; Uitdehaag, J.C.M.; Leemhuis, H.; and Dijkhuizen, L. *J. Biotechnol*., **2002**. *94*, 137–155.
[9] Mansfield, S.D.; Esteghlalian, R. *Applications Of Enzymes To Lignocellulosics, Vol*.855, (Ed: S.D. Mansfield, J.N.Saddler,) ACS Symposium Series pp. 2-29.
[10] Ariza, A.; Eklöf, J.M.; Spadiut, O.; Offen, W.A.; Roberts, S.M.; Besenmatter, W.; Friis, E.P.; Skjot, M.; Wilson, K.S.; Brumer, H.; Davies, G. *J. Biol. Chem.*, **2011**. *286*, 33890-33900.
[11] Grishutin, S.G.; Gusakov, A.V.; Markov, A.V.; Ustinov, B.B.; Semenova M.V.; and Sinitsyn, A.P. *Biochim Biophys Acta Gen Subj*, **2004**. 1674, 268–281.
[12] Edwards, M.; Dea, I. C.; Bulpin, P. V. and Reid, J. S. J. Biol. Chem. **1986**. 261, 9489–9494.
[13] Leonard, R.; Pabst, M.; Bondili, J. S.; Chambat, G.; Veit, C.; Strasser, R. and Altmann, F. *Phytochemistry*, **2008**. 69, 1983–1988.
[14] Gilbert, H. J.; Stalbrand, H.; and Brumer, H. *Curr. Opin. Plant Biol*. **2008**. 11, 338–348.
[15] Cantarel, B.L.; Coutinho, P.M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. *Nucleic Acids Res*., **2009.** 37, D233-D238.
[16] Zechel, D.L.; Withers, S.G. *Acc. Chem. Res*., **2000.** *33*, 11-18.
[17] Zhang, R.; Withers, S.G.http://www.cazypedia.org/, accessed 5 January **2012**.

[18] Schmidt, R.K.; Karplus, M.; Brady, J.W. *J. Am. Chem. Soc.*, **1996**. *118*, 541–546.
[19] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed. Engl.*, **2003**. *42*, 864.
[20] Mayer, B.; Meyer, B. *J. Am. Chem. Soc.*, **2001**. *123*, 6108.
[21] Mari, S.; Serrano-Gomez, D.; Canada, F.J.; Corbi, A.L.; Jimenez-Barbero, J. *Angew. Chem. Int. Ed. Engl.*, **2004.** *44*, 296.
[22] Angulo, J.; Langpap, B.; Blume, A.; Biet, T.; Meyer, B.; Krishna, N.R.; Peters, H.; Palcic, M.M.; Peters, T. *J. Am. Chem. Soc.*, **2006**. *128*, 13529.
[23] Yuan, Y.; Bleile, D.W.; Wen, X.; Sanders, D.A.; Itoh, K.; Liu, H.W.; Pinto, B.M.. *J. Am. Chem. Soc.*, **2008**. *130*(10), 3157-68.
[24] Stenutz, R.; Carmichael, I.; Widmalm, G.; Serianni, A.S. *J. Org. Chem*., **2002**. *67*(3), 949-58.
[25] Picard, C.; Gruza, J.; Derouet, C.; Renard, C.M.; Mazeau, K.; Koca, J.; Imberty, A.; Hervé Du Penhoat, C. *Biopolymers*, **2000.** *54*(1), 11-26.
[26] Nieto, L.; Canales, A.; Giménez-Gallego, G.; Nieto, P.M.; Jiménez-Barbero, J. *Chem. Eur. J.*, **2011**. *17*(40), 11204-9.
[27] Davies, G.J.; Wilson, K.S.; Henrissat, B. *Biochem. J.*, **1997**. *321*, 557-559.
[28] Szczepina, M.G.; Zheng, R.B.; Completo, G.C.; Lowary, T.L.; Pinto, B.M. *Bioorg Med Chem*., **2010**. *18*(14), 5123-8.
[29] *The Nuclear Overhauser Effect in Structural and Conformational Analysis* (Eds.: D. Neuhaus, M. P. Williamson), Wiley-VCH, New York, **2000**.
[30] *NMR Spectroscopy of Glycoconjugates* (Eds.: J. Jimenez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**.
[31] Johnson, M.A.; Pinto, B.M. *Carbohydr Res.*, **2004.** *339*(5), 907-28.
[32]. Enríquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penadés, S. *Chem. Eur. J.*, **2011.** *17*(5), 1547-60.
[33] Kim, Y.W.; Lovering, A.L.; Chen, H.; Kantner, T.; McIntosh, L.P.; Strynadka, N.C.; Withers, S.G. *J. Am. Chem. Soc.*, **2006**. *128*, 2202-2203.
[34] Roldos, V.; Canada, F.J.; Jimenez-Barbero, J. *Chembiochem*, **2011**. *12*, 990.

SECTION IV

EXPERIMENTAL METHODS

Chapter 9

Materials and methods*

9.1 Muropeptides purification

9.1.1 Bacterial growth and peptidoglycan extraction

Bacillus subtilis strain was grown at 37°C in liquid shake culture (200 rpm) in Luria Bertani broth for 18 h (early stationary phase). The broth culture was centrifuged, the cells were washed with H_2O, ethanol, acetone and ethylic ether.
Dried cells were extracted for PGN as described elsewhere [1-3]. Briefly, *B. subtilis* cells were suspended in ice-cold water and added dropwise to boiling 8% SDS and boiled for a further 30 min. After cooling to room temperature, the SDS-insoluble material was collected by centrifugation. The pellets were washed several times with water until no SDS could be detected. High molecular weight glycogen and covalently bound lipoprotein were removed by treatment with α-amylase and trypsin. Further sequential washes with 8M LiCl, 0.1M EDTA and acetone were sequentially performed.

9.1.2 Preparation of muropeptides, GC-MS, HPLC and LC-MS analyses

The isolated peptidoglycan was degraded with muramidase mutanolysin from *Streptomyces globisporus* ATCC21553 (Sigma-Aldrich) at 37°C overnight. The enzyme reaction was stopped by boiling (10 min), and insoluble contaminants were removed by

* - For cristallography, microscopy, homology modeling, docking, CD, SEC-LS, SPR, SAXS and ITC experimental details see ref [7-9, 12]
- Experimental procedures of biomolecular analysis on LysM1-2 (Chapter 7) [11]→personal data Prof. Berisio, CNR of Naples, and Prof. Shibuya, University of Tokyo.

centrifugation. PGN preparation was analysed for sugar and amino acid composition [4,5] by GC-MS. Briefly, 250µg of PGN have been treated with MeOH/HCl 4M for 16h at 85°C and further neutralised and acetylated with 100µl of acetic anhydride and 200µl of pyridine. Finally, the sample was analysed by GC-MS. A SLB capillary column (0.18 mm x 20 m, Supelco) was used and the temperature program was: 100 °C for 5 min, then 5 °C min^{-1} to 300 °C. In order to perform the aminoacid composition, to 100 µL of a 0.1 M HCl solution of muropeptides, 100µL of tert-butyldimethylsilyl-N-methyltrifluoroacetamide were added. After 4h at 100°C, the sample was neutralised and underwent GC-MS analysis. The temperature program was: 100 °C for 1 min, then 35 °C min^{-1} to 300 °C.

The generated muropeptides were dissolved in 0.5 M sodium borate buffer, pH 9.0 and solid sodium borohydride was added immediately. After incubation for 2h at room temperature, excess borohydride was destroyed with 2 M HCl. Finally, the sample was desalted by using a size-exclusion chromatography (Sephadex G10) and adjusted to pH 3-4 with TFA.

Reduced muropeptides were fractionated by reverse phase HPLC as reported elsewhere [2,3]. In order to well separate the muropeptides, the sample was eluted with a gradient from 0% to 30% acetonitrile 50% with 0.05% TFA in 30 min, and to 100% in 40 min. Detection was performed at A210 nm. ESI-MS spectra were obtained by using the same conditions of the elution of HPLC analysis (Agilent 1100MSD). The final mixture of muropeptides **2** and **1** (3:1 ratio) was further purified by gel-filtration chromatography (Sephadex G10 and Biogel P2).

9.2 ECPrkC expression and purification

The plasmidic construct corresponding to the extra-cellular region of *B. subtilis* (EC-PrkC, residues 356-648) was prepared as previously described [6]. The overexpression of EC-PrkC containing

an N-terminal hexa-histidine tag was carried using *E. coli* DH5α cells (Invitrogen). Briefly, an overnight starting culture of 100 mL was prepared to grown 1 L of Luria Bertani (LB) containing 100 ug/mL of ampicillin which was then induced with 0.02% (w/v) L-arabinose at OD600=0.6 and harvested after three hours.
The protein was isolated by sonicating resuspended cells in a binding buffer (5 mM imidazole, NaCl 150 mM, 20 mM TrisHCl, 5%(v/v) glycerol pH 8.0) containing a protease-inhibitor cocktail (Roche Diagnostic). The lysate was cleared by centrifugation at 18,000 rpm, and the supernatant loaded on a 5-mL Ni-NTA column (GEHelthcare), equilibrated with binding buffer. After washing with 10 volumes of binding buffer, a linear gradient of imidazole (5-300 mM) was applied to elute the protein. The protein eluted at 150 mM of imidazole and fractions containing the protein were pooled and loaded on a gel filtration column (Superdex 200, GeHelthcare) equilibrated in NaCl 150 mM, 5%(v/v) glycerol, 20 mM TrisHCl pH 8.0, for a further purification step. The protein eluted in a single peak and was homogeneous, as judged by SDS–PAGE analysis. All the purification steps were carried out at 4°C. The protein was concentrated using a centrifugal filter device (Millipore), and the concentration was determined by BCA protein assay (Pierce).

9.3 NMR analysis

9.3.1. General conditions of NMR analysis

Muropeptides (Chapter 3 [7])

Experiments on the ligands (muropeptide **1**, **2**, try-lys and tri-DAP) were recorded at 283K. All the samples were dissolved in deuterated phosphate (PBS) buffer at pH 7.4. To confirm the amino acid sequence of peptide stems, spectra in PBS H2O/D2O 95:5 were recorded to observe the amide group signals.

For bound state, TR-NOESY experiments were performed with protein-ligand molar ratio 1:25, by using mixing times from 150 to 400 ms. For the acquisition of STD NMR spectra, the protein-ligand molar ratio varied from 1:25 to 1:50 and saturation times between 0.5 s to 2 s were used. The on resonance frequency was set at 0 ppm.

Carbohydrates from bacterial LPS (Chapter 4 [8])

NMR experiments were performed at 283K. All the samples were dissolved in deuterated 0.1 M Tris–HCl buffer with 2 mM $CaCl_2$ (pH 7.5).
For the bound ligands, STD NMR experiments were performed using protein–ligand molar ratio varied from 1:50 to 1:100 and saturation times between 0.5 and 2.5 s were used. The on-resonance frequency was set to −0.7, 0.5 and 6.8 ppm.

Carbohydrates from Ebola virus (Chapter 5 [9])

NMR experiments were performed at different temperatures: 307 K for the *N*-acetylglucosamine and α-D-*O*-methylmannose, 295 K for disaccharides **1** and **2**. All the samples were dissolved in deuterated Tris buffer (pH 7.4, 10 mM) with NaCl (140 mM) and $CaCl_2$ (3 mM). STD NMR experiments on the mixtures were performed with protein/ligand molar ratios varying from 1:50 to 1:100 with saturation times between 0.5 s and 5 s. The on-resonance pulse was at 0 ppm. Tr-NOESY experiments were carried out at a 20:1 product-to-enzyme molar ratio with different NMR mixing times.

O-chain from Burkolderia anthina LPS (Chapter 6 [10])

NMR spectra were recorded at 298K. All the samples (O-chain from *B.anthina* LPS, synthetic tri- and hexasaccharide O-antigen repeating unit) were dissolved in deuterated phosphate buffer (pH 7.4). For STD NMR spectra, the employed protein-ligand molar

ratio varied from 1:25 to 1:60, while the saturation times were varied between 0.5 s to 5 s. The on resonance pulse was set at the aromatic region (6.8 ppm). Tr-NOESY experiments were carried out at a 7:1 to 20:1 product-to-enzyme molar ratio with different NMR mixing times.

Chito-oligosaccharides (Chapter 7 [11])

NMR experiments were performed at 298K. All the samples were dissolved in deuterated PBS (pH 7.4), DTT 10 mM and CHAPS 0.005%. STD NMR spectra were performed with a protein/ligand ratio varying from 1:50 to 1:80, and the saturation times oscillated between 0.5 s to 5 s. The on resonance frequency was set to -1 and 6.8 ppm. TR-NOESY experiments were performed by using protein/ligand molar ratio of 1:10 with different mixing times.

Xyloglucans (Chapter 8 [12])

Experiments on free ligand were recorded at 283K. The samples (hexa- and hepta-saccharides) were dissolved in deuterated phosphate buffered saline (PBS) at pH 7.4. For the acquisition of STD NMR spectra, the protein/ligand molar ratio varied from 1:25 to 1:50, and saturation times between 0.5 and 3 s were used. The on-resonance frequency was
set at 0 ppm. TR-NOESY experiments were performed by using protein/ligand molar ratios of 1:7→1:14 with mixing times of 150–400 ms.

9.3.2 Ligand spectroscopic characterization

The complete assignment of the ^{1}H and ^{13}C resonances, for each free ligand analyzed in this Thesis, was achieved combining the information from 1D-NMR and 2D-NMR DQF-COSY, TOCSY, NOESY, t-ROESY and ^{1}H-^{13}C HSQC experiments. All NMR spectra were recorded on a Bruker 600 MHz DRX spectrometer equipped with a triple resonance probe head (^{1}H, ^{13}C, and ^{15}N) with

a cryo probe. The spectra were calibrated with [D4](trimethylsilyl)propionic acid, sodium salt (TSP; 0.05mM) as an internal reference. ROESY and NOESY spectra were measured using data sets (t1xt2) of 4096x256 points with mixing times between 100 ms and 500 ms. Double quantum filtered phase sensitive COSY experiments were performed using data sets of 4096x512 points; total correlation spectroscopy experiments (TOCSY) were performed with a spinlock time of 100 ms, using data sets (t1xt2) of 4096x256 points. In all homonuclear experiments the data matrix was zero-filled in the F1 dimension to give a matrix of 4096x2048 points and was resolution enhanced in both dimensions by a cosine-bell function before Fourier transformation. Coupling constants were determined on a first-order basis from 2D phase sensitive double quantum filtered correlation spectroscopy (DQF-COSY) [13,14]. Heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond correlation (HMBC) experiments were measured in the ^{1}H-detected mode via single quantum coherence with proton decoupling in the ^{13}C domain, using data sets of 2048x256 points. Experiments were carried out in the phase-sensitive mode according to the described method [15]. A 60 ms delay was used for the evolution of long-range connectivity in the HMBC experiment. In all heteronuclear experiments the data matrix was extended to 2048x1024 points using forward linear prediction extrapolation [15,16].

9.3.3 Bound ligands NMR analysis

For the bound ligands, STD and TR-NOE experiments were performed as following.
TR-NOESY spectra were recorded by using protein/ligand molar ratios of 1:5→1:25 with mixing times of 150–500 ms. The concentration of the protein was 10-50μM.
For the acquisition of STD NMR spectra, the protein/ligand molar ratio varied from 1:25 to 1:100, and saturation times between 0.5

and 5 s were used. A T1ρ filter (spin-lock pulse: 50 ms) to eliminate the unwanted broad resonance signals of the protein was used. The on-resonance frequency was set at different ppm values according to the analyzed system, since it has to fall in a region in which ligand signals must be absent, necessary condition to avoid false-positive results in the STD spectrum. The off-resonance frequency was always maintained at δ=40 ppm, where no protein or ligand signals resonated. For protein saturation, a train of 40 Gaussian shaped pulses each of 50 ms (delay between pulses: 1 ms, field strength: 21 Hz) was employed. The saturated and reference spectra were acquired simultaneously by creating a pseudo-2D experiment. Reference experiments on free ligands, in the same conditions, were carried out to assure the absence of direct irradiation of the ligand. STD spectra were performed with 16 k and 32 k data points. The original free induction decay (FID) was zero-filled to 64 k, and Fourier transformation with use of an exponential window function was applied (lb = 1–2 Hz).

The 2D STD NMR spectra (2D STD-TOCSY and 2D STD-HSQC) were recorded by using data sets of 2048_256 points, and the same parameters used for the STD and TOCSY/HSQC spectra were employed.

When necessary, false positives or artifacts in STD spectra were avoided performing STDD NMR experiment (Section III, Chapter 7). The STDD NMR required the acquisition of an additional STD spectrum of the receptor in absence of ligands, in the same conditions, that is then subtracted from the STD spectrum acquires in the presence of ligands.

To determine the magnitude of the STD effects, the intensity of the signals in the STD NMR spectra were compared with the signal intensities of a reference spectrum (off-resonance). The STD signal with the highest intensity was set to 100% and the other signals were normalized to this peak. In order to refine these qualitative results, STD build up curves were derived [17-19]. The spectra acquired at different saturation times were fitted to a

monoexponential equation: STD= STDmax(1 – exp(-ksat t)), where STD is the STD intensity of a given proton at a saturation time t, STDmax is the asymptotic maximum of the curve, and ksat stands for the observed saturation rate constant. The epitope map of each ligand was obtained by normalizing all the values of different protons ligand to the largest STDfit, giving STDepitopes fit.
The *K*d values were determined with ligand concentrations of 0.3, 0.4, 0.6, 1, 1.3, and 3.5 mm with use of the one site binding model: STD intensity = (STDmax + [L])/*K*d + [L]. Ligand was titrated into protein sample from concentrated stock solutions to minimize dilution effects. STD-AF values, at a saturation time of 2 s, were plotted as a function of the concentration of ligand and in order to yield *K*d, the corresponding curve was fitted to the dose-response equation described above.
Data acquisition and processing were performed with TOPSPIN software.

9.4 Conformational analysis: MM and MD simulations

The determination of the conformation and dynamics of glycans, in free and bound state, can be performed by using molecular modeling techniques supported by NMR spectroscopy data. The individuals monosaccharides show a limited flexibility and in an oligosaccharide, they are assumed as rigid ring structures, so the conformation of a glycan can be performed just defining the behaviour of each glycosidic linkage through molecular mechanics and dynamic simulation.

9.4.1 Molecular mechanics simulations

Since the conformation of a saccharidic chain is mainly defined by the relative orientation of the sugar moieties, i.e. by Φ and Ψ torsion angles, the first step in the conformational analysis was to

build the potential energy surfaces for each disaccharide connected by a glycosidic linkage; Φ represents the torsion angle about H1-C1-O-CX' whereas Ψ about C1-O-CX'-HX'. Molecular mechanics calculations were performed using the MM3* force field as included in MacroModel 8.0. The bulk of the solvent was simulated using a constant dielectric of 80. For each disaccharide both Φ and Ψ dihedral angles were varied incrementally using a grid step of 18 degrees. The corresponding flexible maps were drawn as 2D contours plots using the graphical tools of Macromodel 9.0.
Molecular mechanics provided a first estimation of the conformational regions energetically accessible. The resulting adiabatic energy maps indicated global and local minima.

9.4.2 Molecular dynamic calculations

Molecular dynamics simulations are typically carried out in four steps (Figure 9.1) under
isothermal-isobaric conditions.

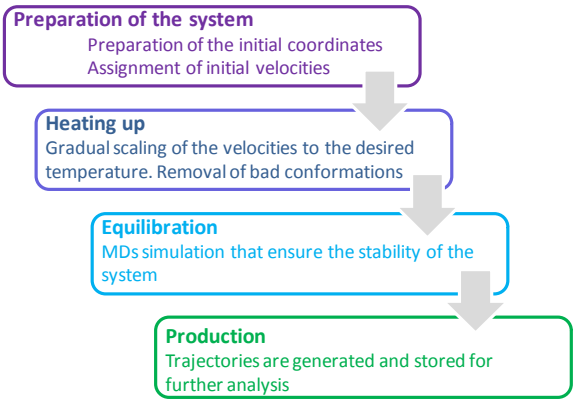


Fig. 9.1: General protocol for running MDs simulations. In the first step, the system, derived from NMR, X-ray or homology modeling, is prepared, adding missing atoms, and submitted to minimization cycles. The second step consists of heating the system in order to remove bad contacts, increasing and assigning new atom velocities. The initial velocities are generally determined from the standard temperature-dependent Maxwell-Boltzmann distribution. The third step is called equilibration, in which energy, temperature and RMSD (Root Mean

Square Deviation) of the system converge to stable values that allow the collection of trajectories (production step)that will be successively analyzed.

Molecular dynamic calculations were performed in explicit water using the Glycam06 force filed and the AMBER 9 software package [20].

All the oligosaccharides were built up and minimized by using Maestro package and the carbohydrate builder utility of the glycam website (www.glycam.com), in order to have a file .pdb with the appropriate atom types as requested by the AMBER 9 software and then the torsional angles were set to the values obtained through the molecular mechanics calculations. The input files, were generated using the antechamber at first (.prepin, .frcmod) and then xleap (.prmtop and .inpcrd) modules of the software package, all the minimization and molecular dynamic calculations were performed using the SANDER module. The calculations employed glycam 06e as force field, using SHAKE for the C-H bonds and 1 fs of integration step. The corresponding molecules were immersed within an octahedral box of water (TIP3P) of the proper size and the remote interactions were calculated using a cut off of 8 angstroms. After the preparation of the input files, a energy minimization process was performed to refine the initial structure. The system was minimized, at first, holding the solute fixed, while a second minimization was performed on the entire system. Furthermore, the whole system was slowly heated from 0 to 300 K using a weak restrain on the solute and then, the system was equilibrated at 300 K using constant pressure and removing the restrains on the solute. The system coordinates were saved and used for the 10 ns simulations using the Sander module implemented in Amber9. The stability of energy, pressure, temperature and other thermodynamic parameters were monitored along the trajectory and then torsions and distances were extracted. Ptraj was the utility used for analyzing and processing trajectories and coordinate files created from the MD simulations. Maestro 9.0

was used as well, after the appropriate conversion of the output files.

Ensemble-average interproton distances were calculated using the NOEPROM program by applying the isolated spinpair approximation as reported [21,22].

References

[1] Bustos, G.J.F.; Dougherty, T.J. *Antimicrob Agents Chemother*., **1987**. 31(2), 178–182.
[2] Glauner, B. *Anal. Biochem*., **1988**. 172, 451-464.
[3] Girardin, S.E.; Boneca, I.G.; Carneiro, L. A.M.; Antignac, A.; Jehanno, M.; Viala, J.; Tedin, K.; Taha, M-K.; Labigne, A.; Zahringer, U. *et al*. *Science*, **2003**. 300, 1584-1587.
[4] Molnar-Perl, I.; and Katona, Zs.F *Chromatographia supplement*., **2000**. 51, 5228-5236.
[5] Gilbart, J.; Harrison, J.; Parks, C.; Fox, A. *J Chromatogr*., **1988**. 441(2), 323-33.
[6] Shah, I.; Laaberki; M. H; Popham, D. L.; Dworkin, J. *Cell*, **2008**. *135*, 486.
[7] Squeglia, F.*; Marchetti, R.*; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *JACS*. **2011.** 133, 20676–20679. *equal contribution.
[8] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*, **2012.** 10,1387-98.
[9] Marchetti, R.; Lanzetta, R.; Michelow, C.J.; Molinaro, A. and Silipo, A. *Eur. J. Org. Chem*., **2012.** 27, 5275–5281.
[10] Marchetti R. et al., NMR binding studies of the interaction between a monoclonal antibody and the LPS O-antigen of *Burkholderia anthina* (Submitted, 2013)
[11] Marchetti R. et al., Molecular basis of chitin recognition and activation of rice chitin elicitor receptor (Manuscript in preparation).
[12] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H; Molinaro, A. *Chemistry*., **2012**. 18(42):13395-404.
[13] Piantini, U.; Sørensen, O.W.; Ernst, R.R., *J. Am. Chem. Soc*., **1982**.104, 6800-6801.
[14] States, D.J.; Haberkorn, R.A.; Ruben, D.J. *J. Magn. Res*. **1982**. 48, 286-292.
[15] de Beer, R.; van Ormondt, D. *Basic Princ. Prog*., **1992**. 26, 201.
[16] Hoch, J.C.; Stern, A.S.; In Hoch, J.C.; and Stern, A.S. (Eds.) *NMR data processing.* Wiley Inc. New York. **1996**. 77-101.

[17] Jayalakshmi, V.; Krishna, N.R. *J. Magn. Reson.*, **2002**. 155, 106-118.
[18] Yan, J.L.; Kline, A.D.; Mo, H.P.; Shapiro, M.J.; Zartler, E.R. *J. Magn. Reson.*, **2003**. 163, 270276.
[19] Szcepina, G.M.; Bleile, W.D.; Pinto, B.M. *Chem. Eur. J.*, **2011**. 17, 11446-11455.
[20] Case, D.A.; Darden, T.A; Cheatham, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Merz, K.M.; Pearlman, D.A.; Crowley, M.; Walker, R.C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K.F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D.H.; Schafmeister, C.; Ross, W.S.;and Kollman P.A. **2006** AMBER 9, University of California, San Francisco.
[21] Asensio, J.L.; Jimenez-Barbero, J. *Biopolymers*, **1995**. 35, 55 –75.
[22] Corzana, F.; Cuesta, I.; Freire, F.; Revuelta, J.; Bastida, A.; Jimenez-Barbero, J.; Asensio, J.L. *J. Am. Chem. Soc*., **2007**. 129, 2849 –2865.

Concluding remarks

The entire item of this Thesis was focused on improving the knowledge on proteins-glycoconjugates molecular recognition events.

Glycoconjugates coat most types of cell in nature and are intimately involved in various biological mechanisms, including cell differentiation, homing to specific tissues, cell adhesion and/or recognition, microbial pathogenesis and cell signaling among other processes [1,2]. Thanks to their high flexibility and structural complexity and diversity, glycoconjugates are a source of biological information written in a specific "language", known as glycocode. It can be deciphered by specific receptors, denoted as GBPs (Glycan Binding Poteins), which recognize carbohydrates and translate their code for the biological community. The glycan binding proteins can be subdivided into two major groups: intrinsic and extrinsic GBPs. The first ones are able to interact with glycans from the same organism mediating cell-cell interactions (Section II [3,4]); while, the extrinsic GBPs recognize sugars from a different organism and are mostly involved in pathogenic processes (Section III [5-8]) although they can also mediate symbiotic relationships. One of the key factors determining the establishment of any particular host-guest interaction, spanning the spectrum from symbiotic to pathogenic, is the pattern of glycoconjugates exposed on cell surfaces. Thus, the basis of the complex interplay between microbial glycoconjugates and their glycan receptors need to be elucidated, at a molecular level, in order to fully explain binding

[1] Haltiwanger, R.S.; Lowe, J.B. *Annu Rev Biochem*, **2004**. 73, 491-537.
[2] Ohtsubo, K.; Marth, J.D.; *Cell*, **2006**. 126, 855-867.
[3] Squeglia, F.*; Marchetti, R.*; Ruggiero, A.; Lanzetta, R.; Marasco, D.; Dworkin, J.; Petoukhov, M.; Molinaro, A.; Berisio, R.; Silipo, A. *JACS.* **2011.** 133, 20676–20679. *equal contribution.
[4] Marchetti, R.; Malinovska, L.; Lameignère, E.; Adamova, L.; de Castro, C.; Cioci, G.; Stanetty, C.; Kosma, P.; Molinaro, A.; Wimmerova, M.; Imberty, A.; Silipo, A. *Glycobiology*, **2012.** 10,1387-98.

and selectivity features and hence to assist in the design of new molecular probes and novel sugar-based therapeutic agents.

In this PhD project, an extended understanding of specific protein-glycoconjugates interactions, crucial in biological phenomena, has been accomplished by the use of NMR, taken together with molecular-modeling techniques.

We mainly focused on the use of two NMR methods, STD NMR and tr-NOESY, which are the most powerful and versatile techniques in the detection and characterization of binding processes between a small substrate and its large receptor. STD NMR experiments allows the detection and characterization of ligand binding to a protein by discriminating binding ligands from non-binding forms, and by providing binding epitopes of these ligands at atomic resolution. On the other hand, tr-NOESY method allows the knowledge of the recognized conformation of the ligand bound to the receptor, which has considerable implication for a rational structure-based drug design.

Thus, NMR experiments, as STD NMR and tr-NOESY, together supply a complete picture of a small ligand binding to its large receptor, providing fundamental insights on essential molecular recognition mechanisms.

[5] Marchetti, R.; Lanzetta, R.; Michelow, C.J.; Molinaro, A. and Silipo, A. *Eur. J. Org. Chem.*, **2012.** 27, 5275–5281.

[6] Marchetti R,Lanzetta R, Nilsson I., Vogel C., Reed D.E., AuCoin D.P., Silipo A., Jimenez-Barbero J., Molinaro A., **2013**, Submitted

[7] Marchetti, R.; Silipo, A.; Berisio, R.; Shibuya, N.; Molinaro, A. Manuscript in preparation.

[8] Silipo, A.; Larsbrink, J.; Marchetti, R.; Lanzetta, R.; Brumer, H.; Molinaro, A. *Chem Eur J*, **2012**. 15;18(42), 13395-404.

Abbreviations

AtCERK1	Chitin elicitor receptor kinase 1 of *Arabidopsis thaliana*
Bcc	Burkholderia cepacia complex
BCLA	*Burkolderia cenocepacia* lectin A
CD	Circular dichroism
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
*Cj*Xyl31A	*Cellvibrio japonicus* α-xylosidase 31 A
CL	C-type lectin
COSY	Correlation Spectroscopy
CRD	Carbohydrate-binding domains
CTL	C-type lectin
CTLR	C-type lectin receptor
DQF COSY	Double-quantum filtered COSY
EC-PrkC	Extracellular Serine/Threonine protein kinase
EG	EGF-like domain
EPS	Exopolysaccharide
FITC	Fluorescein isothiocyanate
GBPs	Glycan binding proteins
GC-MS	Gas chromatography-Mass spectrometry
GHs	Glycoside hydrolases
GL	S-type lectin
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single Quantum Correlation
IG2	Immunoglobulin
IL	L-type lectin
ITC	Isotermal Titration Microcalorimetry
kdo	3-deoxy-D-*manno*-oct-2-ulosonic acid
LC-MS	Liquid Cromatography – Mass Spectrometry
LCO	Lipo-chitooligosaccharide
L-FCN	L-ficolin
LPS	Lypopolysaccharide

LRR-RK	Leucine-rich repeat receptor kinases
LRR-RP	Leucine-rich repeat receptor proteins
LysM	lysine motif
mAb	Monoclonal antibody
MAMP	Microbial associated molecular pattern
MD	Molecular dynamics
MM	Molecular mechanics
MP	P-type lectin
MW	Molecular weight
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
OsCEBiP	Chitin elicitor binding protein of *Oryza sativa*
PAIIL	*Pseudomonas aeruginosa* lectin II
PAMP	Pathogen associated molecular pattern
PASTA	Penicillin binding associated and Serine/Threonine kinase associated
PBP	Penicillin binding protein
PBS	Phosphate buffer
PDB	Protein data bank
PGN	Peptidoglycan
PRR	Pathogen recognition receptor
rf	Radio frequency
rhMBL	Recombinant human mannose binding lectin
RMSD	Root Mean Square Deviation
ROESY	Rotating frame Overhauser Effect Spectroscopy
RP-HPLC	Reverse Phase – High Performance Liquid Cromatography
SAXS	Angle X-ray Scattering
SEC-LS	Triple-Angle Light Scattering
SPR	Surface Plasmon Resonance
STD	Saturation Transfer Difference
STD-AF	STD amplification factor
STDD	Saturation Transfer Double Difference
STPKs	Serine/threonine protein kinases
TLR	Toll like receptor
TM	Transmembrane

TOCSY	Total Correlation Spectroscopy
Tr-NOE	Transferred NOE
T-ROESY	Transverse Rotating-frame Overhauser Enhancement Spectroscopy
TSP	Trimethylsilyl propionic acid
XG	Xyloglucan

Acknowledgments

Desidero ringraziare tutte le persone che mi hanno accompagnato durante il cammino percorso in questi tre anni di dottorato. In modi diversi, hanno tutti contribuito a farmi arrivare fino a qui, suscitando in me nuovi interessi, spronandomi a inseguire nuovi obiettivi, ma soprattutto indicandomi, direttamente o indirettamente, il modo migliore per poterli raggiungere.

I miei più sentiti ringraziamenti vanno ad Alba, che, con la sua immensa bravura, mi ha seguito passo passo durante lo svolgimento del lavoro, con consigli e confronti che mi hanno aiutato ad intraprendere le scelte più appropriate.

A Tony, per aver creduto in me (più di quanto non faccia io stessa) fin dal primo giorno e per avermi aiutato a crescere e a maturare, non soltanto dal punto di vista scientifico. Grazie per la tenacia e la passione che mi hai trasmesso per questo lavoro; grazie anche per i rimproveri dolorosi che mi hanno permesso di imparare dagli errori, diventando più forte e con la consapevolezza che non si può pensare mai di essere arrivati ad un traguardo senza doverne affrontare un altro.

Gracias al Prof. Jesús Jiménez-Barbero por haberme aceptado en su labo enriquendo este doctorado.

Un grazie particolare va a Sara, che con la sua ineguagliabile dolcezza, anno dopo anno, mi insegna che cos'è una vera amicizia. Ti voglio un'infinità di bene amica mia.

Non posso non ringraziare TUTTI i ragazzi del laboratorio che hanno reso meno faticose le interminabili giornate di lavoro. In particolare, un sentito grazie va:
A Flaviana, per aver condiviso insieme le gioie, e soprattutto le difficoltà, che hanno costellato questi tre anni, nonché per l'infinita pazienza e il supporto sia tecnico che morale.
A Gennaro, che dal primo istante in cui è arrivato, ha portato una ventata di allegria rendendo ancora più divertenti i nostri pranzi e le pause caffè; senza dimenticare "Ciruzzo"!!!!
A Alberto, per la sua amicizia, ma soprattutto per le sue grigliate nelle indimenticabili fughe a Zolli e al Parco del Grassano!!

A las chicas y chicos que he encontrado en el CSIC de Madrid, un grupo tan dinámico y trabajador como divertido. Angeles, Pilar, Anita, Khouzaima, Maria, Blanca, Ana, Carmen, Luis y Alvaro, gracias por haber hecho de esta estancia una experiencia inolvidable. ¡Os echo de menos a todos!

Un grazie di cuore va ai miei più cari e sinceri amici: Carlo, Mariateresa, Salvatore, Armando, Daniela, Rosario, Adele, Francesco e Chiara, per tutte le serate e le risate che mi hanno regalato in questi anni.

Grazie al mio piccolo nipotino Andrea, perché nei giorni bui, quando tutto sembrava andare proprio storto e la legge di Murphy prendeva il sopravvento, mi bastava prenderlo in braccio e vederlo ridere per dimenticarsi di tutto il resto.

Ai miei genitori per avermi sempre spinto a realizzare i miei sogni, appoggiandomi in tutto e per tutto, sempre, incondizionatamente. Spero, con questo traguardo, di restituirvi una goccia del mare che mi avete donato.

A mio fratello, o forse dovrei dire "hermanno", perché so che per me ci sarà sempre, ovunque sia.

A Marco, il mio compagno di vita, la mia roccia, il mio sostegno. Grazie per tutto l'amore che mi dimostri ogni giorno.

Thank you!