



**HAL**  
open science

# Deciphering the recognition patterns between lipopolysaccharides and human lectins using nuclear magnetic resonance spectroscopy

Meriem Maalej

► **To cite this version:**

Meriem Maalej. Deciphering the recognition patterns between lipopolysaccharides and human lectins using nuclear magnetic resonance spectroscopy. *Biochemistry [q-bio.BM]*. Université Grenoble Alpes [2020-..]; Università degli studi di Napoli Federico II, 2021. English. NNT: 2021GRALV016 . tel-03267584

**HAL Id: tel-03267584**

**<https://theses.hal.science/tel-03267584>**

Submitted on 22 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES**

**préparée dans le cadre d'une cotutelle entre  
l'Université Grenoble Alpes et Università degli  
Studi di Napoli Federico II**

Spécialité : **Chimie Biologie et Scienze Chimiche**

Arrêté ministériel : le 6 janvier 2005 – 25 mai 2016

Présentée par

**Meriem MAALEJ**

Thèse dirigée par **Jean-Pierre SIMORRE** et **Antonio MOLINARO**  
et co-encadrée par **Cédric LAGURI** et **Roberta MARCHETTI**

préparée au sein des **Laboratoires : Institut de Biologie  
Structurale et Dipartimento di Scienze Chimiche** dans les **Écoles  
Doctorales : Chimie et Sciences du Vivant et Scienze Chimiche**

**Deciphering the recognition patterns  
between Lipopolysaccharides and human  
lectins using Nuclear Magnetic  
Resonance spectroscopy.**

Thèse soutenue publiquement le **16 Avril 2021**,  
devant le jury composé de :

**Madame, Nicole, THIELENS**

DIRECTRICE DE RECHERCHE, UMR5075 – Institut de Biologie Structurale,  
Présidente du jury

**Madame, Delia, PICONE**

PROFESSOR, Department of Chemical Sciences, UNINA, Examinatrice

**Madame, Latifa, ELANTAK**

CHARGEÉE DE RECHERCHE, IMM/CNRS, Rapportrice

**Madame, Laura, RUSSO**

ASSISTANT PROFESSOR, UNIMIB, Rapportrice

**Monsieur, Antonio, MOLINARO**

PROFESSEUR, Department of Chemical Sciences, UNINA, Directeur de thèse

**Monsieur, Jean-Pierre, SIMORRE**

DIRECTEUR DE RECHERCHE, UMR5075 – Institut de Biologie Structurale,  
Directeur de thèse









## Acknowledgements

Being engaged in this collaborative PhD work was exciting and laborious, at once. I gained interesting knowledges on many levels, from humanitarian to professional and scientific.

I would like to thank Pr. Jean-Pierre Simorre and Pr. Tony Molinaro for hosting me in their teams. They welcomed me in Grenoble and in Naples and, they gave me valuable support and assistance. I would like to thank everyone who has helped me with preparing the thesis manuscript. My gratitude goes to Dr. Cédric Laguri who passionately showed me many structural biology tips and helped me with wet lab as well and, I would like to thank Dr. Roberta Marchetti too for her advices and help with wet lab.

I would like to thank all jury members from both French and Italian sides, for their involvement and availabilities for the examination of my dissertation.

I would like to address my thankfulness to our collaborators Pr. Franck Fieschi and Dr. Michel Thepault for their important contribution by providing the requested samples. Many thanks to their PhD, François Bulteau as well.

My gratitude is directed to our collaborators from the cell-imaging platform at IBS, Dr. Jean-Philippe Kleman and Dr. Françoise Lacroix for their help and training.

A lot of thanks to Pr. Anne Imberty and Pr. Nicole Thielens for being the jury of my CST.

I address my gratitude also to the Italian PhD coordinator Pr. Angela Lombardi, to Pr. Delia Picone, to Pr. Alba Silipo and to Dr. Flaviana Di Lorenzo as well for the occasional conversations in Napoli.

I would like to also thank Pr. Catherine Bougault for her advices and instructions regarding chemistry wet lab, and I have special appreciation to the courses given by Pr. Dominique Marion and Dr. Paul Schanda in Grenoble. Many thanks to Dr. Adrien Favier, Dr. Alicia Vallet, Dr. Isabel Ayala, and Dr. Lionel Imbert from IBS for their help whenever I have requested it.

Finally, I own an immense gratitude to all teams' members from both French and Italian laboratories, Cristina, Rosa, Clara, Mateusz, Rim, Undina, Rida, Ricarda, Tiago, Nina, Iva, Faustine, Elisa.. thank you for the amazing times that we have spent together.

I shall thank GlycoAlps network and the Ministry of scientific research in Grenoble and Naples for the financial support of the thesis project. I want also to acknowledge every person from the institute of structural biology in Grenoble (IBS), and from the department of chemical sciences in Naples.

## Remerciement

*Je remercie infiniment mes parents Samira et Sayed qui m'ont toujours encouragé à faire des études supérieures, mes frères et sœurs Mohamed, Ahmed, Sana et Sarah, et leurs petites familles, vous êtes très précieux et j'espère que le bon Dieu nous réunira bientôt tous ensemble. Je vous aime tellement...*

*Je veux également remercier mes amis en Tunisie, en France et en Italie, ou plutôt ma deuxième famille à l'étranger : Fatma, tu es plus qu'une simple amie merci beaucoup pour ton aide et ta présence pendant les moments difficiles, Eya, tu es la meilleure merci beaucoup pour tous, Ines, merci beaucoup tu es plus qu'une amie, Karina, mi manchi tantissimo e spero di vederti presto, Clément, tu étais toujours là pour moi lorsque j'en avais besoin, merci beaucoup.*

*Enfin, je souhaite dédier mon travail à ma grand-mère Jamila décédée au cours de cette thèse, que Dieu te bénisse, tu m'as appris tellement de choses, tes valeurs resteront à jamais gravés dans ma mémoire.*





# Table of contents

Acknowledgements .....	2
Remerciement .....	3
Abstract .....	9
Résumé.....	11
Riassunto .....	13
Liste of figures.....	15
Liste of Tables .....	23
<b>I. Bacterial carbohydrates are Pathogen Associated Molecular Patterns.....</b>	<b>25</b>
I-1. The bacterial cell envelope PAMPS.....	25
I-1-1. Peptidoglycan and Teichoic acids (function and biological activity).....	27
I-1-2. Extracellular polysaccharides (capsular and biofilms) .....	28
<b>I-1-2-1. Role of capsular polysaccharides in bacteria resistance .....</b>	<b>28</b>
<b>I-1-2-2. Polysaccharides of bacterial biofilm are PAMPs.....</b>	<b>29</b>
I-1-3. Other glycolipids .....	31
<b>I-1-3-1. <i>Mycobacteria</i> infection and role of its PAMPs .....</b>	<b>31</b>
<b>I-1-3-2. Bacterial Lipopolysaccharide LPS, the hallmark of bacterial infections .....</b>	<b>33</b>
I-2. Bacteria survival role and molecular diversity of LPS .....	37
<b>I-2-1. LPS heterogeneous structure.....</b>	<b>38</b>
<b>I-2-2. An example of LPS/LOS from pathogenic bacteria .....</b>	<b>44</b>
I-3. Elicitation of host immune response by Lipopolysaccharide.....	45
<b>I-3-1. Innate and adaptive immunity .....</b>	<b>45</b>
<b>I-3-2. Role of LPS in the outbreak of bacterial infection .....</b>	<b>48</b>
<b>I-3-3. Recognition of LPS by the host cell.....</b>	<b>49</b>
<b>II. C-type Lectin Receptors .....</b>	<b>56</b>
II-1. Introduction .....	56
<b>II-1-1. Molecular organisation of C-type lectins: the ECD and CRD domains .....</b>	<b>58</b>
<b>II-1-2. C-type lectins of interest: structural and functional aspects.....</b>	<b>62</b>
<b>II-2-2. Other members of C-type lectins family .....</b>	<b>69</b>
II-3. Applications for the development of sugar-based therapeutics .....	72
<b>III. Structural and functional analyses of carbohydrates ligands and their interactions with lectins ....</b>	<b>77</b>
III-1. Structural elucidation of LPS and LOS molecules.....	77
<b>III-1-1. LPS Extraction methodologies .....</b>	<b>78</b>
<b>III-1-2. Chemical treatments of LPS molecules .....</b>	<b>80</b>
<b>III-1-3. Common methods for LPS structure elucidation .....</b>	<b>84</b>

III-2. Methodologies for studying carbohydrate-lectin interactions.....	97
<b>III-2-1. Binding essays as useful tools to derive interaction information</b> .....	98
<b>III-2-2. Imaging approaches</b> .....	101
<b>III-2-3. NMR spectroscopy for interaction studies</b> .....	105
<b>III-2-4 Other methods to analyze molecular interactions</b> .....	111
<b>IV. LPS-C-type lectins interactions</b> .....	<b>114</b>
IV-1. Examples of up-to-date studies of LPS-lectins interactions.....	114
<b>IV-1-1. Mannose Binding Lectin-LPS interaction</b> .....	114
<b>IV-1-2. DC-SIGN lectin targets bacterial LPS</b> .....	115
<b>IV-1-3. MGL lectin recognizes <i>Campylobacter jejuni</i> LOS</b> .....	116
IV-2. LPS glycoconjugates of interest .....	117
<b>IV-2-1. <i>E. coli</i> Sakai LPS</b> .....	117
<b>IV-2-2. LOSs from <i>E. coli</i> R1 and R3 strains</b> .....	118
<b>IV-2-3. LOS from <i>BC J2315</i> pathogens</b> .....	120
<b>V. Objectives and framework of the PhD project</b> .....	<b>123</b>
<b>VI. Isolation and characterization of LPS ligands for ensuing interactions studies</b> .....	<b>127</b>
VI-1. Results and discussions.....	127
<b>VI-1-1. Extraction and purification of LOSs and LPSs glycoconjugates</b> .....	127
<b>V-1-2. Liquid and solid-state NMR for identification of LPS/LOS residues</b> .....	138
VI-2. Conclusion.....	143
<b>VII. Selective screening of an interactive LPS-Lectin system</b> .....	<b>147</b>
VII-1. Results and discussions.....	147
<b>VII-1-1. Preliminary results of the interaction between intact LPS and the human lectins</b> .....	147
<b>VII-1-2. Interaction between CRDs of human lectins and <i>E. coli</i> OSs</b> .....	150
VII-2 Conclusions .....	152
<b>VIII. Does the human MGL CRD domain interacts with <i>E. coli</i> ligands differently?</b> .....	<b>155</b>
VIII-1. Results and discussions.....	155
<b>VIII-1-1. Backbone assignment of human MGL CRD by NMR spectroscopy</b> .....	155
<b>VIII-1-2 Human MGL interaction profiles with <i>E. coli</i> LPSs by NMR titrations</b> .....	157
VIII-2. Conclusion.....	165
<b>IX. Investigations of the large interaction system offer new findings</b> .....	<b>168</b>
IX-1. Results and discussions .....	168
<b>IX-1-1. Results from ECD-MGL interaction with <i>E. coli</i> OSs through mono-dimensional STD-NMR</b> .....	168
<b>IX-1-2 Bio-Layer Interferometry confirmed MGL binding to the isolated LOS R1</b> .....	177

IX-1-3. Disaggregation of <i>E. coli</i> LOSs vesicles in presence of ECD MGL.....	181
IX-1-4. Human MGL tracking in <i>E. coli</i> bacterium, LOS R1 producer.....	185
IX-2. Conclusions .....	192
<b>Published article.....</b>	<b>194</b>
<b>X. Conclusions and perspectives .....</b>	<b>196</b>
X-1. Conclusions .....	196
<b>X-1-1. Qualitative control of the isolated glycoconjugates .....</b>	<b>196</b>
<b>X-1-2. The CRD of human MGL harbors a potential secondary binding surface for <i>E. coli</i> glycoconjugates .....</b>	<b>197</b>
<b>X-1-3. The ECD of MGL exhibits high affinity for <i>E. coli</i> R1 glycoconjugates .....</b>	<b>197</b>
<b>X-1-4. Proposed Gram-negative membrane models: ECD and CRD MGL binding modes.....</b>	<b>201</b>
X-2. Perspectives .....	203
<b>X-2-1. Mutagenesis studies as ensuing analyses to confirm binding sites on CRD MGL.....</b>	<b>203</b>
<b>X-2-2. Further investigations of the presumable ECD MGL conformation changes.....</b>	<b>204</b>
<b>X-2-3. Further molecular investigations on LOS R1 for growth phase dependency validation ....</b>	<b>205</b>
<b>X-2-4. Immunologic tests as perspectives to study immunity stimulation by <i>E. coli</i> R1.....</b>	<b>205</b>
<b>X-2-5. Sugar-based drugs development as long-term perspectives .....</b>	<b>206</b>
<b>XI. Materials and methods.....</b>	<b>209</b>
XI-1. Production, purification, and chemical analyses of LPSs samples .....	209
<b>XI-1-1. Bacterial growth of <i>E. coli</i> R1 and R3 and <i>B. cenocepacia</i> J2315.....</b>	<b>209</b>
<b>XI-1-2. PCP extraction .....</b>	<b>209</b>
<b>XI-1-3. Hot-phenol water extraction .....</b>	<b>210</b>
<b>XI-1-4. DOC-PAGE analysis.....</b>	<b>210</b>
<b>XI-1-5. Methyl Glycoside and GC-MS analyses .....</b>	<b>210</b>
<b>XI-1-6. Delipidation and chromatography purification of LOSs .....</b>	<b>211</b>
<b>XI-1-6-1. LOS de-<i>O</i>,de-<i>N</i>-acetylation and purification for <i>E. coli</i> glycoconjugates .....</b>	<b>211</b>
<b>XI-1-7. Solid state NMR experiments on LPS O157:H7 .....</b>	<b>212</b>
XI-2. Production of CRD and ECD of human lectins.....	212
XI-3. Interaction studies .....	213
<b>XI-3-1. Buffers .....</b>	<b>213</b>
<b>XI-3-2. NMR spectroscopy .....</b>	<b>214</b>
<b>XI-3-3. BLI experiments .....</b>	<b>215</b>
<b>XI-3-4. Electron microscopy.....</b>	<b>216</b>
<b>XI-3-5. Fluorescence Microscopy and Flow Cytometry .....</b>	<b>217</b>
XI-4. Technical acknowledgments.....	218

**References** ..... 220  
**Annexes** ..... 244

## Abstract

For an equilibrated survival of humans against bacterial infections, immune cells have a well-defined function to distinguish between self and non-self structures. *Escherichia coli* is a Gram-negative bacterium that reside in the gut microbiota, either as a beneficial or harmful microorganism.

The tolerance of pathogenic strains or their clearance is covered by many immune cell receptors such as Pathogen Recognition Receptors (PRRs) including C-type lectin receptors (**CLRs**) that specifically interact with carbohydrates moieties. Lipopolysaccharide (**LPS**), expressed on the surface of bacteria, presents one of the signatures of Gram-negative bacteria by being both a structural and recognition motif by the host cell. The molecular diversity of both LPS (i.e. various carbohydrates, lengths and heterogeneity levels) and lectins (variable binding sites architectures) would confer them the particularity of controlling variable situations and targeting specific interactions. The understanding of the mechanisms of LPS-Lectins interactions is challenging and requires an interdisciplinary approach. On the above basis, we sought to investigate the interaction between a C-type lectin i.e. human Macrophage Galactose-type Lectin **MGL** and LPSs (from *E. coli* **R1**, **R3** mutants and *E. coli* **O157:H7** strain). Bacterial LPSs were extracted, purified, and thereafter considered for the investigation of this large and complex interaction system. The difficult experimental handling of such native biomolecules directed the use of a divided set of approaches. In fact, our scientific strategy includes the use of Nuclear Magnetic Resonance NMR spectroscopy, fluorescence microscopy, and molecular binding essays. By combining these methods, we studied two distinct Lectin-LPS interaction systems: i) the molecular interaction between the Carbohydrate Recognition Domain (**CRD**) of MGL and soluble LPS versions by using NMR titrations and computational methods; ii) the recognition of *E. coli* mutants and LPS glycoconjugates by the Extracellular Domain (**ECD**) of MGL at both molecular and cellular levels by STD-NMR combined with computational analyses, Biolayer Interferometry (BLI), Electron Microscopy (EM) and fluorescence microscopy coupled to flow Cytometry.

When only the CRD of MGL is considered, all tested *E. coli* LPS were found to bind in the millimolar affinity range, to an extended interacting area that includes a **putative secondary binding** region, in addition to the canonical calcium binding site, common in C-type lectins.

Spectroscopic and microscopic investigations provided promising results about the recognition between ECD MGL and *E. coli* LPS/LOS. Trimeric human MGL interacts specifically with *E. coli* **R1**



**LOS** mainly through binding to the terminal di-galactose moiety. In addition, the dissociation constant ( $K_d$ ) was estimated to be in the nanomolar range thus indicating a **strong molecular binding**. Moreover, the specific interaction between fluorescently labelled human MGL and *E. coli* R1 bacteria was investigated by single-cell essays. **MGL-bound *E. coli* R1** bacteria were fluorescent (up to 40% of the bacterial population), exclusively at exponential phase, whereas *E. coli* R3 were not. Here again, binding specificity and selectivity was confirmed for ECD MGL-R1 interaction system.

We showed that MGL strongly and specifically interacts with *E. coli* R1 glycoconjugates at the surface of bacteria through its terminal di-galactose motif. The contribution of a putative secondary binding site on MGL in glycoconjugates recognition remains to be investigated. This PhD work showed that, despite the difficulties that such large system studies may encounter, many findings are attainable by using our scientific strategy. The possibility of investigating data from atomic to cellular scale on LPS-lectin interactions in either modified or native states, opens prominent horizons for the study of bacterial infections.

**Key words:** Gram-negative bacteria, bacterial infections, LPS, human MGL, molecular interactions

## Résumé

Pour que les humains aient une survie stable contre les infections bactériennes, les cellules immunitaires ont une fonction bien définie qui leur permet de distinguer les antigènes du soi et du non-soi. *Escherichia coli* est une bactérie à Gram-négative qui réside dans le microbiote humain en étant un microorganisme bénéfique mais qui peut être aussi nocif (pathogène).

La tolérance des souches pathogène ou leur élimination est assurée par les récepteurs des cellules immunitaires tels que les Pathogen Recognition Receptors (PRRs) incluant les lectines du type-C ou C-type lectin receptors (CLRs). Ces dernières interagissent d'une manière spécifique avec les carbohydrates. Le Lipopolysaccharide (LPS) exprimé à la surface des bactéries, présente une des signatures des bactéries à Gram-négatives en étant à la fois un motif de structure et de reconnaissance par la cellule hôte. La diversité structurale et moléculaire des LPS (i.e. composition variable en carbohydrates avec différentes ramifications et niveaux d'hétérogénéités) et des lectines humaines (sites de liaison avec différentes architectures) leur confère la particularité de contrôler des situations variables en ciblant des interactions spécifiques. Ainsi, comprendre les mécanismes des interactions LPS-Lectine est un défi et ceci exige une approche scientifique multidisciplinaire.

Sur cette base, nous avons cherché à étudier l'interaction entre la lectine humaine du type-C i.e. Macrophage Galactose-type Lectine ou MGL, et, les LPSs (des mutants *E. coli* R1, R3 et la souche O157:H7). Les LPSs ont été tout d'abord extraits, purifiés et ensuite considérés pour les analyses de ce système d'interaction large et complexe. La manipulation expérimentale de ces biomolécules natives est difficile et exige l'utilisation d'un ensemble divisé d'approches. En effet, notre stratégie scientifique comprend l'utilisation de la Résonance Magnétique Nucléaire RMN, la microscopie à fluorescence and les tests d'interaction moléculaire. En combinant ces méthodes, nous avons étudié séparément deux systèmes d'interaction LPS-Lectine : i) l'interaction moléculaire entre le domaine Carbohydrate Recognition Domain (CRD) de la MGL humaine, et, plusieurs versions solubles des LPS bactériens moyennant des titrations par la RMN et des études computationnels; ii) La reconnaissance des épitopes des LPS, sous forme d'assemblages membranaire ou directement sur la surface des bactéries *E. coli*, par le domaine extracellulaire Extracellular Domain (ECD) de la MGL, au niveau moléculaire et aussi cellulaire, à travers la STD-RMN combinée à des analyses computationnel, à l'interférométrie (BLI), à la

microscopie électronique (EM) et enfin la microscopie à fluorescence couplée à la cyrtométrie en flux.

Lorsque le CRD de la MGL est considéré, tous les LPSs exprimés par *E. coli* ont manifesté des interactions, d'une affinité de liaison de l'ordre du millimolaire, sur une surface élargie incluant un second site putatif, en plus du premier site calcique commun aux lectines type-C.

Nos études spectroscopiques et microscopiques ont fourni des résultats prometteurs sur la reconnaissance des LPS d'*E. coli* par le domaine ECD de la MGL humaine. La MGL trimérique interagit d'une manière spécifique avec le LOS, une structure courte du LPS exprimé chez la souche *E. coli* R1, principalement à travers la liaison à l'épitope terminal du LOS constitué d'un di-galactose. De plus, la constante de dissociation (Kd) a été estimée de l'ordre du nanomolaire ainsi indiquant une forte interaction moléculaire. En outre, la spécificité de cette interaction a été encore une fois confirmée par l'étude de la MGL marqué par un fluorophore où les bactéries *E. coli* R1 ont été visualisée, sous le microscope à epi-fluorescence, en temps réel. Les bactéries *E. coli* R1 liées à la MGL étaient fluorescentes (jusqu'à 40% de la population bactérienne), exclusivement en phase exponentielle, tandis que les bactéries *E. coli* R3 ne manifestèrent aucune luminescence. Là encore, le choix de la MGL d'interagir d'une manière sélective et spécifique était orienté vers les glycoconjugués d'*E. coli* R1.

Nous avons démontré que la MGL humaine (ECD) interagit fortement et spécifiquement avec les glycoconjugués d'*E. coli* R1 à la surface bactérienne à travers les motifs di-galactoses des LOSs. La contribution du second site de liaison sur la MGL (le CRD) dans la reconnaissance des glycoconjugués d'*E. coli* reste à analyser minutieusement. Ce travail de doctorat a finalement montré que, malgré la complexité du système d'interaction, il est possible d'étudier la reconnaissance des agents pathogènes, aboutissant à des résultats prometteurs, moyennant notre combinaison de techniques scientifiques. La possibilité d'analyser les données brutes obtenus sur les interactions LPS-Lectine (en état natif ou décomposé), de l'échelle atomique à l'échelle cellulaire, permettrait d'ouvrir de nouveaux horizons pour l'étude des infections bactériennes.

**Mots clés:** Bactérie à Gram-négative, infections bactériennes, Lipopolysaccharide LPS, lectine humaine MGL, interactions moléculaires.

## Riassunto

Per una sopravvivenza equilibrata degli esseri umani contro le infezioni batteriche, le cellule immunitarie hanno una funzione ben definita per distinguere tra strutture del sé e non del sé. L'*Escherichia coli* è un batterio Gram-negativo che risiede nel microbiota intestinale, sia come microrganismo benefico che dannoso.

La tolleranza dei ceppi patogeni o la loro eliminazione è coperta da molti recettori delle cellule immunitarie come i recettori di riconoscimento dei patogeni (PRR), inclusi i recettori della lectina di tipo C (CLR) che interagiscono specificamente con le frazioni di carboidrati. Il lipopolisaccaride batterico (LPS) è una delle firme batteriche ed è il segno distintivo dei batteri Gram-negativi. La diversità molecolare sia di LPS (cioè vari carboidrati, lunghezze ed eterogeneità livelli) che di lectine (architetture di siti di legame variabili) conferirebbe loro la particolarità di controllare situazioni variabili e mirare a interazioni specifiche. La comprensione dei meccanismi delle interazioni LPS-Lectine è impegnativa e richiede un approccio interdisciplinare.

Sulla base di cui sopra, abbiamo cercato di indagare l'interazione tra una lectina di tipo C, cioè lectina di tipo galattosio macrofago umano MGL e LPS (da *E. coli* mutanti R1, R3 e ceppo O157:H7). Gli LPS batterici sono stati estratti, purificati e successivamente considerati per lo studio di questo sistema di interazione ampio e complesso. La difficile manipolazione sperimentale di tali biomolecole native ha diretto l'uso di una serie divisa di approcci. In effetti, la nostra strategia scientifica include l'uso della spettroscopia NMR di risonanza magnetica nucleare, microscopia a fluorescenza e saggi di legame molecolare. Combinando questi metodi, abbiamo studiato due distinti sistemi di interazione Lectina-LPS: i) l'interazione molecolare tra il Carbohydrate Recognition Domain (CRD) di MGL e le versioni di LPS solubili utilizzando titolazioni NMR e metodi computazionali; ii) il riconoscimento di mutanti di *E. coli* e glicoconiugati LPS da parte del dominio extracellulare (ECD) di MGL sia a livello molecolare che cellulare mediante STD-NMR combinato con analisi computazionali, Biolayer Interferometry (BLI), Microscopia elettronica (EM) e microscopia a fluorescenza accoppiato alla citometria a flusso.

Quando si considera solo il CRD di MGL, tutti gli LPS di *E. coli* testati sono risultati legarsi nell'intervallo di affinità millimolare, a un'area di interazione estesa che include una regione di legame secondaria putativa, oltre al sito di legame canonico del calcio, comune in C tipo lectine.

Spettroscopiche e le indagini microscopiche condizionate promettenti risultati riguardo il riconoscimento tra ECD MGL ed *E. coli* LPS / LOS. Trimerico umano MGL interagisce specificamente con *E. coli* R1 LOS principalmente attraverso il legame alla porzione terminale di-galattosio. Inoltre, la costante di dissociazione ( $K_d$ ) è stata stimata nell'intervallo nanomolare, indicando così un forte legame molecolare. Inoltre, l'interazione specifica tra MGL umano marcato in modo fluorescente e batteri *E. coli* R1 è stata studiata mediante saggi unicellulari. I batteri *E. coli* R1 legati a MGL erano fluorescenti (fino al 40% della popolazione batterica), esclusivamente in fase esponenziale, mentre *E. coli* R3 non lo erano. Anche in questo caso, la specificità e la selettività del legame sono state confermate per il sistema di interazione ECD MGL-R1.

Abbiamo dimostrato che MGL interagisce fortemente e specificamente con i glicoconiugati R1 di *E. coli* sulla superficie dei batteri attraverso il suo motivo terminale di-galattosio. Resta da studiare il contributo di un putativo sito di legame secondario su MGL nel riconoscimento dei glicoconiugati. Questo lavoro di dottorato ha dimostrato che, nonostante le difficoltà che possono incontrare studi di sistema così ampi, molti risultati sono ottenibili utilizzando la nostra strategia scientifica. La possibilità di indagare i dati dalla scala atomica a quella cellulare sulle interazioni LPS-lectina negli stati modificati o nativi, apre importanti orizzonti per lo studio delle infezioni batteriche.

**Parole chiave** : batteri Gram-negativi, infezioni batteriche, lipopolisaccaride LPS, MGL letina umano, interazioni molecolari.

## Liste of figures

<b>Figure 1. Schematic representation of Gram-positive and Gram-negative bacteria with their associated PAMPs.</b> PNG: peptidoglycans, TA: teichoic acid, LTA: lipoteichoic acid, CPS: capsular polysaccharide, PM: periplasmic membrane, OM: outer membrane. ....	26
<b>Figure 2. PNG structure and its interactions with NOD 1 and NOD 2 across the cell wall of Gram-negative bacteria.</b> Adapted from <sup>5</sup> .....	27
<b>Figure 3. Schematic representation of Reg III lectin mode of action for Gram-positive bacteria lysis.</b> RegIII $\alpha$ first interacts with peptidoglycan via carbohydrate binding engendering then hexameric proteins cluster in the bacterial membrane. Adapted from <sup>19</sup> .....	28
<b>Figure 4. Biofilm formation and host type I interferon (IFN) innate immune response activation through sensing of biofilm-specific PAMPs i.e. c-di-NMPs.</b> Taken from <sup>36</sup> .....	30
<b>Figure 5. Main components of mycobacteria cell wall.</b> .....	31
<b>Figure 6. Common structure of Enterobacterial LPS gathering different types of sugars.</b> Sugar residues are presented following the symbolic representation with the rules set-forward from Symbol Nomenclature for Glycans (SNFG) <sup>67</sup> which compose the O-Antigen and the core oligosaccharide, covalently attached to Lipid-A moiety that is essentially composed of fatty acids and 2 glucosamines. Linkages of sugars shown in this scheme are for diagrammatic purposes only. ....	34
<b>Figure 7. Biosynthesis steps of LOS glycoconjugate, from E. coli R1 bacteria followed by its transport to the OM.</b> Lipid-A-KDO <sub>2</sub> stepwise biosynthesis is insured by (LpxA,C,D,H,B,K,L,M) enzymes machinery followed by core construction. Glycosyltransferases needed for Inner Membrane (IM) and Outer Membrane (OM) assembly are colored in purple and green, respectively. In red are labelled the enzymes required for inner core modifications. For clarity, dashed arrow is pointed to further steps for LOS transport and the blue glycosyltransferase could intervene in by O-Antigen addition in the periplasm, in case of strain mutation. Inspired from <sup>75,76</sup> .....	35
<b>Figure 8. Schematic representation of LPS transport across the cell envelope and assembly at the OM.</b> For clarity, green rectangles refer to enzymes involved in LOS and LPS biosynthesis in the OM and IM, respectively; and proteins involved in LPS/LOS transport are depicted in red circles (i.e. Lpt transport protein complexes and MsbA). One of the recently published X-ray structures of MsbA is presented. Adapted from <sup>81</sup> .....	36
<b>Figure 9. E. coli LPS transport Lpt machinery.</b> Structure of the LPS molecule is detailed on the right part of the figure to show each part of LPS interacting with the transporters (i.e. LptC, LptA) while being translocated. For clarity, n refers to the number of LptA proteins, and distance between IM and OM is indicated in nanometers nm. Taken from <sup>84</sup> .....	37
<b>Figure 10. LPS (A) and mixture of LPS and LOS molecules (B) assemblies in the lipid bilayer constructed using MD simulations.</b> Calcium ions are depicted in yellow for clarity. Adapted from <sup>93</sup> .....	39
<b>Figure 11. Chemical structure of Lipid-A from Escherichia. coli and Bulkholderia Genus.</b> Taken from <sup>95</sup> . ....	40
<b>Figure 12. General presentation of E. coli K12 core oligosaccharide.</b> Dashed arrows highlight all possible substitutions.....	42
<b>Figure 13. Chemical structures of 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo), L-glycero-D-manno-heptose (L,D-Hep), and D-glycero-D-manno-heptose (D,D-Hep), the features of the enterobacterial inner cores.</b> Taken from <sup>108</sup> .....	42
<b>Figure 14. Representation of inner core structure distribution in E. coli and Salmonella.</b> Dotted lines indicate the linkages formed by specific gene products and dashed arrows refer to all possible	

substitutions among mentioned strains. Ethanolamine phosphate (EtnP) aminoethyl diphosphate (PPEtn) and phosphate (P) are phosphate groups branched to the inner core. Taken from <sup>75</sup> .	43
<b>Figure 15. Innate and adaptive immune systems facing microbial PAMPs by communicating through immune cells.</b> Taken from <sup>120</sup> .	46
<b>Figure 16. Triggering of the immune response by bacterial infection.</b> Macrophages encounter pathogen invaders and from that step on, a cascade of intercellular events implicates the immune system including secretion of chemokines and cytokines and changes in the stickiness of the endothelial cells of the blood vessels necessary for immune cells migration. Adapted from <sup>119</sup> .	47
<b>Figure 17. TLR-4 activation by LPS and intracellular signalling pathways involving several effectors.</b> Taken from <sup>133</sup> .	50
<b>Figure 18. (TLR-4/MD-2/LPS)<sub>2</sub> complex structure showing the primary and dimerization interfaces.</b> Pdb: 3FXI .	51
<b>Figure 19. Schematic domain structures of the cytosolic LPS receptors Caspase-4, Caspase-5 and Caspase-11.</b> Taken from <sup>147</sup> .	53
<b>Figure 20. Non-canonical inflammasome formation in humans and mice.</b> (a) and (b) show molecular bindings from single mouse and human caspases, respectively, to the Lipid-A portion of LPS. The oligomerization of complexes is displayed in (c). p20 and P10 refer to catalytic domains of caspases. Taken from <sup>148</sup> .	53
<b>Figure 21. Graphical view of the different interactions mediated by lectins expressed on the surface of dendritic cells, that are cell-cell adhesion and pathogen recognition.</b> Taken from <sup>151</sup> .	56
<b>Figure 22. Cartoon representation of the distinct domains of some CLR's round the immune cell membrane showing direct and indirect interactions with signalling effectors during innate immune response to a pathogen.</b> A frame is added to the original figure, showing CLR's are the C-type lectins studied within this research work. Taken from <sup>155</sup> .	58
<b>Figure 23. General organisation of C-type lectin receptor on the membrane of myeloid cell.</b> Adapted from pdb 1RTM .	59
<b>Figure 24. An example of the entire structure of human C-type lectin ECD, P-selectin.</b> P-selectin constructs, derived from X-ray data, include CRD or lectin domain (B) (residues close to Lewis <sup>x</sup> binding site are depicted from P-selectin superimposed with E-selectin), epidermal growth factor-like (EGF) domain and transmembrane domain (A). Calcium ion is highlighted in yellow. For clarity, P-selectin ribbon representation (blue) shows overall similarity with its homologous E-selectin (green). Taken from <sup>159</sup> .	59
<b>Figure 25. Generic CRD or CTLD domain of animal lectins.</b> Black and white representations, middle and left structures, display examples of CRDs showing Ca <sup>2+</sup> binding sites locations in human ASGPR and mice MBP-A with their respective PDBs. For clarity, the four cysteines are presented in atomic presentation. Secondary structural elements are described in the text. Taken from <sup>161</sup> .	61
<b>Figure 26. Graphical overview of the recognition of pathogens by the two CLR's of interest and downstream stimulation of immune response.</b>	62
<b>Figure 27. Organization of the upper part of human ECD DC-SIGN and DC-SIGNR.</b> Arrangements of CRDs from the two lectins were derived from SAXS and X-ray data of one DC-SIGNR fragment (pdb 1XAR). Taken from <sup>168</sup> .	63
<b>Figure 28. View of the upper surface of CRD DC-SIGN in Apo state and the domain organisation of the ECD.</b> Four residues (Glu347, Asn349, Glu354 and Asn365) that bind a calcium ion from the second binding site, are responsible for carbohydrate local recruitment. Taken from <sup>170</sup> .	64
<b>Figure 29. Complex formation following DC-SIGN CRD binding to an oligosaccharide containing Lewis<sup>x</sup> determinant at the nonreducing end (LNFPIII).</b> Carbohydrates are represented by universal sugars codes and DC-SIGN CRD 3 structure is displayed in cartoon representation. PDB: 1SL5 .	65

<b>Figure 30. Super-resolution images and density ratios of DC-SIGN and MR lectins during interaction with microbial PAMPs.</b> (A) cell surface densities of DC-SIGN and MR CD206 (white box) in proximity of fugal contact site (green circle) and non-contact site (yellow box) and (B) the density ratio of both lectins (labelled with antibodies conjugated with AlexaFluor <sup>647</sup> ) bound to glycan motifs of the zymosan particles over their densities at non-contact site. Taken from <sup>181</sup> .....	66
<b>Figure 31. Human CRD MGL crystallographic structure before (A) and after (B) complex formation with GalNAc residue.</b> Respective PDBs: 6PUV and 6PY1.....	68
<b>Figure 32. Similarities in sugar preferences of DC-SIGN and Langerin.</b> Taken from <sup>216</sup> .....	71
<b>Figure 33. Schematic representation of the proposed model of interaction between cow Mincle and Trehalose TDM from mycobacteria.</b> Taken from <sup>57</sup> .....	71
<b>Figure 34. Strategy used for glycomimetic development as human lectins inhibitors.</b> The X-ray structure of DC-SIGN (pdb 2XR5) bound to the final second-generation inhibitor is displayed at the bottom. Adapted from <sup>227</sup> .....	73
<b>Figure 35. Chemical structure of selectin (P-selectin, PDB 1G1R) bound to sle<sup>x</sup> and its inhibitors rivipansel and uproleselan.</b> Fucose sugar essential for the binding process is depicted in red. Adapted from <sup>227</sup> .....	74
<b>Figure 36. Examples of carbohydrates that could be found in human and in some LPS molecules.</b> The figure shows the different isomerism in sugars ring sizes (b), region-isomerism (c), anomeric configuration of the glycosidic linkage (d) and positions (e) that they may adopt in a given polysaccharide sequence. Up to sixteen different isomers could be formed by one monosaccharide. Taken from <sup>232</sup> .....	77
<b>Figure 37. Combined extraction methods strategy for LOS/LPS molecules purification.</b> Phenol/EDTA/TEA is another extraction protocol that can be inserted among the steps. Taken from <sup>235</sup> .....	79
<b>Figure 38. Elution order corresponding to each type of derivatized sugars in the GC-MS chromatogram.</b> Adapted from <sup>235</sup> .....	80
<b>Figure 39. Overview of the two main reactions for LPS transformation into acetylated O-methyl glycosides (MGA).</b> Taken from <a href="http://glycopedia.eu/">http://glycopedia.eu/</a> .....	81
<b>Figure 40. Preparation of octyl glycoside standard from D-glucose.</b> A mixture of diastereoisomers (D) is produced in the second reaction which have the same RT as the enantiomers (L). Taken from <sup>235</sup> .....	82
<b>Figure 41. Example of derivatization protocol used to deduce the monosaccharide sequence of a polysaccharide.</b> For clarity, the sugars that are shown correspond to a part of the polysaccharide molecule linked through glycosidic linkages (highlighted in blue arrows), and the values are indicative of the fragmentation pattern of each part of the residue (m/z). Adapted from <sup>235</sup> .....	82
<b>Figure 42. Schematic representation of cleavage sites established during the two chemical reactions used for solubilisation of LOSs molecules.</b> The acetic acid treatment occurs by breaking the glycosidic linkage between Kdo and the GlcN moiety of Lipid-A and De-O-de-N-acetylation reactions release N- and O-linked fatty acids.....	84
<b>Figure 43. An example of negative-ion MALDI-TOF mass spectrum of LPS from E. coli O164.</b> Taken from <sup>255</sup> .....	85
<b>Figure 44. Spin systems (A-E) of intra and inter-residues assignment of anomeric signals of a <sup>13</sup>C-enriched O-Antigen polysaccharides of E. coli O142 by HSQC and NOESY NMR spectra.</b> Taken from <sup>275</sup> .....	93
<b>Figure 45. General descriptions about the NMR parameters and inputs related to each of the classical experiments recorded for carbohydrates analyses.</b> Taken from <sup>273</sup> .....	94
<b>Figure 46. Example of detectable NOEs connectivities in a disaccharide by NMR spectroscopy.</b> Inter-NOEs contacts are colored in red and intra-NOEs in blue. ....	95



<b>Figure 47. The interaction between LPS O-Antigen by EM and ss-NMR INEPT experiments.</b> (a) EM images of LPS before and after addition of Gentamicin. (b) $^1\text{H}$ - $^{13}\text{C}$ NMR signals (i.e. O-Antigen B-Band, the most perturbed) are shown. Taken from <sup>268</sup> .	96
<b>Figure 48. Solid state <math>^{31}\text{P}</math> MAS NMR spectra from (top) the Outer Membrane of commensal <i>E. coli</i> and (bottom) LPS/DOPC mixtures, in presence and absence of nisin.</b> Adapted from <sup>278</sup> .	97
<b>Figure 49. DC-SIGN binding profiles from Glycan array and SPR experiments.</b> Top: Screening of DC-SIGN binding to various glycoconjugates ligands, at different concentrations of the lectin. Bottom: SPR results i.e., sensorgram (left) and binding curve (right) of LPS core terminal heptoside following glycan array analyses. Adapted from <sup>281</sup> .	99
<b>Figure 50. Typical titration curve from Me-<math>\alpha</math>-Fuc binding to PA-IIL showing thermodynamics parameters extracted from ITC data.</b> Adapted from <sup>286</sup> .	101
<b>Figure 51. Snapshot of stepwise membrane curvature formation following lectin binding to carbohydrate structures of GSLs</b> (a) and their cross-link to GSLs (b), generating plasma membrane tubules with many GSLs clusters (c). Taken from <sup>298</sup> .	104
<b>Figure 52. Use of two imaging approaches to study LPS interaction with Caspase4.</b> Visualisation of LPS/caspase4 and LPS/LBP complex formation at different molar ratios by negative staining using electron microscopy and the effect of C258A mutated caspase4 on the interaction with LPS (both biomolecules labelled with fluorophores) by fluorescence microscopy. They used different versions of mutated caspase4 including a (1-80 amino acids) truncated constructs ( $\Delta$ 80 caspase-4 (C258A)). Adapted from <sup>147</sup> .	105
<b>Figure 53. Illustration of intermolecular saturation transfer during protein-ligand binding in fast exchange regime.</b> 1D $^1\text{H}$ -STD experiments applied to the system and the derived STD (difference) spectrum are shown. Taken from <sup>303</sup> .	107
<b>Figure 54. Dependency of NMR cross-peaks shapes and trajectories to slow and fast exchange regimes during protein binding to the ligand.</b> Adapted from <sup>309</sup> .	110
<b>Figure 55. Binding interactions of terminal mannose oligosaccharide at the canonical binding site of rat MBL CRD.</b> For clarity, atoms are colored as following: Oxygen in black, Carbon in white, Nitrogen in grey and calcium in light grey. Adapted from <sup>320</sup> .	115
<b>Figure 56. Confocal microscopy images with fluorescently conjugated antibodies specific for TLR4 and DC-SIGN produced by HK-2 cells, before and after LPS addition.</b> Taken from <sup>326</sup> .	116
<b>Figure 57. Summarized <i>E. coli</i> O:157 LPS structures that have been resolved to date.</b>	118
<b>Figure 58. Structures of the R1 and R3 outer core OSs from the LPS/LOSs of <i>E. coli</i> and genetic organization of the central waaQ operon (core OS biosynthesis) loci.</b> The asterisks indicate the points of attachment of O-antigen in smooth-type LPS bacteria. Adapted from <sup>328</sup> .	120
<b>Figure 59. Core oligosaccharide structures after alkaline degradation of <i>E. coli</i> LOS R1 and R3 with the corresponding minor and major forms.</b> Taken from <sup>108</sup> .	128
<b>Figure 60. Chemical structure of the core oligosaccharide from BC J2315 strain.</b> Taken from <sup>259</sup> .	128
<b>Figure 61. Right panel: Different electrophoretic profiles of the extracted materials from <i>E. coli</i> and <i>B. cenocepacia</i> strains using the DOC-PAGE electrophoresis.</b> Left panel: PCP precipitate formation prior to LOS R1 and R3 extractions. Lanes <b>1</b> and <b>4</b> were loaded with LOS R1, <b>3</b> with LOS R3 and <b>2</b> with LOS from BC J2315 strain. Ref corresponds to the reference used as pure LPS isolated from <i>E. coli</i> O127:B8.	130
<b>Figure 62. CG spectra of acetylated methyl glycosides from LOS R1 and R3.</b> The reporter ions masses corresponding to distinctive sugar moieties are shown in the table.	131
<b>Figure 63. CG spectrum of acetylated methyl glycosides from the isolated LOS BC J2315.</b>	132

<b>Figure 64. GC and MS partial chromatograms of LOS BC J2315 zoomed region of Kdo and Ko retention time peaks and the occurrences of characteristic fragmentation patterns of Kdo corresponding to 153 and 375 m/z reporter ions.</b> .....	132
<b>Figure 65. Phenol and aqueous phases from PCP followed by hot-phenol water extraction of LPS O157:H7 and their corresponding gel profiles.</b> LPS from <i>E. coli</i> O127:B8 was used as reference. ....	134
<b>Figure 66. Examination of <sup>13</sup>C LPS O157:H7 purity through (A) DOC-PAGE analyses and (B) 600 MHz <sup>1</sup>H-NMR.</b> A) PCP and PCP/Hot-phenol water extracted LPS gel profiles, 1 mg/mL of the isolated <sup>13</sup> C-LPS were solubilised in H <sub>2</sub> O for electrophoresis (10 µg/lane). B) NMR spectra recorded at 298K of the isolated LPS samples prepared with different protocols and containing 5 mg of <sup>13</sup> C-LPS extracts suspended in D <sub>2</sub> O.135	
<b>Figure 67. 600 MHz <sup>1</sup>H-NMR spectra of 2.7 mM final concentration of OS J2315 fractions in D<sub>2</sub>O at 298K after P4 then P2 columns chromatography.</b> The four fractions resulted from the second chromatography showed different NMR profiles highlighted by the presence of well-defined peaks in the anomeric region (4.6-5.5 ppm) of fraction 1 with respect to fractions 2-4. ....	137
<b>Figure 68. 600-MHz-<sup>1</sup>H NMR spectrum of 2.3 mM final concentration of deacylated LOS from <i>E. coli</i> R1 and R3.</b> NMR samples were prepared in 25mM Tris-d <sub>11</sub> , 4mM CaCl <sub>2</sub> , 150 mM NaCl pH 7 in D <sub>2</sub> O (buffer A) and experiments were run at 298K. In the high field region of the spectrum 2 pairs of signals were observed originating from 3-deoxy protons with chemical shifts characteristic of Kdo residues. Signals of anomeric protons from the sugar sequence of the core oligosaccharide were also present at the region between 6.0-4.5ppm together with their ring protons between 3.0-4.3 ppm.....	138
<b>Figure 69. Zoom in the [<sup>1</sup>H-<sup>13</sup>C]-2D-HSQC spectrum of 5 mg of the isolated deacylated LOS R1.</b> The NMR sample was prepared in buffer A and experiments were run at 298 K. The most relevant heteronuclear correlations are reported. Correlations corresponding to CH <sub>2</sub> groups are colored in pink whereas CH groups are in blue. OS R1 structure is highlighted in the figure. ....	139
<b>Figure 70. 950 MHz MAS <sup>13</sup>C-CP-1D-NMR spectrum of 2 mg of <sup>13</sup>C LPS from <i>E. coli</i> O157:H7 strain.</b> The sample was prepared in 25 mM Tris 150 mM NaCl 4 mM CaCl <sub>2</sub> H <sub>2</sub> O (buffer B) and ss-NMR experiments were recorded at 270K and 25KHz.....	140
<b>Figure 71. 950MHz MAS [<sup>1</sup>H-<sup>13</sup>C]-hCCH-2D-INEPT spectrum (zoomed) of 2 mg of <sup>13</sup>C LPS from <i>E. coli</i> O157:H7 strain.</b> The sample was prepared in buffer B and ss-NMR experiments were registered at 270K and 50KHz.....	142
<b>Figure 72. 950MHz MAS [<sup>13</sup>C-<sup>13</sup>C]-hCC-DARR-NMR spectrum focused on sugars resonances of 2 mg of <sup>13</sup>C LPS from <i>E. coli</i> O157:H7 strain.</b> The sample was prepared in buffer B and ss-NMR experiments were recorded at 293-298K and 15KHz. For simplicity, the O-Antigen structure is shown using symbol nomenclatures from <sup>346</sup> and <sup>13</sup> C- <sup>13</sup> C correlations from lipids are not shown in this spectrum.....	143
<b>Figure 73. Sugar sequences within the isolated <i>E. coli</i> LPS scaffolds for further interaction studies with human lectins.</b> Symbol nomenclatures of sugar residues has been used from <sup>346</sup> . Red and blue squares refer to structural differences and similarities within the isolated LPS species, respectively. ....	144
<b>Figure 74. Disaggregation of <sup>13</sup>C-LPS O157:H7 in presence of human lectins including DC-SIGN observed in [<sup>1</sup>H-<sup>13</sup>C]-2D-HSQC NMR spectra (A) and TEM (B) graphs.</b> Electron microscopy micrographs are magnified to 23000x with 100 nm scale bar. ....	149
<b>Figure 75. Superimposition of <sup>1</sup>H-<sup>15</sup>N backbone resonances of human CRD DC-SIGN (blue) in presence of 20 eq. of OS R1 (green) and 20 eq. of OS R3 (pink) in buffer B at 303 K on the 850 MHz spectrometer.</b> .....	151
<b>Figure 76. Superimposition of [<sup>1</sup>H-<sup>15</sup>N] backbone resonances of human CRD MGL (blue) in presence of 20 eq of OS R1 (purple) and 20 eq of OS R3 (red) in buffer B at 303 K on the 850 MHz spectrometer.</b>	152
<b>Figure 77. (A) Graphical description for transfer of magnetization during a [<sup>1</sup>H-<sup>15</sup>N]-2D-Best-TROSY-HSQC experiment applied to proteins (B) 950 MHz [<sup>1</sup>H-<sup>15</sup>N]- BEST-TROSY-HSQC spectrum of human <sup>15</sup>N-</b>	

<b><sup>13</sup>C -CRD MGL in buffer B at 303K.</b> For clarity, A- and B- labels in residues nomenclatures (on the entire spectrum) refer to CRD MGL NMR chains.....	156
<b>Figure 78. Illustration of amino acid sequence backbone assignments of human MGL CRD.</b> The CRD MGL backbone assignment is overlaid with the one from Diniz et al <sup>350</sup> who have used different experimental conditions (lower temperature and higher concentrations of lectin and of Ca <sup>2+</sup> ions in the buffer). In our CRD MGL sequence, the 15 initial residues were added by our collaborators for the purification of the protein and for functional tests belonging to them.....	157
<b>Figure 79. 700 MHz (top) [<sup>1</sup>H, <sup>15</sup>N]-2D-Best-TROSY-HSQC superimposed spectra of 50 μM <sup>15</sup>N-CRD MGL at the different OS R1 molar ratios.</b> Backbone resonances are showed during stepwise addition of the ligand in buffer B at 303K.....	158
<b>Figure 80. 700MHz [<sup>1</sup>H, <sup>15</sup>N]-2D-Best-TROSY-HSQC superimposed spectra (zoomed at specific regions) of 50 μM <sup>15</sup>N-CRD MGL at the different OS R3 and OS R1 molar ratios.</b> Backbone resonances are showed during stepwise addition of the ligand in buffer B at 303K.....	159
<b>Figure 81. 850MHz [<sup>1</sup>H, <sup>15</sup>N]-2D-Best-TROSY-HSQC superimposed spectra of 50 μM <sup>15</sup>N-CRD MGL at the different LPS O157:H7 molar ratios.</b> Backbone resonances are showed during stepwise addition of the ligand in buffer B at 300K.....	160
<b>Figure 82. Chemical shift changes of human <sup>15</sup>N-CRD MGL active residues following LPS O157:H7 binding.</b> .....	161
<b>Figure 83. Representation of chemical shifts scatters of the four studied ligands over the backbone of human MGL CRD.</b> Data derived from NMR titrations considering single <sup>1</sup> H and <sup>15</sup> N chemical shift values for each residue, considering the corresponding threshold values (2*σ) delimited by the red line.....	163
<b>Figure 84. Potential binding surfaces following the interaction of E. coli glycoconjugates ligands deduced from NMR titrations with human CRD MGL.</b> Threshold values were calculated from two times standard deviation (σ) from <sup>1</sup> H and <sup>15</sup> N, for all four ligands, giving these values 0.72, 0.16, 0.035 and 0.066 for GalNAc, LPS O157:H7, OS R1 and OS R3, respectively. Residues that are involved in ligand binding are depicted in green and those who were not assigned in white for clarity sake and, calcium ions are colored in pink. ....	164
<b>Figure 85. Cloning and purification of recombinantly expressed MGL-ECD.</b> A) Domain organisation of MGL. TM: transmembrane domain. Cyto.: cytoplasmic domain. B) Construct for the overproduction of MGL-ECD. Strep: StrepTag II. Xa: factor Xa protease cleavage site. 3G: tri-glycine linker. C) SDS-PAGE analysis of MGL-ECD overexpression. Lane M: Page Ruler unstained protein ladder; lane 1: total proteins before induction; lane 2: total proteins after induction. A band at about 28 kDa corresponding to MGL-ECD MW is overexpressed. D) SDS-PAGE analysis of purified MGL-ECD. Lane 1: A unique band at about 28 kDa corresponding to pure MGL-ECD is observed.....	169
<b>Figure 86. 600 MHz-2D DOSY spectrum of 20 uM final concentration of ECD MGL.</b> Lectin sample was prepared in 25mM Tris-d <sub>11</sub> , 4mM CaCl <sub>2</sub> , 150 mM NaCl pH 7 in D <sub>2</sub> O (buffer A). Diffusion signals from ECD MGL are present in the upper part of the spectrum between -10 and -11 log (m <sup>2</sup> /s). In the lower part of the spectrum, diffusion signals from molecules which move fast (i.e. small molecules like Tris at 3.45 ppm and water molecules HOD at 4.77 ppm) are also detected. ....	170
<b>Figure 87. 600 MHz <sup>1</sup>H-1D-NMR spectra of ECD MGL (27.8 μM) in buffer A.</b> The efficiency of on-resonances 7.5 ppm, 0 ppm and -1 ppm in STD-diff spectra (B, C and D, respectively) were compared. Spectrum A corresponds to the <sup>1</sup> H-NMR profile of the lectin.....	171
<b>Figure 88. 600 MHz <sup>1</sup>H-1D-STD-NMR spectra of OS-R1 alone in buffer A (top) and reference (bottom).</b> .....	171

Figure 89. <b><sup>1</sup>H-1D-STD-NMR spectroscopy analysis of ECD MGL: OS R3 mixture.</b> On top STD NMR spectrum (up) of the 1:100 mixture of MGL–OS R3 is displayed and the related <sup>1</sup> H reference spectrum is shown at bottom.....	172
<b>Figure 90. STD NMR spectroscopy and epitope mapping analyses of ECD MGL:OS R1 mixture.</b> A) <sup>1</sup> H-NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:100 mixture of MGL: OSR1. Some key proton resonances are labelled. B) STD-derived epitope mapping of the MGL: OSR1 interaction, with colour coding from the highest (red) to lowest (yellow) observed STD effects. C) Schematic structure of OS R1. Phosphate groups are depicted as white letter P in orange circles. ....	173
<b>Figure 91. Homology modelling of human MGL. a)</b> Multiple sequence alignment of human MGL with the homologous protein ASGR. The sequence intervals from Cys157 to Leu284 was considered as CRD MGL and as a template used for the homology modeling, namely ASGR (Cys154- Leu181) are highlighted in grey. The residues important for galactoside binding are highlighted in cyan. <b>b)</b> Superimposition of MGL-CRD homology-based model (in blue) and the X-ray crystal structure from Asialo-glycoprotein (in magenta). ....	175
<b>Figure 92. The 3D model building of the MGL- outer core R1 complex.</b> Gradual molecular docking was done by using Autodock program, starting from the Di-Galactose unit (a and b) with two different sugar orientations, to the pentasaccharide of OS R1. ....	176
<b>Figure 93. A 3D model of the MGL–OSR1 complex.</b> A) Binding poses of the OS R1 docked into the binding pocket of the MGL starting from the 3D model of the MGL–pentasaccharide complex and based on NMR spectroscopy data. B) Specific interactions between binding-site residues and the OS are depicted by dotted lines for polar contacts. ....	177
<b>Figure 94. BLI sensorgram traces from (top) human ECD MGL binding to streptavidin tips and (bottom) from binding to 1.6 μM LOS R1 at 25°C in buffer C.</b> Colored lines represent data traces at different ECD MGL concentrations. Before LOS R1 loading, MGL charged tips were moved to a fresh buffer for the second baseline permitting the standardization of the initial binding value to zero. Binding was followed from t= 0 to 1200 seconds and BLI steps are mentioned for each measurement.....	179
<b>Figure 95. BLI analyses of human ECD MGL (53 nM) binding to different concentrations of LOS R1 ligand at 25°C in buffer C and B.</b> Colored lines represent sensorgrams traces at different LOS R1 concentrations. Binding was followed from t= 0 to 1800 seconds and BLI steps are mentioned for each measurement. ....	180
<b>Figure 96. DLS analyses of LOSs R1 and R3 size distributions before and after extrusion.</b> The size distribution is shown as bimodal peaked with the appropriate diameter sizes in nm, while after extrusion the size became monomodal for both LOSs. ....	182
<b>Figure 97. TEM micrographs showing human ECD MGL disaggregation effects on LOS R1.</b> Complex formation of LOS R1-MGL, at 1:0.01 molar ratio, is manifested in the corresponding image by small aggregates whereas no morphological changes were observed in the case of LOS R3 in presence of the protein. TEM profiles following GalNAc addition showed significant insights for LOS R1-MGL binding. ....	184
<b>Figure 98. Structural differences between major LOSs exposed extracellularly by E. coli R1, R3 and B bacteria tested by fluorescence microscopy.</b> Identical LOS portions are colored in blue and terminal di-galactoses of LOS R1 are depicted in red. The scaffold of E. coli B LOS with a di-Glc instead of an outer core <sup>364,365</sup> . Ethanamine phosphate (EtnP) aminoethyl diphosphate (PPEtn) and phosphate (P) are phosphate groups branched to the inner core. ....	186
<b>Figure 99. Phase contrast and fluorescence microscopy images of E. coli R1, R3 and K12 bacteria mixed with 2 μM ECD MGL<sup>A647</sup>.</b> Cells are concentrated to 19*10 <sup>8</sup> cells/mL and incubated with fluorescently labelled human MGL in buffer C. ....	187

<b>Figure 100. Microphotographs taken for E. coli R1 in presence of human MGL at different growth stages.</b>	
Exponential growth phases are labelled Expo., for early and mid-exponential phases; stationary phase is labelled Stat. corresponding to optical density (OD) values of 0.3, 1.2 and 4, respectively.....	188
<b>Figure 101. Flow cytometry analyses for fluorescent E. coli R1 and E. coli R3 quantification.</b>	
A) MGL-bound cells Fluorescence index showing different estimations by comparing both E. coli strains. B) dot plot distribution of E. coli R1 and R3 populations at a concentration of $9 \times 10^7$ cells/mL. FSC: Forward scattering; SSC: Side scattering. ....	190
<b>Figure 102. Growth phase dependency and Ca<sup>2+</sup> addition effect on human MGL bound E. coli R1.</b>	191
<b>Figure 103. Calcium addition intensifies MGL-bound E. coli R1 fluorescence.</b>	
Left panel: Micrographs of E. coli R1 on phase-contrast (a) and under 635 nm (b and c). Right panel: flow cytometry measurements corresponding to each condition. ....	192
<b>Figure 104. Prospective view of human ECD MGL interactions with E. coli glycoconjugates depending on MGL conformational organisation in immune cell membranes.</b>	
In the membrane, the view of the two biological assemblies of trimeric human ECD MGL is imagined as human MBL (pdb 1HUP) and rat MBL (pdb1KWV) or DC-SIGN (pdb1XAR) for closed and open structures, respectively. Binding sites locations are indicated by the black circles. ....	199
<b>Figure 105. Overview of molecular interactions between CRD/ECD of human MGL and the studied LPS/LOS at the surface of Gram-negative bacteria.</b>	
The image shows essentially LPS and LOS as constituents of the bacterial outer membrane (OM) considering well-defined strains serotypes (rough- and smooth-type). The homology model of CRD MGL was fitted into the ECD of rat MBL (PDB 1KX0) given that the ECD MGL structure is unavailable. Blue and red arrows indicate glycoconjugates interactions taking place at the first and the second binding sites, respectively. ....	202
<b>Figure 106. OS R1 docking into CRD MGL structure predicted from PDB 6PY1.</b>	
Calcium ions are colored in green, OS R1 backbone in cyan with ball and stick representation and, CRD MGL in grey with surface and cartoon representation. ....	203
<b>Figure 107. in vivo and in vitro experiments as future experiments for the evaluation of E. coli R1 immuno-stimulation.</b>	206

## Liste of Tables

<b>Table 1. Overview on some examples of transmembrane C-type lectins.</b> .....	70
<b>Table 2. Chemical shift regions for <math>^{13}\text{C}</math> and <math>^1\text{H}</math> NMR spectroscopy of carbohydrates.</b> Inspired from <sup>273</sup> ..	91
<b>Table 3. Examples of fundamental equations associated to protein-ligand binding kinetics.</b> Adapted from <sup>279</sup> .....	98
<b>Table 4. Typical 3D-NMR experiments used for protein backbone assignment.</b> Colored nuclei correspond to the ones that are experiencing magnetization, and for clarity, $\text{C}\alpha$ are in red and $\text{C}\beta$ in brown; R signifies main chain of residues; and curved arrows indicate the magnetization transfer evolving chemical shift or J-coupling. Depending on the experiment, the chemical shift might be evolved for some nuclei and not for the others. inspired from <sup>305–307</sup> .....	109
<b>Table 5. Core OS types distribution in VTEC E. coli isolates from humans and animal.</b> Taken from <sup>328</sup> .	119
<b>Table 6. Fatty acids components of the isolated LOS from R1 and R3 E. coli strains</b> showing the most relevant fatty acid species (bold) of E. coli strain and their corresponding retention times and molecular masses of the fragmented molecules obtained from fatty acid methyl esters. ....	133
<b>Table 7. Fatty acids components of the isolated LOS from BC J2315</b> showing the most relevant fatty acid species (bold) of BC J2315 strain and their corresponding retention times and molecular masses of the fragmented molecules obtained from fatty acid methyl esters. ....	133

## ***Chapter I.***

## I. Bacterial carbohydrates are Pathogen Associated Molecular Patterns

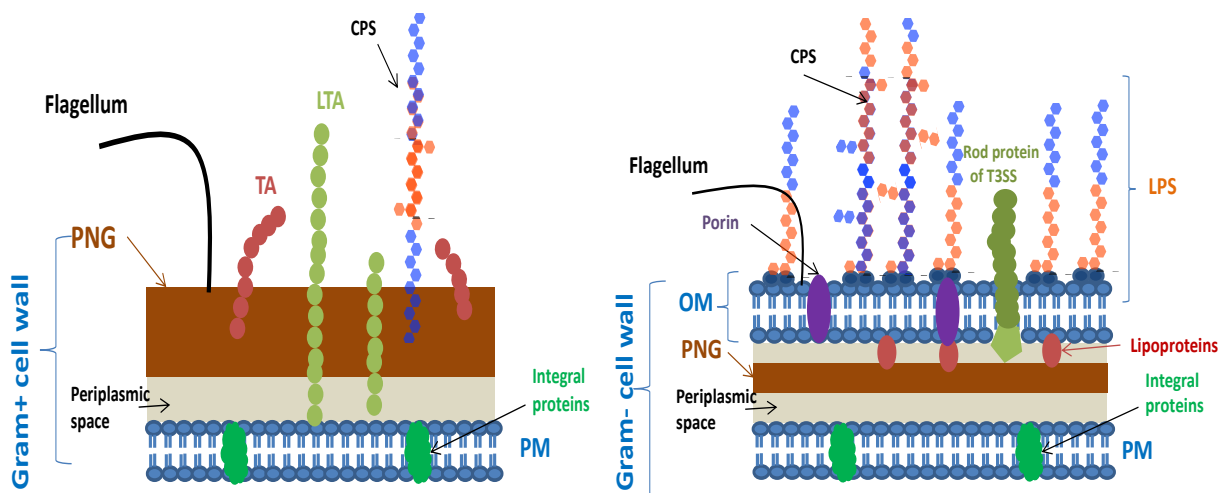
A great attention is increasingly given from scientists to infectious diseases caused by microorganisms like viruses and bacteria. The ability of microbes to adapt to changes in the external environment overtime causes microbial resistance to antibiotics leading to drastic consequences on public health<sup>1</sup>. Bacteria are known to be constantly evolving and some of them could reside in a host-organism (the gut microbiota for example) and they can even transfer genetic material (e.g., new resistance genes) between them while cross-talking. They use up the organism's nutrients and energy to grow and multiply generating wastes like endotoxins<sup>2</sup>. Bacterial endotoxins could be either released by bacteria in the milieu, as said, or naturally present on the surface of the cell wall. Being naturally harmful, some bacteria cause pathogenesis essentially due to their virulence factors. These are surface attached glycans (or glycoconjugates) extremely important for pathogen survival as structural motifs and energy source. Those carbohydrate-containing epitopes (or glycans) are the motifs that immune cells specifically recognize during bacterial infection.

### I-1. The bacterial cell envelope PAMPS

Pathogen Associated Molecular Patterns namely **PAMPs** are characteristic molecular motifs of pathogenic microorganisms. They are expressed by pathogens and are recognized by Pathogen Recognition Receptors **PRRs** produced by innate immune cells, thus triggering the immune response. Certainly, PAMPs could only be found in pathogens (like bacteria and viruses), the host cells do not express any type of PAMP or MAMPs (a classification of Microbial molecular patten referring to non-pathogenic microbes).

Lipoproteins, peptidoglycans, teichoic acids, surface exposed polysaccharides and glycolipids including **Lipopolysaccharide LPS** are some examples of PAMPs<sup>3,4</sup>. Bacterial cell wall possesses different chemical and physical properties depending on whether the bacteria is Gram-positive or **Gram-negative**<sup>5</sup> (figure 1).





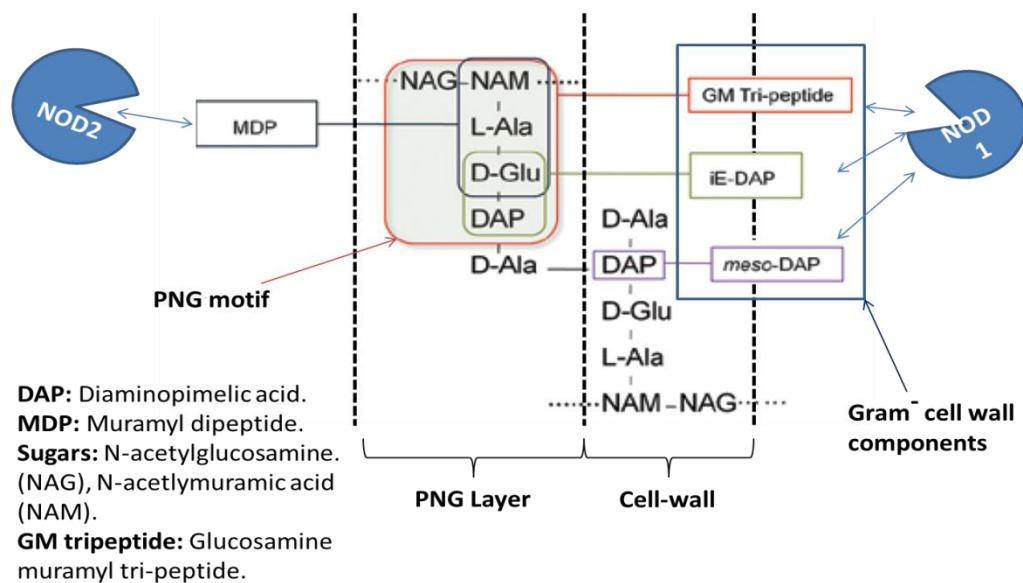
**Figure 1. Schematic representation of Gram-positive and Gram-negative bacteria with their associated PAMPs.** PNG: peptidoglycans, TA: teichoic acid, LTA: lipoteichoic acid, CPS: capsular polysaccharide, PM: periplasmic membrane, OM: outer membrane.

These differences are highlighted by the presence of various molecular structures in each type of bacteria. Gram-positive bacteria's cell wall contains an inner plasma membrane made up with phospholipid bilayer and a thick layer of peptidoglycan (PNG) that is the primary component. These bacteria stain purple after Gram staining because the thick PNG layer enables to retain the crystal violet dye causing them to appear purple<sup>6</sup>. Other essential component, that extend from the plasma membrane to outside the cell through the PNG layer, are Teichoic Acid (TA), Lipoteichoic acid (LTA), and flagella. These cell wall components are glycoconjugates and play essential roles in maintaining cell structure and in proper cell division.

Like Gram-positive, Gram-negative bacteria do express PNG. However, they are characterized by the presence of a double membrane composed of inner plasma membrane separated from the Outer Membrane (OM) by a thin PNG layer, which does not keep the violet crystals during staining (with "Gram's method") but picks up the pink color of the counterstain. The outer-membrane OM is an asymmetrical membrane essentially composed of phospholipids and glycolipids like LPS. Embedded in the OM, lipoproteins and  $\beta$ -barrel proteins can be found<sup>7</sup>. The presence of the OM makes Gram-negative bacteria harder to treat with antibiotics and multiple drugs because of the presence of LPS, increasing their resistance to the host immune system. Thus, a deep study of Gram-negative PAMPs may lead to a better understanding of how eventually control these bacteria in healthcare and to improve therapeutics against these pathogens.

### I-1-1. Peptidoglycan and Teichoic acids (function and biological activity)

In Gram-positive bacteria, peptidoglycans PNG are present in extremely high amounts comparing to Gram-negative and they are mostly considered as PAMPs exclusively from Gram-positive pathogens<sup>5</sup>.



**Figure 2. PNG structure and its interactions with NOD 1 and NOD 2 across the cell wall of Gram-negative bacteria. Adapted from<sup>5</sup>**

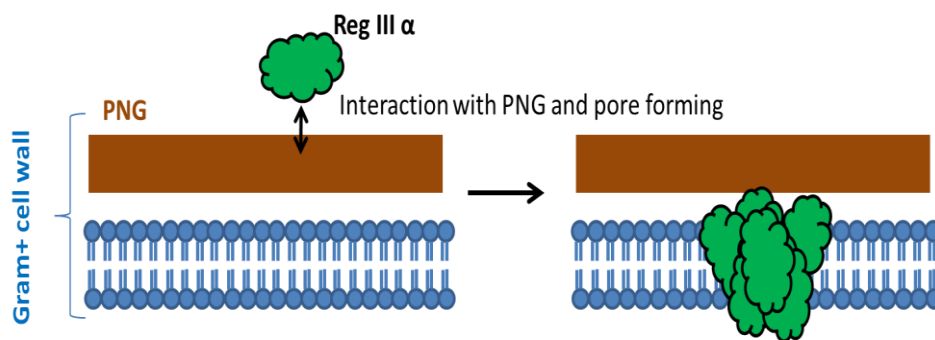
Chemically, PNG's structure is composed of a glycan  $\beta$  (1-4) main chain that is bridged by peptides through N-acetyl-D-glucosamine (2-acetamido-2-deoxy-D-glucose, GlcNAc) and N-acetylmuramic acid (MurNAc) (figure 2). The latter together with peptides present the immunostimulative principle of bacterial peptidoglycan<sup>8</sup>.

These PAMPs are individually sensed by different PRRs. NOD1 and NOD2 are cytosolic sensors belonging to the NOD-like receptors NLR family that specifically recognize PNGs from Gram-negative (NOD1 and NOD2) (figure 2) and Gram-positive (NOD2) pathogens<sup>9,10</sup>.

Another PAMP sensor termed TLR-2 belongs to Toll-like Receptors family (**TLR**), is capable to recognize not only PNG but also LTA when associated with TLR-1 or TLR-6<sup>11,12</sup>. Other studies using *Streptococcus pneumoniae* and *Haemophilus influenzae* as infectious pathogens have shown that TLR-2 together with TLR-4 downregulate cell-cell interactions and accommodate PNG translocation across the epithelium, respectively<sup>13</sup>. TLR-4 also participates to the sensing of another immunostimulant PAMP i.e. LPS; this recognition process will be described later in this chapter. As for TLR-2, its agonists could be both LTA and TA. It has been shown that

Pneumococcal LTA and TA, which are quite chemically complex and share the same structural repeating units, are also targeted by a C-reactive protein, which induces proinflammatory cytokines like TNF- $\alpha$  and IL-1 $\beta$  in human monocytes<sup>14,15</sup>. This protein is very reactive to polysaccharides and found to be calcium dependent and specific for phosphocholine (PC) groups which are present in LTA and TA<sup>16</sup>.

In addition to immunologic tests, the interactions of PNG and teichoic acids were investigated using structural analyses. Lehotzky and colleagues used Nuclear Magnetic Resonance (NMR) spectroscopy as an approach for identifying protein residues involved in peptidoglycan binding. Proteins designated RegIII are C-type lectins belonging to antimicrobial proteins and the presence of Gram-positive pathogen in the intestine is detected by those receptors. RegIII $\alpha$  binds to the bacterial peptidoglycan and subsequently forms a membrane-penetrating pore during pathogen recognition to induce osmotic lysis of the bacteria<sup>17,18</sup> (Figure 3).



**Figure 3. Schematic representation of Reg III lectin mode of action for Gram-positive bacteria lysis.** RegIII $\alpha$  first interacts with peptidoglycan via carbohydrate binding engendering then hexameric proteins cluster in the bacterial membrane. Adapted from<sup>19</sup>.

A wide range of PAMPs are recognized by TLRs and ten human-TLRs members have been identified to date<sup>20</sup>. TLRs usually associate together by homo- or hetero-dimerization resulting from their activation by PAMPs. Taken together, structural analyses of PNG interactions with different proteins concluded from several studies, suggest that large sized PNG fragments facilitate the accessibility of immune receptors for their activation during pathogen recognition.

## **I-1-2. Extracellular polysaccharides (capsular and biofilms)**

### **I-1-2-1. Role of capsular polysaccharides in bacteria resistance**

*Escherichia coli* and *Neisseria meningitidis* Gram-negative pathogens provide well-studied examples for investigating the structure and function of extracellular polysaccharide chains like capsular polysaccharides or CPSs. These bacteria possess capsules organized as long chain polysaccharides that form a dense additional layer around the bacterial cell and are important

virulence factors<sup>21,22</sup>. Interfering with the innate immune system at different levels following pathogen invasion, bacterial capsules participate to complement pathway activation. The function of capsule from pathogenic *E. coli* O75:K5 mutants was estimated by using human-serum assays suggesting that the capsule K5 plays relatively minor role for bacteria survival in serum compared to the O-antigen O75 contribution<sup>23</sup>. Another study made by the same research group, was established to evaluate the role of *E. coli* K5 capsule during pathogen recognition and phagocytosis<sup>24</sup>. K5 capsules have been shown to reduce not only attachment of bacteria to immune cells but also their subsequent internalization. Capsules from *H. influenzae* type-b have shown similar behaviour explained by their crucial role in resistance to phagocytosis over time<sup>25</sup>. These polysaccharide capsules of *H. influenzae* are attractive as a vaccine antigen since their related disease is restricted to one serotype contrary to other strains<sup>26</sup>.

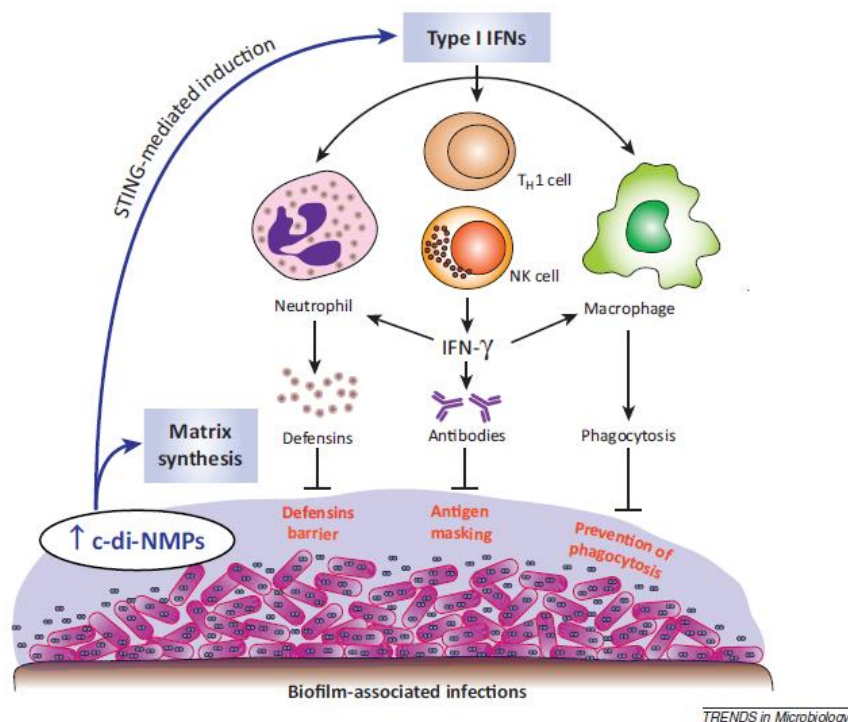
### **I-1-2-2. Polysaccharides of bacterial biofilm are PAMPs**

When an infection tends to persist, this could be due to biofilm formation<sup>27,28</sup>. It is noteworthy to mention that biofilms do not necessarily elicit an immune response since in human microflora, for instance, the host cells live in “peaceful” acquiescence with commensal bacteria. However, the immune tolerance to an intestinal biofilm could change as soon as its residence privileged area moves and/or new bacteria populations appear. That will perturb the faithful relations between host cells and biofilms and lead to host immune response activation<sup>29</sup>.

Biofilm formation is one of the mechanisms that bacteria use to resist to antimicrobial peptides AMPs<sup>30</sup>. A way to make a well-built community, bacteria produce extracellular components that form the biofilm matrix. Each of those components has a role in biofilm structure, protection, and activity toward antimicrobial compounds (AMPs)<sup>28</sup>. Largely studied biofilm components are polysaccharides (including CPSs)<sup>31</sup>.

Contributions from biofilm polysaccharides are essential at different levels and led to an equilibrated mode of growth. Polysaccharides manage the formation of biofilms by participating to each step. Those steps include multiple stages: initial attachment, a subsequent maturation phase, and detachment or disassembly<sup>32</sup>. *Staphylococci* are frequent commensal bacteria which cause biofilm-associated infections including diseases generated from biofilms residing into devices or implants<sup>33</sup>. Those strains are considered as examples of study for *Staphylococcal* biofilm formation investigations<sup>34</sup> and particularly *S. epidermidis*, that is one of the best studied relevant biofilm-forming organisms.

The mechanisms that take place during biofilm formation from the attachment of bacteria to the epithelial surface to biofilm maturation, require some factors such as the so-called aggregative polysaccharides and adhesive proteins. Those polysaccharides play essential roles in formation of complex structures by promoting microbial interactions and disconnecting these interactions promoting dissolution of the biofilm<sup>35</sup>. Biofilm-associated PAMPs like cyclic dinucleotides (c-di-NMPs) have been widely discussed for their importance in inducing type I interferon (IFN) innate immune response (figure 4) and more importantly for the uncertainties related to their recognition at the first stage of infection<sup>36</sup>.



**Figure 4. Biofilm formation and host type I interferon (IFN) innate immune response activation through sensing of biofilm-specific PAMPs i.e. c-di-NMPs.** Taken from<sup>36</sup>.

*Pseudomonas aeruginosa* and *Burkholderia cepacia* are common biofilm bacterial pathogens found in Cystic Fibrosis pneumonia patients contributing to chronic infections. Mutagenesis studies showed that the genes responsible of *P. aeruginosa* polysaccharides, Psl, production are essential for cells attachment<sup>37</sup>. The structures of Psl are not well known but believed to be a mixture of mannose Man, glucose Glc, and rhamnose Rha and possibly galactose Gal<sup>38</sup>. Psl possesses multiple roles including its involvement in bacteria adhesion and interactions but also it acts as a signalling<sup>39</sup> and protective<sup>40</sup> molecule against some antibiotics, in the biofilm matrix. Exopolysaccharides from *P. aeruginosa* PAO1 for example are highly required for biofilm formation which is not the case in PA14 strain<sup>41</sup> meaning that there is a strain-specific polysaccharides production during biofilm formation<sup>31</sup>. Each pathogenic strain forming a biofilm

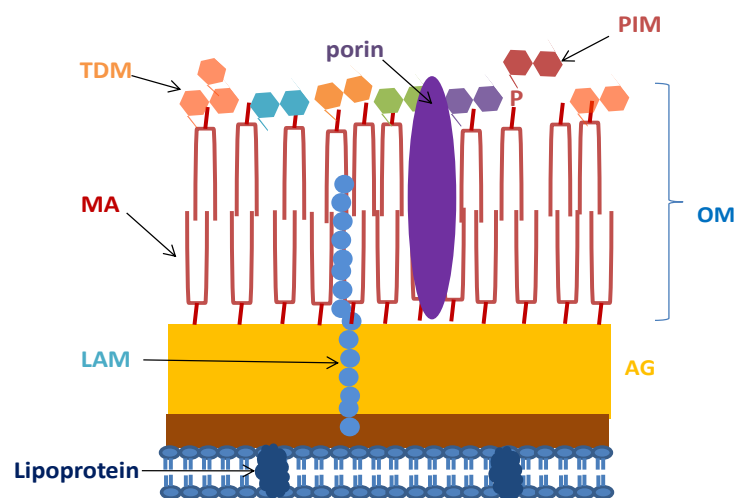
has its own PAMPs which makes it interesting to investigate the involvement of biofilm polysaccharides in specific immune response activation.

### I-1-3. Other glycolipids

In this part of the chapter, well-studied examples of glycolipids from *Mycobacteria* and Gram-negative bacteria are discussed. For clarity sake, and because recognition mechanisms involving PAMPs and immune cell receptors are presented in this part, the details related to innate and adaptive immunity are described in the third part of this chapter (part I-3-1).

#### I-1-3-1. *Mycobacteria* infection and role of its PAMPs

Unlike other Gram-positive and Gram-negative bacteria, *Mycobacteria* are enveloped by a high content of motifs including Trehalose dimycolate (TDM) in their Outer Membrane OM, rich in Mycolic acid (MA) and Arabinogalactan (AG), Lipoarabinomannan (LAM), Peptidoglycan and Phosphatidylinositol mannose (PIM) (figure 5).



**Figure 5. Main components of mycobacteria cell wall.**

Notoriously, mycobacteria have a clever infective way by which they could evade the immune response. After the recognition of their immunomodulatory ligands (e.g. TDM) by alveolar macrophages, mycobacteria reside in the phagolysosome of these macrophages and dorm inside for some time, thereby escaping from the immune response. There are overly complex mechanisms by which mycobacteria run off immune response, sometimes achieved with the help of immune system effectors themselves or even ions like hydrogen and iron from macrophages which indirectly inhibit lysosome formation. These mechanisms include suppressing phagolysosomes formation and hence phagocytosis, inhibiting macrophages formation, resistance to reactive intermediates like oxygen reactive species ROSs secreted by neutrophils and many other mechanisms that involve also PAMPs recognition through TLRs. Theses

participate in immune response to *Mycobacterium Tuberculosis* Mtb and concomitantly helps Mtb to get away from immunity. As said, this microorganism vigorously takes advantage of any situation during the infection<sup>42</sup>.

#### **I-1-3-1-1. Examples of immune pathways activated by *M. Tuberculosis* PAMPs**

Several studies have discussed the pathogenesis of Mtb strain causing Tuberculosis infection<sup>43-45</sup>. Recognition patterns between Mtb PAMPs and immune cell receptors have been studied. TLR-4, TLR-2 and TLR-9 among the TLRs family are involved<sup>45</sup>. This recognition is accompanied by downstream recognition pathways involving other receptors in addition to TLRs, i.e. myeloid differentiation primary response protein 88 (MyD88) which plays an essential role in innate immune response to Mtb<sup>46</sup>. A second pathway in response to Mtb infection involves TLR4 and the adaptor molecule Toll/IL-1R domain containing adapter inducing interferon IFN- $\beta$  (TRIF pathway).

##### **I-1-3-1-1-1. Recognition of Mtb PAMPs by immune system**

Mycobacterial cell wall glycolipids like LAM, LM, mycobacterial glycoprotein, PIM, and acylated lipoproteins are PAMPs for TLR2 which forms heterodimers with either TLR1 or TLR6 during pathogen recognition<sup>47</sup>. Stimulatory effects of TLR-2 on pro-inflammatory cytokines TNF- $\alpha$  production<sup>48</sup> and IL-12<sup>49</sup> in macrophages throughout Mtb infection, have been demonstrated. In line, interleukin receptor associated kinases (IRAK1/4) are subsequently recruited with TLR-2 to activate the so-called Hippo pathway -which has been classified as reprogrammed signalling cascades to furnish specific defensive measures alongside immunologically active pathways- thus triggering modulation of chemokines expression and secretion<sup>50</sup>.

NOD-2 sensor, previously presented as a receptor for Gram-positive, acts in synergy with TLR-4 to boost the release of pro-inflammatory cytokines by dendritic cells (DCs) during innate response against Mtb<sup>51</sup>. A study established by Khan et al<sup>52</sup> aimed to understand this synergism between NOD-2 and TLR-4 in response to Mtb using a combination of immunological tests. The activation of both receptors induces the level of "autophagy" in DCs which is a vital process taking place after phagocytosis and involves fusion of the phagosome with autophagosome<sup>53</sup>. The authors showed that NOD-2 and TLR-4 activation leads to the enhancement of antigen presenting ability of DCs to T cells which activates, in turn, the adaptive immunity against Mtb. In the same context of autophagy process, a study established by Xu et al<sup>54</sup> highlighted the importance of LPS stimulation in the activation of Mtb-induced autophagy through a shared signalling pathway.

Knowing that Mtb does not have LPS in its cell-wall, the authors assumed that modulation of autophagy through Toll-like receptors activation (TLR-4) by bacterial LPS may stimulate certain important pathways against Mtb.

Apart from Mtb PAMPs recognition by TLRs, dectin-2, DC-SIGN and Mincle myeloid cells receptors (or **C-type lectins**) have been considered for their involvement in the interaction with Man-LAM<sup>55</sup> and Lipomannans<sup>56</sup> and TDM<sup>57</sup>, respectively. Although it has been shown that the recognition between Man-LAM and dectin-2, for instance, is mainly mediated by multivalent interactions through di-mannoside-LAM caps, the molecular mechanisms behind their recognition have not been studied yet. On the other hand, the recognition mechanism between a mouse Mincle lectin and small molecule analogs of mycobacterial TDM has been elucidated<sup>58</sup> and it is described in the next chapter of this manuscript (II-2-2-1), as an example of lectin-glycoconjugates interactions.

### **I-1-3-2. Bacterial Lipopolysaccharide LPS, the hallmark of bacterial infections**

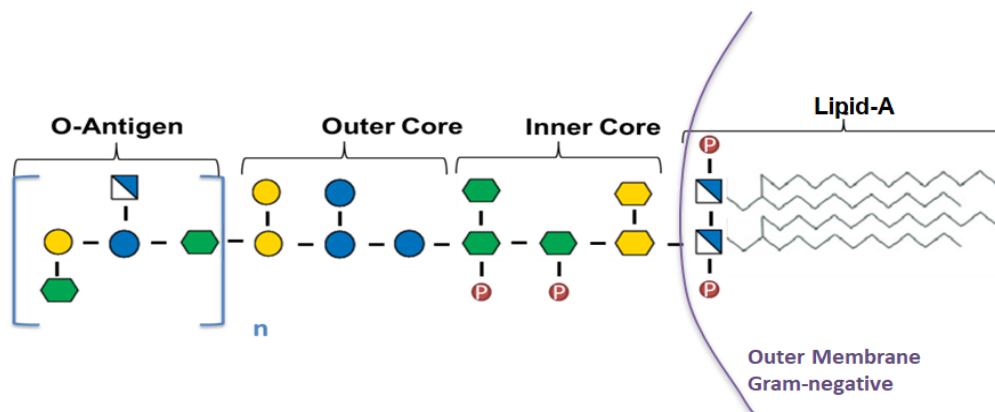
**LPS** is classified as an endotoxin, that, once released into the body by lysed bacteria, plays a key role in the pathogenesis of Gram-negative infections. Besides, as soon as it remains on the cell wall, it serves as a PAMP. The term endotoxin was first discovered by the Danish physiologist Peter Ludvig Panum in 1856<sup>59</sup> who has carefully performed experiments on "putrid poison" - a cell wall product of Gram-negative bacteria - that was suggested to be an endotoxin responsible for inflammation symptoms in patients with sepsis. Later, from 1888 to 1892, several researchers have defined the concept of endotoxin through numerous experiments using died pathogens and their supernatants as immune response stimulators, until the discovery of LPS in 1943 by the American researcher Murray J. Shear<sup>60</sup>. Thenceforth, researchers from all over the world have enforced their knowledge to study LPS endotoxin structure and function. To date, several LPS-induced inflammatory responses were elucidated to figure out how LPS could stimulate the production of pro-inflammatory substances and activate certain pathways during the immune response<sup>61-65</sup>. However, few studies have been established regarding its structure-function relationship due to, not only its heterogenous scaffold and its typical molecular assembly in solution, but also its hard/demanding chemical synthesis, making relevant recognition mechanisms of LPS by immune cells receptors hard to investigate deeply.



### I-1-3-2-1. LPS structural aspect

The cell wall of Gram-negative constituent **LPS** is a potent pro-inflammatory PAMP<sup>4,66</sup>. Among the PAMPs described along this PhD manuscript, LPS from *E. coli* mutants and wild-type, are our main molecules of interest. LPS is a glycolipid with an amphiphilic structure gathering several types of sugars and fatty acids (figure 6). Structurally, LPS in its Smooth form, consists of three distinct domains: (1) an O-specific polysaccharide (O-Antigen), (2) a core oligosaccharide (core OS) and (3) the lipid A moiety which anchors LPS to the outer leaflet of the bacterial cell wall. In Rough form bacteria, LPS is dubbed as Lipooligosaccharide or **LOS** that lacks the O-Antigen portion.

The existence of such heterogenous structure of LPS influences its capability to interact and activate host immune response to pathogens including for example *Haemophilus influenzae*, *E. coli*, *Salmonella enterica*, *Klebsiella pneumoniae*, *Bordetella pertussis*, *Pseudomonas aeruginosa*, *Chlamydia psittaci*, and *Legionella pneumophila*<sup>4</sup>.



**Figure 6. Common structure of Enterobacterial LPS gathering different types of sugars.** Sugar residues are presented following the symbolic representation with the rules set-forward from Symbol Nomenclature for Glycans (SNFG)<sup>67</sup> which compose the O-Antigen and the core oligosaccharide, covalently attached to Lipid-A moiety that is essentially composed of fatty acids and 2 glucosamines. Linkages of sugars shown in this scheme are for diagrammatic purposes only.

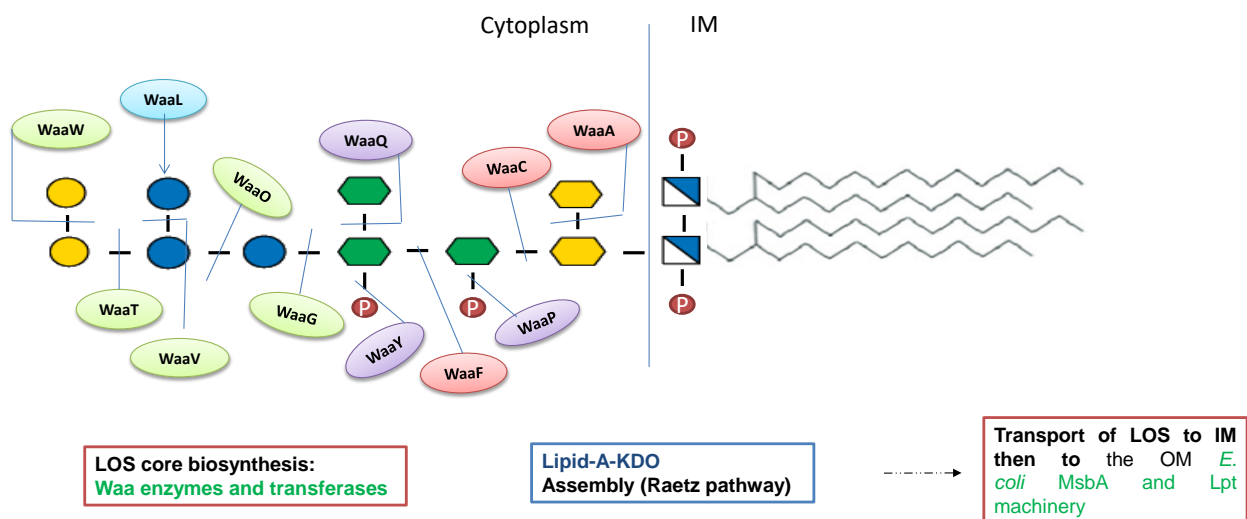
### I-1-3-2-2. Biosynthesis and transport of LPS

Being the major component of the cell wall and asymmetrically anchored to the Outer Membrane OM<sup>68</sup> and bridged by divalent cations ( $Mg^{2+}$  and  $Ca^{2+}$ ) in the lipid bilayer, LPS molecules form a very well-ordered OM structure in Gram-negative. Many well performed reviews describe the bacterial OM, its biogenesis and function<sup>69,70</sup>. OM proteins (e.g. protein channels) are produced in the cytoplasm, whereas phospholipids and LPS are synthesized at the cytoplasmic leaflet of the Inner Membrane IM<sup>71</sup>. In the early 1970s, biosynthesis of OM

constituents posed a special issue in that the final product is found in a molecule physically separated from its cytoplasmic precursors (e.g. nucleotide sugars) by the OM, which raised the question whether the location of LPS synthesis is in situ or elsewhere. Osborn and collaborators have proved that LPS final location at the OM occurs by its unidirectional transfer<sup>72</sup>.

### I-1-3-2-2-1. Biosynthesis of LPS variants

The enzymatic synthesis of LPS in *Salmonella* and *E. coli* strains is guaranteed by enzymes localized in the cell envelope. Kadmas and coworkers revealed the activity of the enzymes responsible of the synthesis of each part of LPS in *E. coli*, like Heptosyl transferase I, which adds heptose residue to the Kdo<sub>2</sub>-Lipid-A, thus forming the inner core LPS<sup>73</sup>. The biosynthetic pathway of LPS, also referred as Raetz pathway, has been extensively elucidated by Christian Raetz and collaborators<sup>74</sup>. Briefly, the authors demonstrated that all components of LPS (lipid A, core oligosaccharide and O-antigen) are synthesized at the Inner Membrane IM, where lipid A and core are ligated for Rough-LPS (LOS) formation.



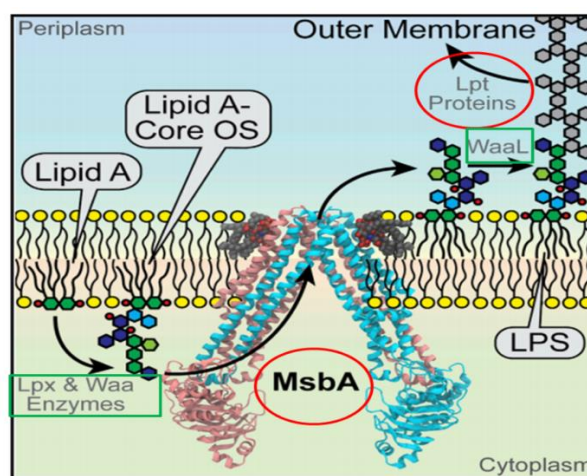
**Figure 7. Biosynthesis steps of LOS glycoconjugate, from *E. coli* R1 bacteria followed by its transport to the OM.** Lipid-A-KDO<sub>2</sub> stepwise biosynthesis is insured by (LpxA,C,D,H,B,K,L,M) enzymes machinery followed by core construction. Glycosyltransferases needed for Inner Membrane (IM) and Outer Membrane (OM) assembly are colored in purple and green, respectively. In red are labelled the enzymes required for inner core modifications. For clarity, dashed arrow is pointed to further steps for LOS transport and the blue glycosyltransferase could intervene in by O-Antigen addition in the periplasm, in case of strain mutation. Inspired from<sup>75,76</sup>.

Figure 7 shows an example of **LOS** biosynthesis, produced by *E. coli* R1 strain, in which we are interested. In case of LPS production, LOS together with the O-antigen are flipped across the IM and ligated by the WaaL ligase at the outer leaflet of the IM. Notably, *E. coli* K-12, a daughter

strain, only synthesizes LOS because the gene responsible for O-antigen biosynthesis is mutated<sup>77</sup>.

#### I-1-3-2-2-2. Transport of LPS across Gram-negative membranes

After synthesis of LPS/LOS in the inner face of IM, it is translocated to the outer leaflet of the IM (figure 8). LPS flipping across the IM is guaranteed by the ATP-binding cassette ABC transporter MsbA, which is involved in lipids trafficking in *E. coli*<sup>78</sup>, although in *Neisseria meningitidis*, MsbA is not engaged for this function<sup>79</sup>. This protein is the best understood LPS transporter and it has been largely studied so that several X-ray structures have been proposed<sup>80,81</sup>.

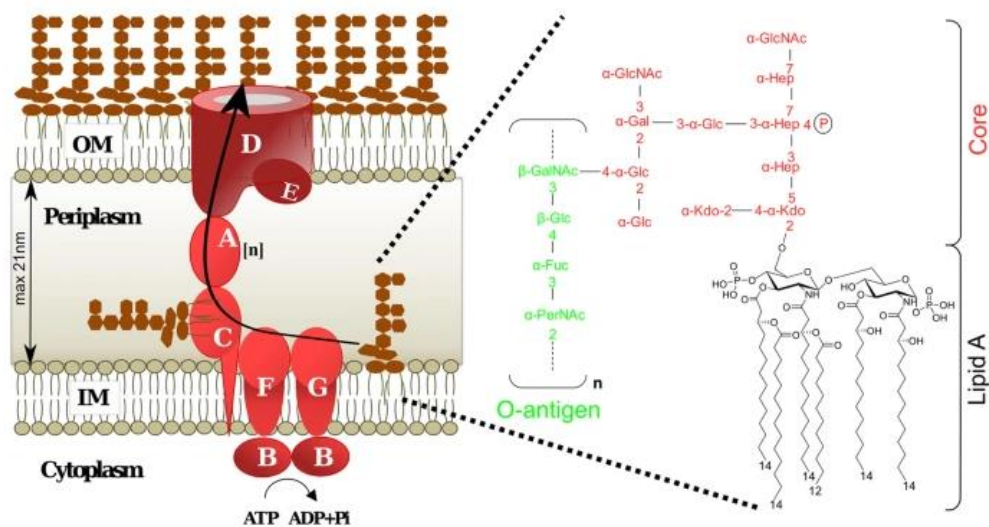


**Figure 8. Schematic representation of LPS transport across the cell envelope and assembly at the OM.** For clarity, green rectangles refer to enzymes involved in LOS and LPS biosynthesis in the OM and IM, respectively; and proteins involved in LPS/LOS transport are depicted in red circles (i.e. Lpt transport protein complexes and MsbA). One of the recently published X-ray structures of MsbA is presented. Adapted from<sup>81</sup>.

The ABC transporter superfamily includes other important proteins such as LptB, LptF and LptG, and the membrane protein LptC, are all involved in LPS transport pathway. Summarily, LPS is first extracted from the IM by the previously mentioned ABC transport proteins and it secondly crosses the periplasm to the OM thanks to the periplasmic protein LptA, the OM lipoprotein LptE and the  $\beta$ -barrel OM protein LptD which also belong to the ABC transporter superfamily<sup>82</sup>. A review written by Ruiz et al discussed some proposed models of LPS transport through the IM and OM and its assembly<sup>70</sup>. Protein mediated transport of LPS is insured by this huge complex of proteins that are extended from the IM (LptB<sub>2</sub>CFG) to the OM (LptCADE) (figure 9) where LptDE sub-complex mediates LPS assembly at the OM surface. The first IM subcomplex uses ATP

hydrolysis to aliment LPS translocation. The second periplasmic complex LptCA forms  $\beta$ -jellyroll folds forming a bridge with a hydrophobic groove<sup>83</sup> and interacts with LPS during its transport<sup>84</sup>.

The full mechanism behind LPS transfer from the IM to the OM remains complex since several transport models have been proposed with many emerged questions regarding the interactions between proteins responsible for that function and the source of energy for LPS assembly at the OM. Although we know every Lpt factor involved in LPS transport, we still lack information about their elemental functions. Parenthetically, our research group is currently working on that subject, to figure out, how LPS transport occurs through the membranes at the atomic level<sup>84,85</sup>.



**Figure 9. *E. coli* LPS transport Lpt machinery.** Structure of the LPS molecule is detailed on the right part of the figure to show each part of LPS interacting with the transporters (i.e. LptC, LptA) while being translocated. For clarity, n refers to the number of LptA proteins, and distance between IM and OM is indicated in nanometers nm. Taken from<sup>84</sup>.

## I-2. Bacteria survival role and molecular diversity of LPS

The barrier function of the Outer Membrane OM was demonstrated in 1958 first by MacGregor and Elliker. They noticed a high sensitivity of *P. aeruginosa* to quaternary amines upon treatment of the OM with ethylenediaminetetraacetic acid (EDTA)<sup>86</sup>. Few years later the research group of Loretta Leive have shown that the observation of increased sensitivity of the OM was because EDTA promotes the entry of several compounds into *E. coli* cells<sup>87,88</sup>. The same group proved afterwards, using biochemical and genetic studies, that EDTA is responsible of LPS release from the cell surface<sup>89</sup>. This observation is explained by EDTA chelation to the divalent cations that are intercalated between LPS molecules ( $\text{Ca}^{2+}$  and  $\text{Mg}^{2+}$ ), and this results in LPS repulsion and release from the OM. Moreover, the barrier quality of OM from rough LPS mutants (expressing LOS) was studied, showing more permeable structures than their wild-type

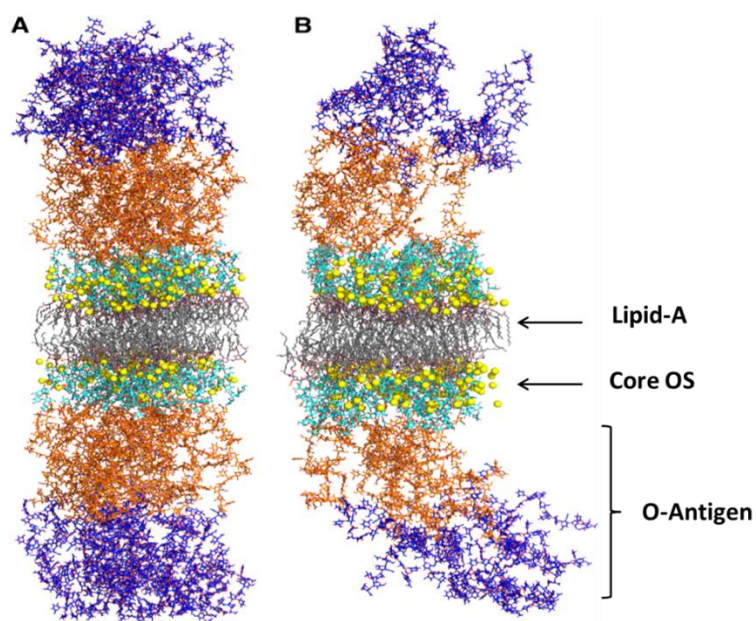
counterparts<sup>90</sup>, which might prove the contribution of O-Antigen capping in the OM stability. In addition to the O-Antigen, the trademark responsible for the impermeability of the OM is the high degree of packing of LPS molecules into the outer leaflet of OM which confers the low fluidity<sup>91</sup>.

The tight packing of LPS molecules into the OM is accomplished by its structural features. The LPS structure is rich of fatty acids chains which results in the establishment of many hydrophobic interactions between LPS molecules aligned on the OM; this, together with the presence of phospholipids, guarantee a good impermeability of the OM. In addition, fatty acids chains from enteric bacteria are usually saturated, which allows tighter packing of LPS molecules. The presence of unsaturated lipids, at low temperatures, increases the fluidity of bacteria membrane. As said, the role of divalent cations is also crucial for maintaining the integrity of the LPS into the OM.  $Mg^{2+}$  and  $Ca^{2+}$  neutralize the negative charges provided by phosphate groups present in Lipid-A which facilitates lateral interactions between LPS molecules, thus forming bridging interactions that protect against antibiotics.

Consequently, altering the structure of LPS molecules compromises the protection of the bacteria and the barrier function of the OM indicating the vital role of LPS in maintaining the integrity of the OM structure, the control the influx and efflux of solutes and the protection of Gram-negative against antibiotics. For example, if the asymmetry between the IM and the OM is lost, the balance in phospholipids would be altered because they are mislocalized but there is a regulation system (i.e. periplasmic proteins) dedicated for their retrograde and anti-retrograde transport<sup>92</sup>. The structure of OM could be lost or altered for many reasons, for instance, when the LPS transport machinery is absent and consequently LPS balance is affected at the OM or when a detergent is added and then it competitively disrupts LPS/LOS/phospholipids lateral bridges at the OM.

### **I-2-1. LPS heterogeneous structure**

As presented in figure 6 and as discussed throughout this chapter, all LPS molecules possess conserved architectures characterized by the presence of three biosynthetically and genetically distinct domains including the lipid A, core oligosaccharide and O-polysaccharide<sup>4</sup>. The structural arrangement of LPS and LOS molecules in solution or at the cell surface is defined as membrane-like structures (figure 10) due to the presence of fatty acids that constitute Lipid-A portion.



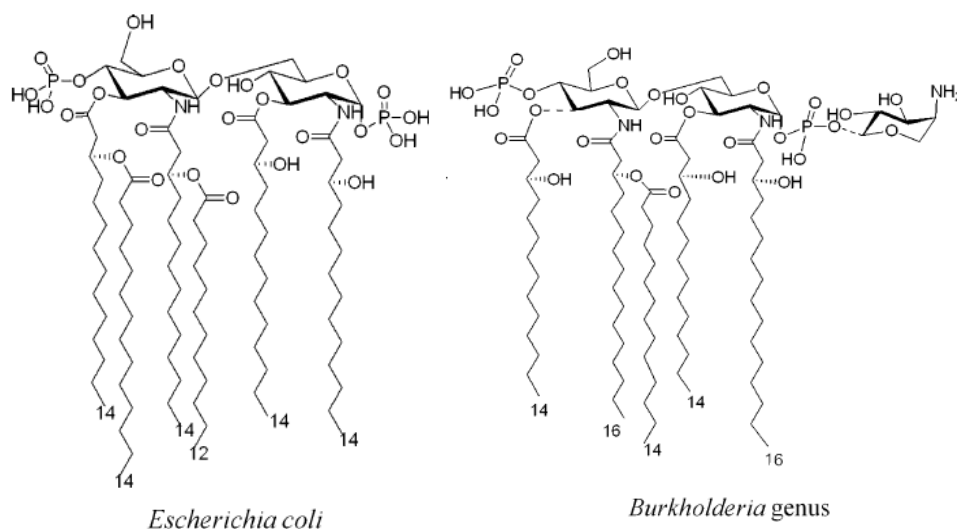
**Figure 10. LPS (A) and mixture of LPS and LOS molecules (B) assemblies in the lipid bilayer constructed using MD simulations.** Calcium ions are depicted in yellow for clarity. Adapted from<sup>93</sup>.

Notably, LPS diversity is exhibited by the fine structural and functional heterogeneity of each portion of the molecule. This molecular diversity provides challenges in the understanding of its behaviour during the immune response to pathogen which may be highly dependent on each part of the molecule. In the next part of the chapter, I will introduce some modifications within each LPS portion in an increasing order of heterogeneity starting from the Lipid-A to the O-Antigen parts.

### **I-2-1-1. The lipid-A moiety**

A review published in our research group summarized very well the chemical aspects regarding lipid A and its role in innate immunity<sup>94</sup>. In terms of general scaffold, Lipid-A is consisted of a disaccharide backbone formed by a  $\beta$ -(1 $\rightarrow$ 6) glucosamine backbone, phosphorylated at position 1 of the proximal  $\alpha$ -GlcN (GlcN I) and position 4 of distal  $\beta$ -GlcN (GlcN II) and acylated at position 2 and 3 by “primary” fatty acids, which are further acylated at hydroxyl position with amide and ester “secondary” fatty acids, usually not hydroxylated and with different lengths as reported in figure 11 showing two examples of Lipid-A.





**Figure 11. Chemical structure of Lipid-A from *Escherichia. coli* and *Burkholderia Genus*.** Taken from<sup>95</sup>.

In addition to the global structural variations in Lipid-A structures through Gram-negative strains, this LPS portion also presents a microheterogeneity due to the presence of subtle chemical differences originating from a wide number of factors including bacterial adaptation and incomplete biosynthesis. Different modifications, presumably in any region of the Lipid-A, can be present<sup>96</sup>. Microheterogeneity has been detected in the acylation patterns concerning the number, types and symmetrical (e.g., *N. meningitides*) and asymmetrical (e.g., *E. coli*) distribution of fatty acids, which is in turn strictly connected to the 3D shape of Lipid-A backbone. Phosphate groups that decorate the Lipid-A from both sides of the molecule can also be substituted by additional phosphates thus forming pyrophosphate (PP), or other polar substituents such as Amino-Arabinose (Ara4N), typically found in *Burkholderia* strains<sup>97</sup>, or in *Bdellovibrio bacteriovorus*<sup>98</sup>.

Moreover, some structural variations are beneficial and strongly contribute to bacterial resistance and protection, thus becoming even more harmful for the host. For example, a study established by Hamad and coworkers, demonstrated that the presence of Ara4N in the lipid A of *B. cenocepacia* LPS significantly increases the resistance to cationic antimicrobial peptides by shielding the negatively charged residues<sup>99</sup>. Conversely, *Francisella* species possess a Lipid-A without phosphates instead of Ara4N which may result in relatively affected membrane stability<sup>4</sup>.

Finally, despite that the overall LPS structure influences its function and recognition by host cell, the Lipid-A intrinsic conformation is the most responsible for its agonistic and antagonistic activity in innate immune system<sup>100</sup>. In fact, the distribution and length of fatty acids constituting

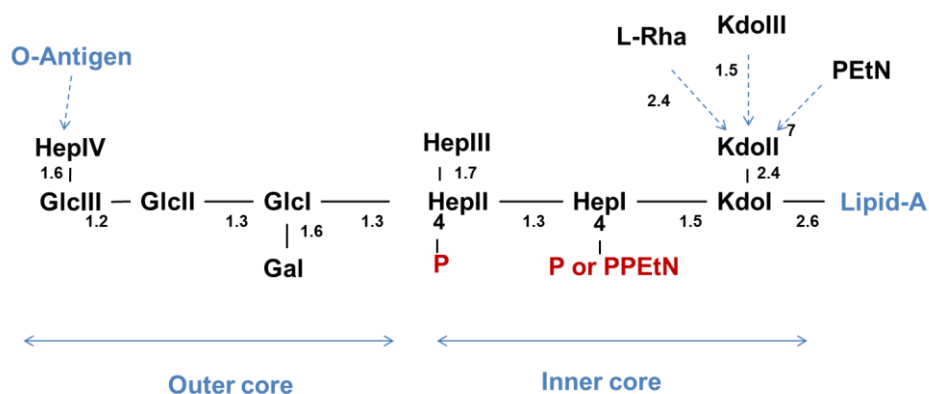
the lipid-A as well as the number of phosphate groups define its endotoxic activity. Lipid-A with a “canonical” shape (i.e. hexa-acylated lipid-A from *E. coli*) renders LPS agonistic with relatively high endotoxicity; whereas tetra-acylated lipid-A (Lipid IVa from *Y. Pestis*) or long fatty acid chains acts as an antagonist abolishing the activity of endotoxically active LPS in humans. Although the latter does not exhibit TLR4/MD2 activation in humans, it can activate caspase-4 or caspase-11 pathways depending on the strain and/or on Lipid-A modifications (i.e. addition of Aminogalactose GalN residue at the phosphate group)<sup>101</sup>. Commonly, bacteria mutants producing only lipid-A lacking outer core neither O-Antigens (i.e. expressing deep rough LPS) are more susceptible to most clinical antibiotics<sup>102</sup>.

### **I-2-1-2. The core oligosaccharide**

In LPS molecule, the Lipid-A portion is covalently linked to the core oligosaccharide **OS**. The latter is composed by different monosaccharide units, approximately up to 12, in linear or, in branched architectures in strains possessing non-carbohydrate components presented in a non-stoichiometric fashion<sup>103</sup>. An excellent review written by Caroff and Karabian<sup>104</sup>, describes the structure-function relationship of LPS components with precise insights on the OS portion. The latter is composed of an **inner core**, that is directly bound to the glucosamine residues of Lipid-A, and an **outer core** which is either ligated to the O-Antigen in smooth type LPS, or freely exposed to the outside of OM in rough type LOS as reported in figure 12 showing general structure of core oligosaccharide in *E. coli K12* strain. This strain produces mostly LOS since it does not synthesize the O-antigen because of mutation, in its related operon, that has evolved during lab culturing. However, it is possible that *E. coli K12* expresses, by correcting that mutation, the appropriate ligase which would attach the O-antigen to the core<sup>105</sup>.

Structural analyses by Muller-Loennies and coworkers deeply characterized, for the first time, several glycoforms expressed by this strain<sup>106</sup>, exhibiting various substitutions from the inner to the outer core accompanied by some truncations.

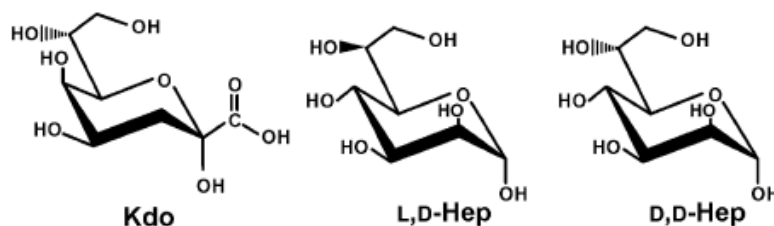




**Figure 12. General presentation of *E. coli* K12 core oligosaccharide.** Dashed arrows highlight all possible substitutions.

### I-2-1-2-1. The inner core composition

The **inner cores** of most studied LPSs do have characteristic sugars like L-glycero-D-manno-heptose, D-glycero-D-manno-heptose and  $\alpha$ -3-deoxy-D-manno-oct-2-ulosonic acid (**Kdo**) covalently linked to Lipid-A (figure 13). The latter residue connects the inner core oligosaccharide to lipid A backbone through glucosamine GlcN II and it is considered as LPS marker<sup>107</sup>.



**Figure 13. Chemical structures of 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo), L-glycero-D-manno-heptose (L,D-Hep), and D-glycero-D-manno-heptose (D,D-Hep), the features of the enterobacterial inner cores.** Taken from<sup>108</sup>.

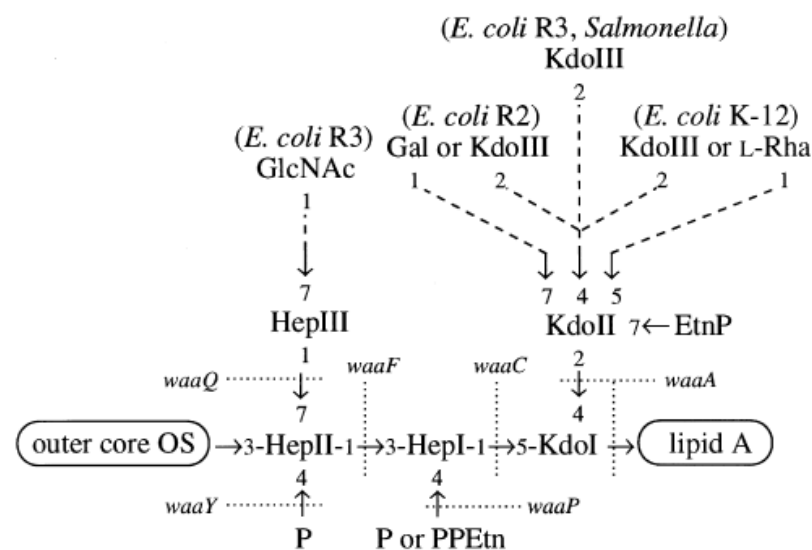
Kdo residue has been extensively studied as LPS tracer by considering, for instance, its derivative 8-amino-3,8-dideoxy- $\alpha$ -D-manno-oct-2-ulopyranosonic acid (Kdo8N), for studying LPS biosynthesis and transport<sup>109</sup>. Furthermore, some other strain-specific ulosonic acids have been found in *Burkholderia*, *Acinetobacter*, or *Yersinia* species, including D-glycero-D-talo-oct-2-ulopyranosonic acid (Ko)<sup>103,110</sup>. A considerable variety of LPS glycoforms (variants) with different substituents linked to heptoses I and II from the inner core have been observed in many microorganisms such as strain RM7004 of *H. influenzae*<sup>111</sup>.

Some significant substitutions in the inner core, by negatively charged substituents including phosphate (P), pyrophosphoryl-2-amino-ethanol (PPEtN), phospho-arabinosamine (PAra4N) and uronic acids have been detected in bacterial LPS<sup>112</sup>. The presence of these moieties in this part

of the LPS molecule confers the decrease of permeability of the OM, by the formation of electrostatic interactions with divalent cations, thus enhancing cell stability. Moreover, positively charged groups might sometimes be present, including 2-amino-ethanol (EtN) and Ara4N or GlcN residues, replacing cations and interacting with the phosphate groups, thus favouring LPS molecules packaging and blocking the entrance of positively charged molecules. It has also been reported that the inner core of *Francisella tularensis* LPS lacks heptoses<sup>99</sup> and it is composed by mannose residues instead, directly linked to Kdo sugar.

### I-2-1-2-2. The outer core structure defines LOS molecules

In *E. coli* bacteria, five core types have been so far distinguished by means of structural investigations (i.e. **R1**, **R2**, **R3**, **R4** and **K12**). These *E. coli* prototypes share similarities in their inner core structures (figure 14).



**Figure 14. Representation of inner core structure distribution in *E. coli* and *Salmonella*.** Dotted lines indicate the linkages formed by specific gene products and dashed arrows refer to all possible substitutions among mentioned strains. Ethanolamine phosphate (EtnP) aminoethyl diphosphate (PPEtn) and phosphate (P) are phosphate groups branched to the inner core. Taken from<sup>75</sup>.

Even though the inner core is already variable among distinct strains with the previously mentioned examples of substitutions, the outer core exhibit more diversity within the same strain. Mainly composed of hexoses like Glc, GlcN, Gal and GalN, the outer core has been vastly investigated because of not only their structural diversity but also their involvement in pathogen recognition by immune cells. As an example, Vinogradov and collaborators have studied a wide range of structure changes touching the outer core, that include side chain substitutions like a Kdo linked to HepIII, or GalA amide-linked to an aliphatic amine of certain amino acid or to the

$\alpha$ -amino group of L-lysine<sup>113</sup>. The different outer core sugars substitutions of the oligosaccharides in which we are particularly interested i.e. *E. coli* LOSs R1 and R3 are presented in chapter IV.

### **I-2-1-3. The O-polysaccharide or O-Antigen**

The outermost portion of LPS molecule is constituted by a sequence of numerous repeating units (there may be up to 50 identical and up to 100 different segments that have been identified so far) forming the O-Antigen. The latter is characterized by the highest level of structural diversity, not only between species, but also between bacteria serotypes<sup>74</sup>. Cellularly, the location and the high hydrophilicity of O-Antigen portion are the main reasons why LPS is extended out of the bacteria.

Being antigenic and in some cases pathogenic, O-Antigens present the basis for serotype classification among the various bacterial strains. The need of bacteria to express the O-Antigen is strictly controlled. Furthermore, it was proposed that bacteria like *Shigella flexneri* and *Klebsiella pneumoniae* can use their O-Antigens to hide themselves from the host immune system by escaping from histones that have antibacterial activities besides their main function in covering DNA<sup>114</sup>. In such bacterial colonies, the O-antigen “cap” would be subject to a hard selection by the immune system since that this LPS extra-portion contains variable residues that could be continuously altered.

Finally, the different functions of the O-Antigen have not been investigated (except with microarray analyses for antibodies detection<sup>115</sup>) at the molecular level, due to the high heterogeneity of such molecules, a drawback that could be resolved through various approaches that are explained in the first part of chapter III.

### **I-2-2. An example of LPS/LOS from pathogenic bacteria**

#### **I-2-2-1. *P. aeruginosa* produces an original LPS O-Antigen**

*P. aeruginosa* Gram-negative is an opportunistic pathogen meaning that it only causes infections for individuals that have weakened immune defense with higher susceptibility for cystic fibrosis patients to develop pulmonary infections, as *B. cenocepacia* pathogens. LPS contribution in *P. aeruginosa* pathogenesis as PAMP has been demonstrated to come mainly from O-Antigen portion<sup>116</sup>. The particularity of the later is that it exists as two bands, the so-called A-band (homopolysaccharide) and B-band (heteropolysaccharide). The former is a common antigen and the later is strain specific. While A-band is composed of 23 trisaccharides repeating units of Rhamnose, B-band is longer with several repeating units of di- to penta-

oligosaccharides<sup>117</sup>. Depending on the single expression of either A-band or B-band O-Antigens and based on their surface charges, LPS from *P. aeruginosa* was extensively studied through the use of mutants (e.g. A<sup>+</sup>B<sup>-</sup>, A<sup>-</sup>B<sup>+</sup>), showing that A-band is more hydrophobic than B-band with high abundance in chronic phases of infection (following biofilm formation). Although B-band is strongly involved in surface adherence for biofilm formation (initial step), this O-Antigen is predominant in early chronic stage<sup>118</sup>. Given its structural complexity, especially regarding the B-band long polysaccharide portion that has higher affinity for antibiotics than A-band O-Antigen, LPS PAO1 (serotype O5, laboratory strain) was studied by our group, by means of solid-state NMR spectroscopy and electron microscopy, and this high sensitivity of B-band to an antibiotic (gentamicin) was confirmed<sup>84</sup>.

### **I-3. Elicitation of host immune response by Lipopolysaccharide**

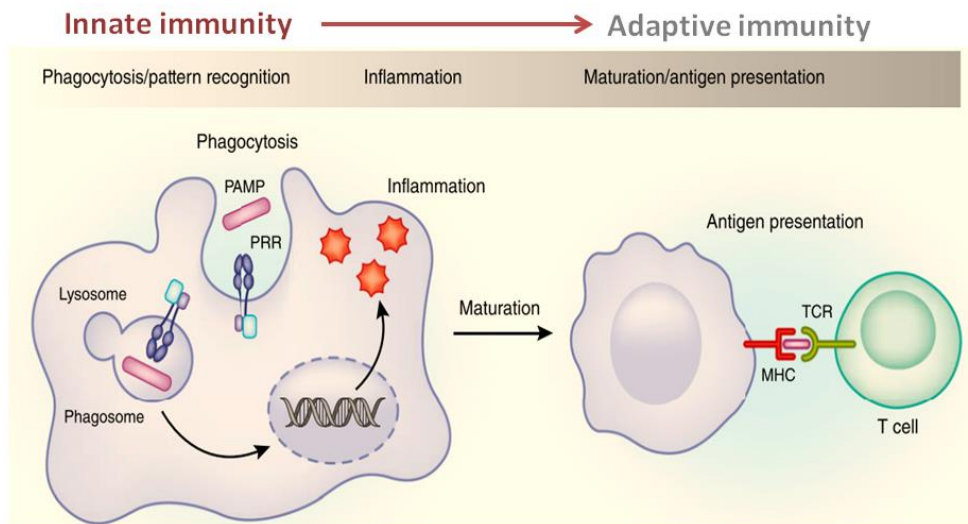
In the first part of this chapter, glycoconjugates from different bacterial types were presented describing their recognition by some PRRs without detailing how the immune response initiates upon the presence of a pathogen in the body. This part is meant to describe how the immune system works under bacterial infection and it is worthy to mention that this chapter is strongly related to chapter II in which some examples of dendritic cell receptors, namely lectins, are discussed.

#### **I-3-1. Innate and adaptive immunity**

Any living organism naturally has a protective system presented by a barrier against attacks from the external environment. From micro- to multi-cellular organisms, this safeguarding is insured by an ensemble of systems. Micro-organisms, for instance, use their structurally well-defined cellular membranes to exchange selectively with the outside; and the same way, multicellular organism like humans possess the skin which plays a similar role. Those are the very first barriers preserving a safe environment inside the organisms.

In case of failure in protection, vertebrates including humans have additional more advanced protective systems called immune system. Consisted of a complex network of molecules, organs and cells, the immune system intervenes in resistance and elimination of the exogenous pathogens from the organism. In mammals, this network is composed of two systems that are interconnected: the innate and the adaptive immune systems<sup>119</sup> (figure 15). The innate immune system is transmitted through generations whereas the adaptive immune system can develop memory thanks to immunisation and therefore adjust itself in response to pathogens.

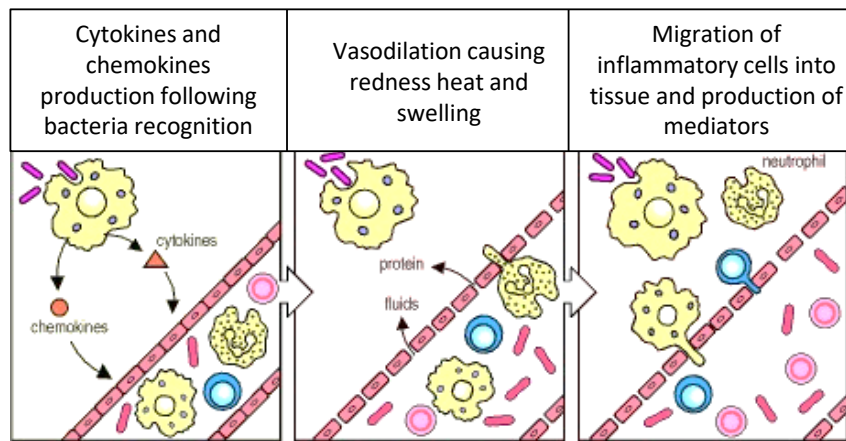
In case pathogens manage to escape from the first protective barrier, they will surely be “trapped” to be completely degraded by an immune system that function properly. The innate immune system intervenes first to recognize the invaders through Antigen Presenting Cells APCs, relying on means like PAMPs (described in the first part of chapter I) including bacterial LPS; it is directed against any pathogens that enter the body.



**Figure 15. Innate and adaptive immune systems facing microbial PAMPs by communicating through immune cells.** Taken from<sup>120</sup>.

Distinguishing self from non-self-patterns is the main mission of innate immune system through activated immune cells to eliminate pathogens. However, in some cases, the innate system cannot handle with pathogen elimination and so the adaptive immune system is activated. The latter works together with innate immunity to provide the strongest resistance to microbes and other foreigners in a specific way<sup>121</sup>. Adaptive immunity is very efficient, it invites specific active components such as white blood cells like APCs, B- and T- lymphocytes to neutralize pathogens. Within this chapter, we will essentially focus on **innate immune response**, involving APCs and PAMPs, adaptive immune regulation aspects will not be detailed.

The innate immunity network is made of many immune cells among them **Macrophages**, **Dendritic Cells** (DCs), mast cells, neutrophils, eosinophils, and the Natural Killer NK cells; and of course, their secreted molecules, such as histamine and cytokines. In case of inflammation due to pathogen invasion, a sequence of events occurs in very well-organized manner.



**Figure 16. Triggering of the immune response by bacterial infection.** Macrophages encounter pathogen invaders and from that step on, a cascade of intercellular events implicates the immune system including secretion of chemokines and cytokines and changes in the stickiness of the endothelial cells of the blood vessels necessary for immune cells migration. Adapted from<sup>119</sup>.

When a tissue becomes infected with a given pathogen, a cell mediated innate immunity response takes place (figure 16). PRRs expressed by mast cells recognize **PAMPs** exposed on bacterial surface and start secreting histamines as inflammation response which in turn increases the permeability of the blood vessels i.e., vasodilation, allowing a fast recruitment of the other immune cells. Similarly, the recognition between macrophages and PAMPs stimulates the release of cytokines which, together with histamines, enable leukocytes (typically Neutrophils) to enter the inflamed area and begin phagocytizing these pathogens<sup>122</sup>.

Pathogen recognition is supported by a second interaction between complement proteins attached to the surface of pathogens and immune cell complement receptors, that helps in forming the phagolysosome inside the cell in which the pathogen is thereafter degraded into small antigens. With this, macrophages take out these antigens to their surfaces (hence their name APCs) to stimulate the adaptive immunity. Dendritic cells and macrophages are APCs that capture then present antigens to the lymphocytes (T cells) using Major Histocompatibility Complex (MHC) molecules and they also produce different cytokines, once activated after pathogen invasion. Some macrophages get infected by the pathogen during the innate immune response and they should be then eliminated by the Natural Killer NK cells through apoptosis<sup>119</sup>.

Cytokines produced by macrophages, including Interleukins II-1b and II-6 and pro-inflammatory TNF- $\alpha$ , stimulates the liver which produce specific proteins, such as mannan-binding lectin MBL, as important component that activate a certain pathway of the complement system. Cytokines are also responsible of uplifting the production of more neutrophils to insure a quick and efficient innate immunity. TNF- $\alpha$ , produced by macrophages, possess a key role in the initiation of the

adaptive immunity by stimulating DCs to migrate into the lymph node. DCs are also able to directly recognize and phagocytose the pathogen, and undergo licensing enabling them to move into lymph node.

### **I-3-2. Role of LPS in the outbreak of bacterial infection**

Many infections are caused by pathogenic Gram-negative bacteria causing bloodstream infections, pneumonia, urinary tract infections, gastrointestinal diseases, and acute bacterial meningitis. It is not surprising to know that Gram-negative cause serious or persistent infections since their capsules and outer membranes prevent immune cells, designed to fight the infection, from ingesting the bacteria and protect them from antibiotics.

As in any host-pathogen relationship, a pathogen must succeed infecting the host cell while conserving its cellular stability, both functions are perfectly ensured by **bacterial LPS**. According to what is known so far and generally, the **first contact** with innate immune system arms is accomplished through potent PAMPs and available specific PRRs expressed by immune cells at the cell membrane. For example, LPS glycoconjugate from Gram-negative is recognised by TLR4, the so-called Toll-like Receptor belonging to PRRs family<sup>123,124</sup> and this recognition mechanism is detailed in the next part of this chapter. Mutational studies by substituting certain PAMPs expressed by pathogens led to different pathways activations depending on the TLR being involved. TLR4 for instance, which in principle recognizes LPS, can also capture a specific fimbriae from the same bacteria but this requires the participation of a co-interaction with glycosphingolipids<sup>125</sup>.

PAMPs recognition by PRRs is pathogen-specific and it also depends on PRRs expression level by immune cells, PAMPs cell surface distribution and concentrations and, importantly, the invading strain (beneficial, commensal, or pathogenic)<sup>126</sup>. Chiefly, PAMPs recognition by PRRs relies on many factors that include the accessibility of PRR to the specific PAMP and, the manifestation of other present virulence factors can stimulate PRRs activities<sup>125</sup>.

#### **I-3-2-1. LPS induced sepsis**

After pathogen invasion or breakdown of a strain tolerance, LPS recognition, by macrophages and dendritic cells, leads to phagocytosis and degradation. Since then, LPS molecules are released in the primary site of infection and could even move to sterile tissues thus triggering new inflammations. The most known LPS-mediated syndromes are **sepsis** unluckily

followed by **septic shock**, and **healthy cells dysfunction**, irrespective of the involvement of immune cells.

Sepsis takes place when an infection is not cured, and it is manifested by high disorders at different levels which may lead to death. Reinaldo and colleagues have extensively studied the role of LPS during sepsis, at different levels i.e., genetical and cellular<sup>127</sup>. They observed various immune response outcomes from both LPS tolerance and hypersensitivity including dramatically decreased TNF- $\alpha$  cytokine production in monocytes cells for patient with either severe sepsis or septic shock, with respect to healthy persons. PRR expression level was evaluated besides and, a little increase in TLR-4 and TLR-2 expression was observed few hours after LPS administration in healthy patients. Conversely, production rate of these TLRs was not changed in sepsis patients. However, the same treatment either with LPS or with *P. aeruginosa* pathogens enhanced the production of reactive oxygen species ROS and nitric oxide NO (that amplify immune response) for those patients<sup>128</sup>. All these events are modulated in septic patients after LPS induction especially that low doses of LPS triggers cells death like human monocytes<sup>129</sup>. Finally, many informations related to the pathophysiology of sepsis could be concluded from the mentioned LPS studies to improve the care of septic patients.

#### **I-3-2-2. Healthy cells alterations following LPS induced sepsis**

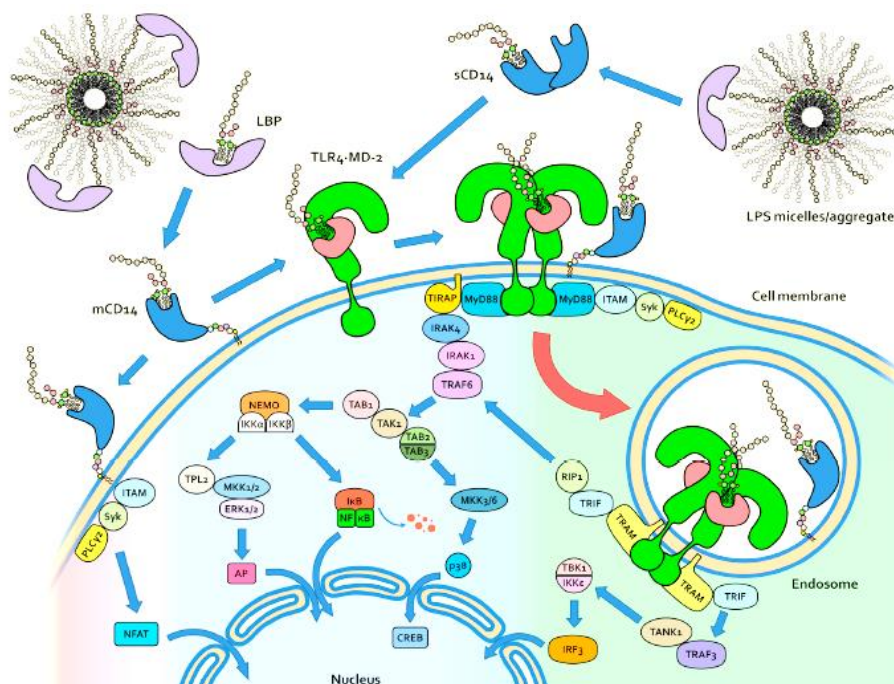
Several studies have reported the key role of freely circulating LPS in inducing an inflammatory response. Yucel and al have described LPS capacity to induce **cardiovascular cells dysfunction** in patients with septic shock<sup>130</sup>. The authors demonstrated that LPS is the major inducer of septic cardiomyopathy by using in their studies an endotoxin-induced inflammatory model. They showed that LPS administration of high doses to human cardiac myocytes induce an excessive expression of cytokines (TNF- $\alpha$ , IL-1 $\beta$ , IL-6, CCL2, CCL5, IL-8) and pro-inflammatory factors (IL-10 and IL-6) resulting from exaggerated apoptosis and leading to cells injury besides. Complementarily, effects of LPS on channel expression levels in cardiac myocytes were also evaluated showing changes in calcium, sodium, and potassium currents upon the addition of LPS, thus suggesting **cellular electrical alteration**.

#### **I-3-3. Recognition of LPS by the host cell**

The LPS recognition process in mammalian hosts is initially mediated by LBP/CD14/TLR4-MD2 interaction system. That is, LPS interaction with Toll-like Receptor 4 (TLR-4) and its associated proteins, which are all expressed by immune cells such as monocytes, macrophages and dendritic cells. This recognition mechanism leads to conformational changes of TLR4



receptor which accelerates afterwards LPS-induced signalling pathways essentially MyD88 (myeloid differentiation primary response 88) at the plasma membrane and the TRIF (TIR-domain-containing adaptor protein inducing interferon- $\beta$ ) pathway that takes place after LPS internalization and endosome formation<sup>131,132</sup>. Many immune cells actors are recruited for LPS sensing and for intracellular signalling leading to nuclear factors activation and production of inflammatory cytokines (Figure 17). For simplicity sake, these intracellular signalling events are not described in this manuscript. This part of the chapter is meant to describe: LPS molecular recognition by TLR4 and related proteins complexes; and cytosolic LPS interaction with intracellular proteases, the so-called caspases, which takes place generally during late stage of infection<sup>132</sup>.

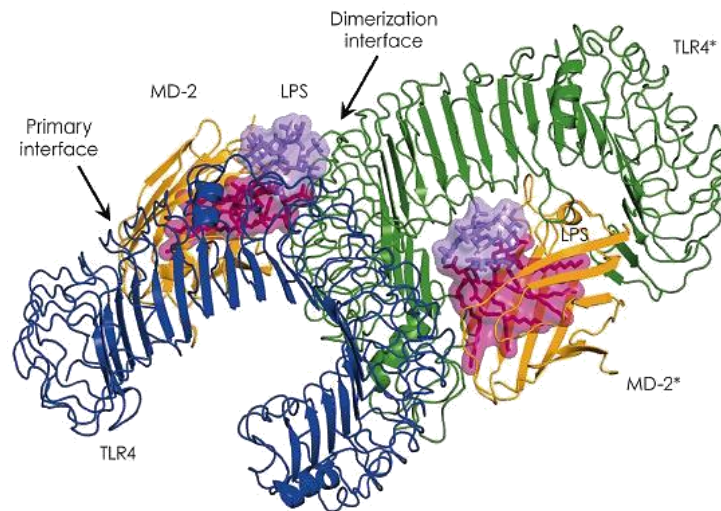


**Figure 17. TLR-4 activation by LPS and intracellular signalling pathways involving several effectors.** Taken from<sup>133</sup>.

### I-3-3-1. LPS Recognition through LBP/ TLR-4/ MD-2 system

The current LPS internalization view, concluded so far by several research groups, is that LPS is initially extracted from bacteria aggregates and transferred by the Lipopolysaccharide Binding Protein (LBP) to the cluster of differentiation 14 (CD14) which in turn transfer the LPS monomer to the Toll-like receptor (TLR-4)/ myeloid differentiation factor (MD-2) heterodimer<sup>134</sup>. LBP binding to LPS aggregates in strong and occurs at the N-terminal domain of LBP whereas CD14 binds at the C-terminal part of the elongated LBP. This structurally orchestrated interaction is important for physical carry and transfer of LPS to TLRs<sup>135</sup>.

TLRs, in response to LPS loaded by CD14, form **canonical inflammasome** complexes for activating a certain pathway afterwards. Lee research groups proposed the interaction mechanisms between *E. coli* LPS and the human heterodimer TLR-4/ MD-2, by establishing the crystal structure of the complex, showing a multimer composed of two copies of the TLR4–MD2–LPS complex arranged symmetrically with multiple structural components involved in LPS recognition<sup>136,137</sup> (Figure 18).



**Figure 18. (TLR-4/MD-2/LPS)<sub>2</sub> complex structure showing the primary and dimerization interfaces. Pdb: 3FXI.**

From a structural point of view, TLR-4 receptor naturally adopts a particular horse-shoe shape. Before binding to agonistic LPS, this structure firstly binds to MD-2 forming a primary interface of the complex. The second interface appears following dimerization of the complex when the TLR-4/MD2 dimer binds to the LPS, thus forming (TLR4–MD2–LPS)<sub>2</sub> complex. The authors compared then the behaviour of these receptors upon the interaction with antagonistic LPS (bearing less fatty acid chains) and observed remarkable versatility of the ligand recognition mechanisms by TLR-4/MD2. Figuring out the molecular mechanisms behind different LPS activation processes is essential to distinguish toxic from non-toxic LPS/Lipid-A variants activities. The TLR-4/MD-2 heterodimer has strict ligand specificity, but it can apparently be activated by diverse LPS structures since it has been demonstrated that minor structural changes in derivatives of LPS can abolish their endotoxic capability<sup>138,139</sup>. Still, this picture of TLR-4/MD-2 complex being the sole receptor for LPS molecules was doubted since many groups were able to show unexpected responses to their LPS preparations that were independent of TLR-4<sup>140</sup>. In fact, it has been shown that LPSs from *Porphyromonas gingivalis* and *Leptospira interrogans* are recognised not by TLR-4, but by TLR-2<sup>141</sup>.

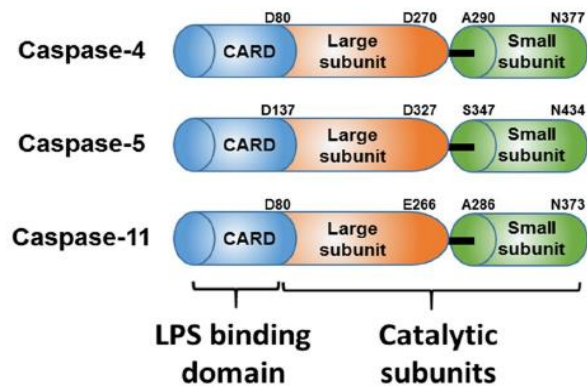
As a reminder, native LPS derivatives originate from Lipid-A modifications and from carbohydrates composition. Cochet and colleagues have shown the importance of Kdo and heptose residues from the inner core LPS as structural motifs that contribute to LPS/TLR4 complex formation<sup>133</sup>. One of our interests to this PhD project revolves around the understanding of how slight structural changes, primarily in the detailed composition of the saccharidic parts of LPS molecule, influences its binding to immune cells receptors like C-type lectins.

### **I-3-3-2. LPS involvement in the caspase pathway, non-canonical inflammasome formation**

Caspases or cysteine-aspartic proteases are enzymes recruited during the immune response for cleaving precursor proteins to create biologically active cytokines. There are also some caspases which have special missions in the regulation of programmed cell death or apoptosis. In this part of the chapter, several studies concerning LPS-mediated inflammasome formation involving cytosolic caspases are presented.

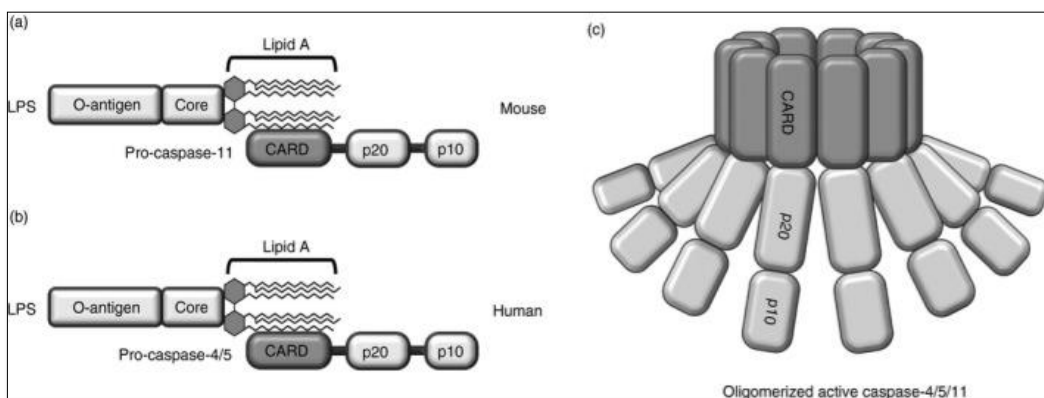
In 1998, Schumann et al have evaluated the induction of cytokine (Il-1 $\beta$ ) secretion by the presence of *Salmonella minnesota Re595* LPS. They showed that the release of mature Il-1 $\beta$  is associated with caspase-1 activation by LPS in peripheral blood monocytes and epithelial cells<sup>142</sup>. No details about possible direct molecular mechanisms between caspase-1 and LPS were declared at that time. Few years later, it has been found that another cytosolic caspase, termed caspase-11, is activated by cytosolic LPS from *E. coli* O111:B4 and leads to **non-canonical inflammasome** activation<sup>143,144</sup>, which is another type of inflammation caused by LPS in addition to canonical inflammasome. A year later, another research group showed that the non-canonical inflammasome activation takes place right after a direct interaction between LPS and caspase-11 and caspase-4 (another cysteine protease belonging to caspase family) through their CARD (caspase activation and recruitment domain) and they suggested that LPS binding induces oligomerization of one of them<sup>145</sup>. Another cytosolic LPS receptor named caspase-5 have been identified and seems to be involved in LPS binding through the CARD domain in human monocytes<sup>146</sup>.

LPS-mediated cytokine release seems to be dependent on LPS direct and indirect interactions with all these cytosolic caspases which could be explained by the architecture of their structures (Figure 19).



**Figure 19. Schematic domain structures of the cytosolic LPS receptors Caspase-4, Caspase-5 and Caspase-11.** Taken from<sup>147</sup>

A recent publication by An et al<sup>147</sup>, presented an updated and new picture of the LPS recognition by innate immunity participants highlighted by the formation LPS aggregates between LPS and human caspase-4. Briefly, they expressed and purified important amounts of monomeric CARD domain variants of caspase-4 in *E. coli* system and they established *in vitro* studies showing that caspase-4 domains bind to large LPS micelles and break them into small molecular weight complexes. Young-Su Yi has published a review<sup>148</sup> which describes molecular mechanisms, based on some paradigms, that occur during LPS interaction by oligomerization with caspases 11, 4 and 5 in response to Gram-negative pathogens infections (figure 20).



**Figure 20. Non-canonical inflammasome formation in humans and mice.** (a) and (b) show molecular bindings from single mouse and human caspases, respectively, to the Lipid-A portion of LPS. The oligomerization of complexes is displayed in (c). p20 and p10 refer to catalytic domains of caspases. Taken from<sup>148</sup>.

The detailed mechanisms of LPS recognition by all these caspases and the events happening during the interaction with LPS at the molecular level remain unclear. From these puzzling observations, our point of view is that this could be due to the structural heterogeneity of LPS

and to the complexity of the “huge” caspases complex which all come over LPS during the immune response.

## ***Chapter II.***

## II. C-type Lectin Receptors

At project start, we believed on C-type lectin receptors' (CLRs or CTLs) capability to interact with ligand containing carbohydrates like bacterial LPS during innate immune response, being part of pathogen recognition receptors PRRs. In addition, those lectins are immune cell receptors that possess structural features which not only stabilize their organisation into the cell membranes but also mediate somehow their interactions with specific and various carbohydrate structures. This chapter is dedicated to an introduction to CLRs as well as the two main human lectins considered in this study. The chapter is also meant for mentioning the importance of structure-function relationship within lectins structural domains and it ends by citing therapeutic applications.

### II-1. Introduction

In 1954, the term lectin, which is derived from the *Latin* verb “*legere*” meaning “to gather or select”, was introduced by William C. Boyd and Elizabeth Shapleigh<sup>149</sup>. By the end of the 19<sup>th</sup> century, researchers started to have the evidence that these lectins are erythrocytes agglutinating proteins. Agglutination of cellular structures or precipitation of glycoconjugates are the two phenomena give rise after lectin binding. Lectins are ubiquitously expressed in mammals, plants, and microorganisms, in different cellular locations, in both membrane-associated and/or free in solution. Acting as a fundamental messenger, lectins that are either released in the extracellular environment (soluble) or expressed as transmembrane receptors (insoluble) on Antigen Presenting Cells (APCs), mediate contacts with other cells<sup>150</sup>. In the context of bacterial infections, they are the sensors of pathogens during host-pathogen interactions (figure 21).

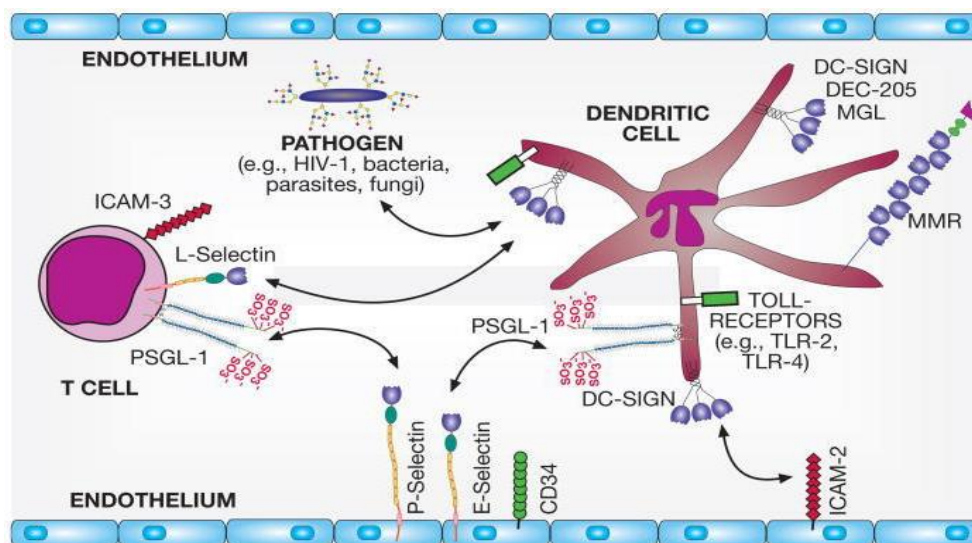


Figure 21. Graphical view of the different interactions mediated by lectins expressed on the surface of dendritic cells, that are cell-cell adhesion and pathogen recognition. Taken from<sup>151</sup>.

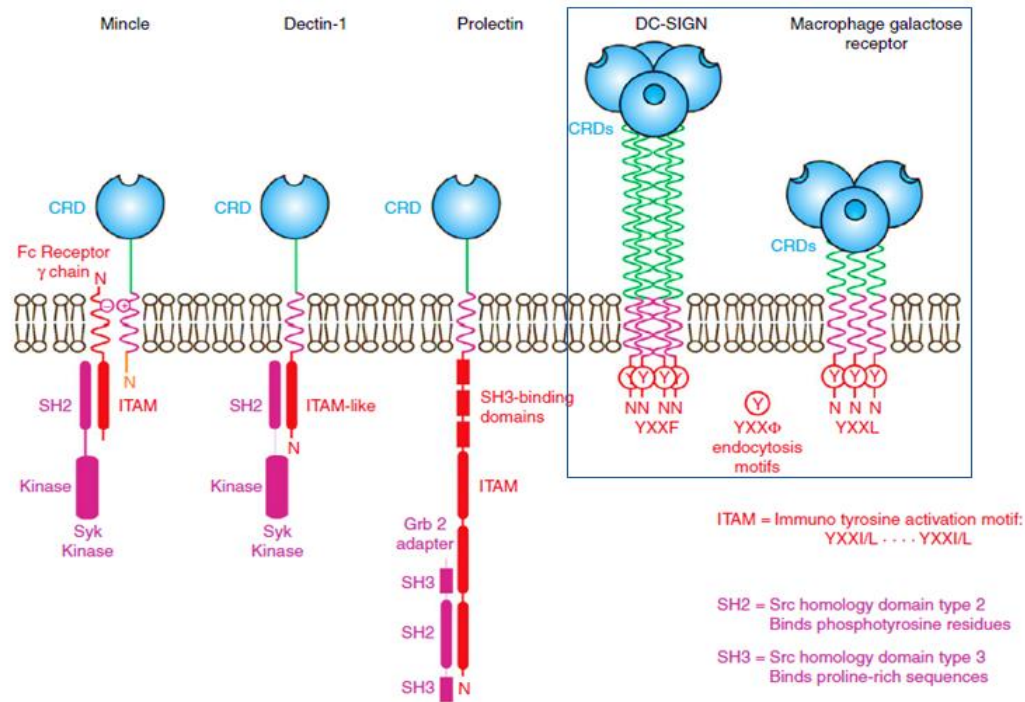
As Pathogen Recognition Receptors (PRRs), CLRs have a glycan binding activity and belong to Glycan Binding Proteins (GBPs) which recognize essentially and specifically various carbohydrate scaffolds from bacteria and viruses<sup>152</sup>. As introduced above, CLRs have common roles (such as adhesion molecules or PRRs) and they are localized on APCs, as membrane receptors or freely circulating, in many tissues in contact with external environment such as epithelial cells of intestines, stomach, and lungs. Immature APCs found in the blood stream also express a variety of CLRs.

Oligomeric CLRs such as circulating Mannan-Binding lectin or MBL possessing a C-type lectin-like domain, interact with carbohydrate-containing ligands depending on their secreted forms. In other words, the overall oligomerization state influences MBL binding to microbial carbohydrates. In fact, tetrameric human MBL exhibits slightly higher affinity and increased stability level (through an avidity phenomenon) than trimeric MBL, when bound to specific carbohydrate-containing ligand<sup>153</sup>, despite their identical binding profiles to associated proteins or MBL-associated serine proteases (MASPs). The increased binding stability provoked by MBL structure organisation (tetramer vs. trimer) might empower the resulting signalling pathway (e.g., complement system activation). In a similar context of competitive interactions, plant lectins like Wheat germ agglutinin (WGA) inhibit monoclonal antibodies binding via competitive bindings to cell surface carbohydrates<sup>154</sup>.

In addition to their functions as adhesion molecules and receptors for pathogens, CLRs participate in the signalling pathways during the immune response<sup>155</sup>. In the case of bacterial infections, this function is least well understood for the CLRs of interest (i.e. **DC-SIGN** and **MGL**)<sup>156</sup> and will not be discussed in this manuscript. Still, DC-SIGN is certainly believed to act as an endocytic receptor that deliver the up-taken ligands to lysosomes<sup>151</sup>.

The endocytosis pathway is exclusively dependent on the bound ligand (partially or completely glycan-based), MGL might also display the same activity as it has a cytoplasmic endocytosis motif as DC-SIGN (figure 22). Importantly, pathogen recognition via specific sugar binding at the cell surface, for instance by lectin conformational changes<sup>157</sup>, would potentially lead to various signalling events on the cytoplasmic side of the immune cell membrane.

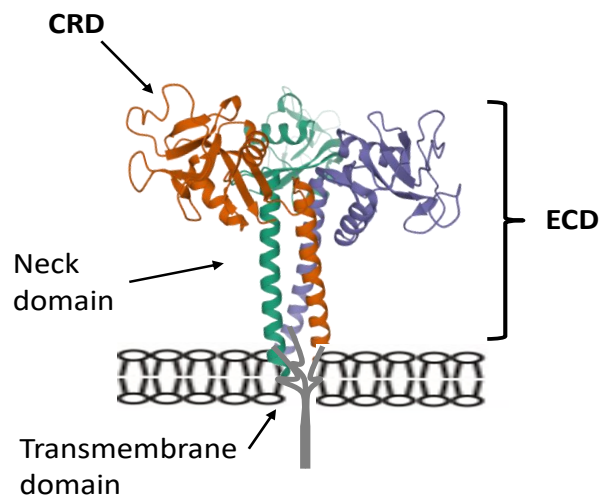




**Figure 22. Cartoon representation of the distinct domains of some CLRs round the immune cell membrane showing direct and indirect interactions with signalling effectors during innate immune response to a pathogen.** A frame is added to the original figure, showing CLRs are the C-type lectins studied within this research work. Taken from<sup>155</sup>.

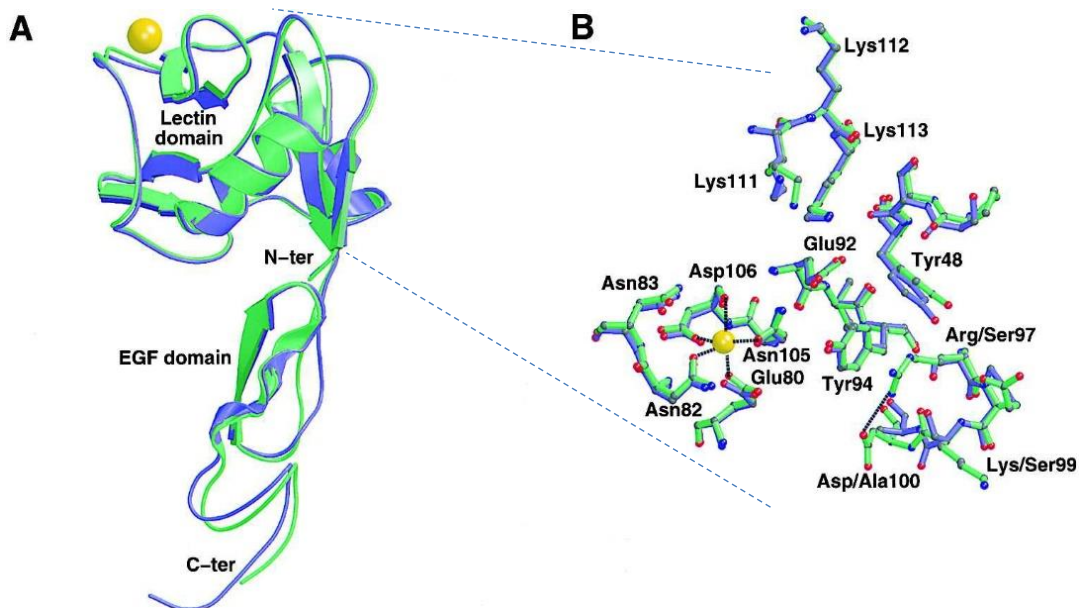
**II-1-1. Molecular organisation of C-type lectins: the ECD and CRD domains**

Pathogen recognition processes, prior to signalling reaction cascades are established by immune cells (see Chapter I-3.). They are localized around cells surfaces and governed by immune cell receptors including different types of lectins. The structural organisation of CLRs as transmembrane receptors is structured as an **Extracellular Domain (ECD)** extended to the external environment. ECD is constituted of a **Carbohydrate Recognition Domain (CRD)** linked to the cytoplasmic domain through a neck domain. Figure 23 represents a ribbon representation of a trimeric CLR.



**Figure 23. General organisation of C-type lectin receptor on the membrane of myeloid cell.**  
Adapted from pdb 1RTM.

CLRs recognize their carbohydrate ligands in a calcium ( $\text{Ca}^{2+}$ ) dependent way for most of them. It is noteworthy to say that the structural diversity together with the phylogenetic distribution and various cellular locations and functions of CLRs, make it hard to define certain criteria for their classification. Paradoxically, the classification of CLRs occasionally misses the criteria of calcium dependency knowing that for some C-type lectins the presence of a calcium binding site is not necessary for carbohydrate binding<sup>158</sup>. Even so, this classification is maintained thanks to the high structural similarity of the CRDs among the C-type lectins. As for the ECD of the C-type lectins, it appears that there are only few X-ray data and computational modelling available<sup>159</sup>, and an example is shown in figure 24.



**Figure 24. An example of the entire structure of human C-type lectin ECD, P-selectin.** P-selectin constructs, derived from X-ray data, include CRD or lectin domain (B) (residues close to Lewis<sup>x</sup> binding

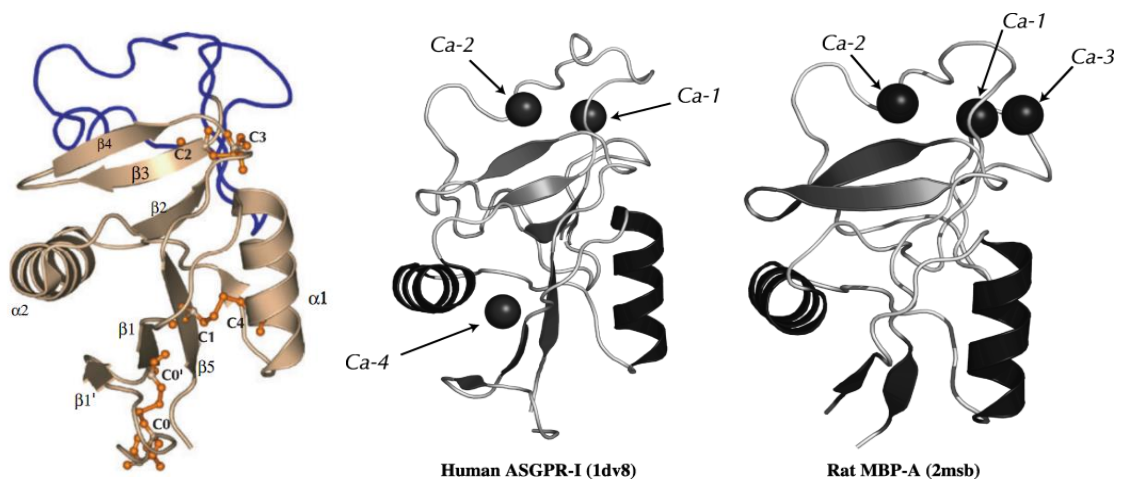
site are depicted from P-selectin superimposed with E-selectin), epidermal growth factor-like (EGF) domain and transmembrane domain (A). Calcium ion is highlighted in yellow. For clarity, P-selectin ribbon representation (blue) shows overall similarity with its homologous E-selectin (green). Taken from<sup>159</sup>.

### **II-1-1-1. The Carbohydrate Recognition Domain, a crucial C-Type Lectin element**

The mechanism of action of CLRs is based on several types of possible interactions such as binding mediated by salt bridges, aromatic stacking, van-der-walls interactions and hydrogen bonding and metals coordination. The implication of metal groups in the binding/recognition mechanism, like calcium ions (Calcium dependent binding for “C-type” lectins), is strongly dependent to the **Carbohydrate Recognition Domain (CRD)** structure, defining certain binding motifs. In fact, the CRD is more generally called as the C-type Lectin-like Domain (CTLN) since there is no strictly attributed interaction mechanism or binding mode specific to the CRD of CLRs even with structural homology<sup>158,160</sup>. But for common usage, the term “CRD” will be utilized in this manuscript instead of “CTLN” when mentioning the carbohydrate recognition domain.

The functional versatility of CRD scaffolds, within this superfamily of lectins, is described in a review written by A. Zelensky and J. Gready<sup>161</sup>. CLRs share in their CRDs **common secondary structure elements** including the terminal  $\beta$ -strands organized into two anti-parallel  $\beta$ -sheets and edged by two  $\alpha$ -helices, and the four most conserved cysteines forming the salt bridges in between the loops. CRD structures have up to four canonical binding sites (figure 25).

The **number and architecture of binding sites** may vary but at least one canonical binding site is usually found among the CRDs. This calcium-dependent binding site is responsible for carbohydrate binding because it harbours specific motifs like the **EPN** (Glu-Pro-Asn) motif that leads to the recognition of mannose, fucose, glucose-containing ligands whereas another motif namely **QPD** (Gln-Pro-Asp) targets N-Acetylated galactose, galactose-type glycans<sup>162</sup>.



**Figure 25. Generic CRD or CTLD domain of animal lectins.** Black and white representations, middle and left structures, display examples of CRDs showing  $\text{Ca}^{2+}$  binding sites locations in human ASGPR and mice MBP-A with their respective PDBs. For clarity, the four cysteines are presented in atomic presentation. Secondary structural elements are described in the text. Taken from<sup>161</sup>.

The existence of various CRD architectures among some of the C-type lectins, with shared and different elements, brings out the questions of how these proteins could distinguish different types of glycan structures in order to help antigen presenting cells to discriminate between harmless (self-antigens) from pathogens? Do they carry different binding pockets (i.e. canonical and non-canonical sites) in their structures?

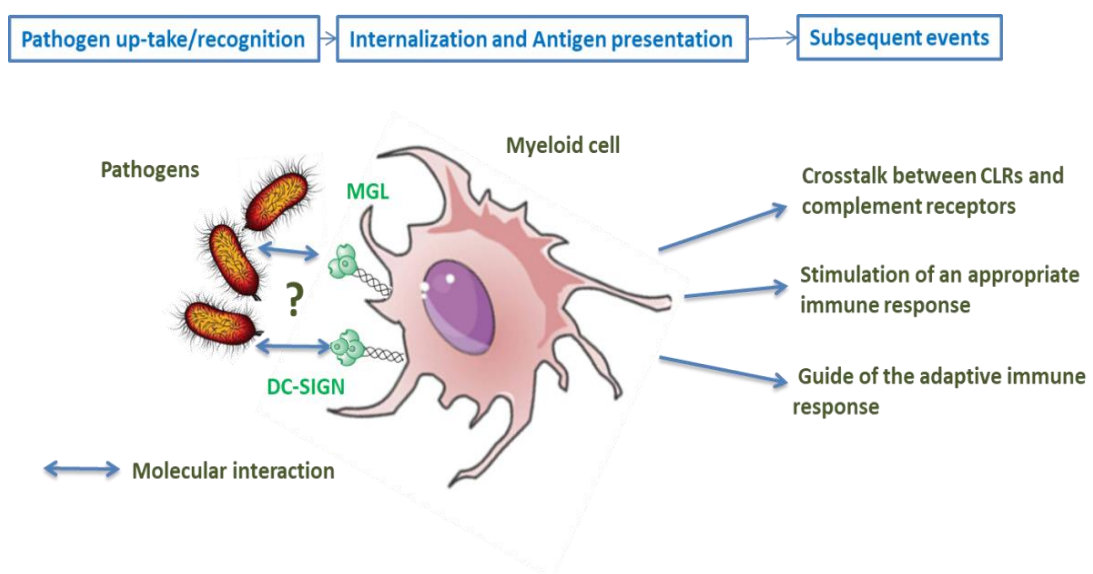
Those interrogations are difficult to answer in a straightforward way because some of the canonical sites, in a receptor with multiple binding sites for instance, may ensure structural stability of the protein and not ligand binding. Else, the evolution of CLR domains among species makes this matter harder to resolve, even if knowing that some well-conserved sugar binding motif do exist<sup>162</sup>. For instance, some CLR like DC-SIGN (a lectin of interest that is introduced in the next part of this chapter) underwent evolutionary “structure-function” changes which result in lack of informations in the orthology between human and mice<sup>163</sup>. Those changes are possibly incurred by the ability of this lectin to recognize glycoproteins from viruses which may have selected for loss or changes in some functional genes in DC-SIGN, due to viruses mutations<sup>164</sup>. However, the fact that canonical binding sites, that intervene in and mediate pathogen recognition, are overall conserved, this might help understanding CLR interactions with pathogens. In addition, ligands binding may be influenced by the structural organization of CLR which is controlled by the neck domain.

Differences in calcium binding sites number and locations, together with the slight variations of potential amino acids in the CRD among human and mice, make their studies particularly

challenging which may give rise to important scientific questions regarding bacterial LPS recognition by these CLRs in humans. The two human lectins, in which we are interested, are believed to have these structural features based on what has been published so far about their ligand specificities described in the next part of the chapter.

### II-1-2. C-type lectins of interest: structural and functional aspects

This part of the chapter focuses on two human CLRs the most considered in my PhD, **DC-SIGN** and **MGL**. From biological and immunological standpoints, these two lectins are believed to mediate common physiological roles in maintenance of homeostasis, cell-cell adhesion, endocytosis, antigen uptake and host-pathogen interaction<sup>165</sup>.



**Figure 26. Graphical overview of the recognition of pathogens by the two CLRs of interest and downstream stimulation of immune response.**

Being part of the same group of C-type lectins i.e., Type II **transmembrane receptors**, these two membrane-bound lectins showed a significant relevance in development and progression of many diseases such as viral infections and cancer related diseases that are not studied in this PhD project, and **bacteria related diseases** (for **DC-SIGN** and **MGL**) directed to the center of our attention (figure 26).

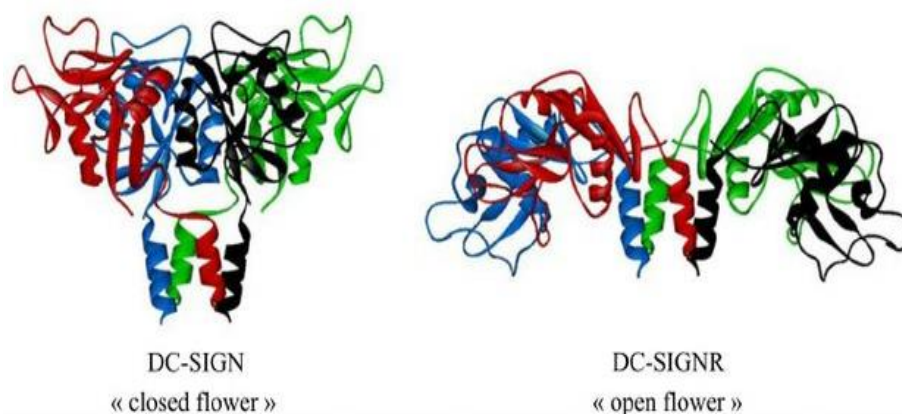
From a structural perspective, DC-SIGN and MGL share common points especially regarding the overall flower bouquet conformation of their ECD and CRD structures which are discussed below. These two CLRs possess similar neck domain role in controlling the overall structure oligomerization. Importantly, the neck domains of these lectins are responsible for their

extracellular oligomerization and one may think that these structural changes/oligomerization states could also be influenced by the ligand being accommodated.

## II-1-2-1. DC-SIGN lectin

### II-1-2-1-1. Structure and cellular location of DC-SIGN

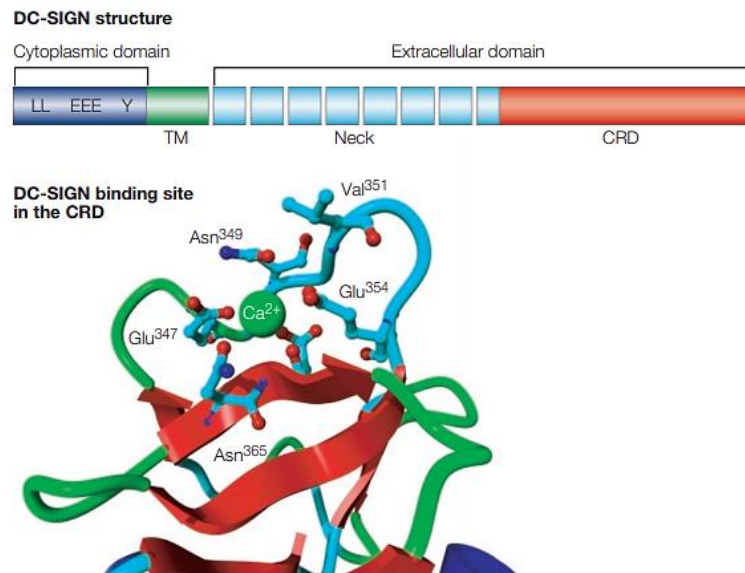
Dendritic-cell-specific intercellular adhesion molecule-3-grapping non-integrin (**DC-SIGN**) also named CD209 (Cluster of Differentiation 209) is a CLR expressed by macrophages and dendritic cells<sup>166</sup>. DC-SIGN has a closely related homolog called DC-SIGNR that is produced in the liver by endothelial cells and in the lymph node<sup>167</sup>, also termed Liver/Lymph node Specific Intercellular adhesion molecule-3-Grabbing integrin (L-SIGN). In fact, the key differences between the two homologs comes from their different structure conformations (open and closed flower bouquet structures) as shown in figure 27. These weak contrast between the two lectins may explain their similar, but not identical, affinities toward certain ligands.



**Figure 27. Organization of the upper part of human ECD DC-SIGN and DC-SIGNR.** Arrangements of CRDs from the two lectins were derived from SAXS and X-ray data of one DC-SIGNR fragment (pdb 1XAR). Taken from<sup>168</sup>.

From the molecular structure view, four homologous C-terminal CRDs of DC-SIGN are linked to the neck domain constituted of 23-amino acid repeats responsible of its oligomerization, thus forming a tetramer<sup>169</sup>. The presence of at least one canonical binding site in the CRD structure is essential for carbohydrate recognition by DC-SIGN (figure 28).





**Figure 28. View of the upper surface of CRD DC-SIGN in Apo state and the domain organisation of the ECD.** Four residues (Glu347, Asn349, Glu354 and Asn365) that bind a calcium ion from the second binding site, are responsible for carbohydrate local recruitment. Taken from<sup>170</sup>.

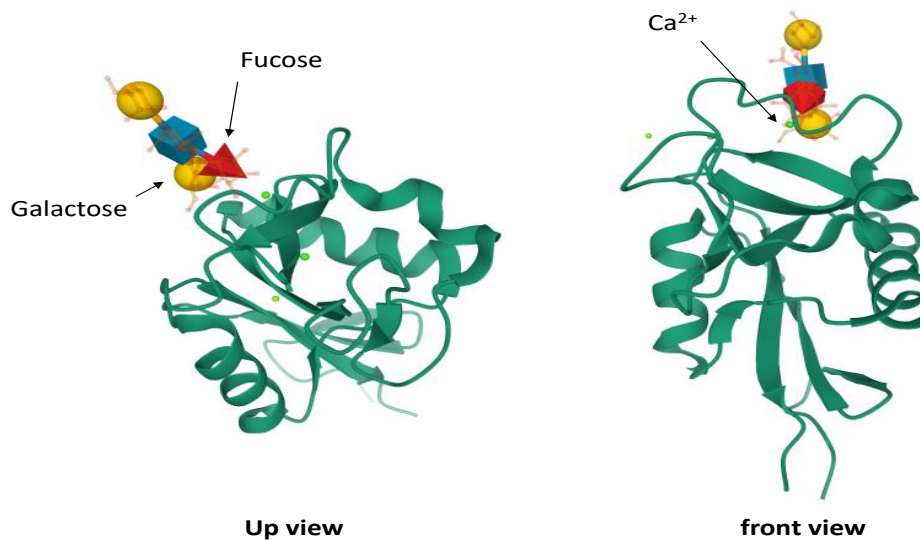
## II-1-2-1-2. Biological functions of DC-SIGN

As its name indicates, this CLR has the highest affinity to ICAM-3 glycoprotein, an adhesion molecule expressed by all leukocytes<sup>171,172</sup>. ICAM-2, another glycoprotein expressed on CHO cells, also interacts with DC-SIGN<sup>173</sup>. DC-SIGN lectin acts as an adhesion molecule and as a PRR at once, by employing its structural and functional capabilities to interact with different types of self and non-self molecules. This CLR has multiple functions that take place during bacterial infections such as primarily the recognition of PAMPs, the up take of these microbial signatures and the stimulation of signalling pathways thus influencing DC functioning for the destruction or clearance of pathogens, even knowing that some of them are capable of exploiting DC-SIGN functions to subvert the immune response<sup>174</sup>. DC-SIGN is one of the most studied CLRs due to its important implications in many diseases that might be explained by the relatively high number of binding sites exposed by this lectin, and many crystallographic structures have been determined for DC-SIGN in complex with several ligands.

### II-1-2-1-2-1. DC-SIGN carbohydrate ligands

DC-SIGN from *Homo sapiens* has been extensively investigated for its biological significance, by acting as both, an adhesion and pathogen recognition molecule. Many relevant glycan ligands of human DC-SIGN have been identified including blood group antigens, high mannose-containing glycans, Lewis<sup>X,Y</sup> structures and Ebola and influenza viruses related antigens. Generally, the bound conformation of mannose and fucose epitopes from Lewis structures was

believed to form compact and rigid complexes (figure 29) that do not necessarily exhibit conformational changes upon binding to the receptor<sup>175</sup>. However, it has been demonstrated that for Lewis<sup>x</sup>-bearing structures, DC-SIGN adopts different conformations while binding to those ligands. This might even involve remote interactions such as ligand bridging between two CRD molecules or clustering of DC-SIGN CRDs, for instance, via the hydrogen bonding between galactose of Le<sup>x</sup> and Glu286 of the adjacent DC-SIGN monomer<sup>176,177</sup>.



**Figure 29. Complex formation following DC-SIGN CRD binding to an oligosaccharide containing Lewis<sup>x</sup> determinant at the nonreducing end (LNFPIII).** Carbohydrates are represented by universal sugars codes and DC-SIGN CRD 3 structure is displayed in cartoon representation. PDB: 1SL5.

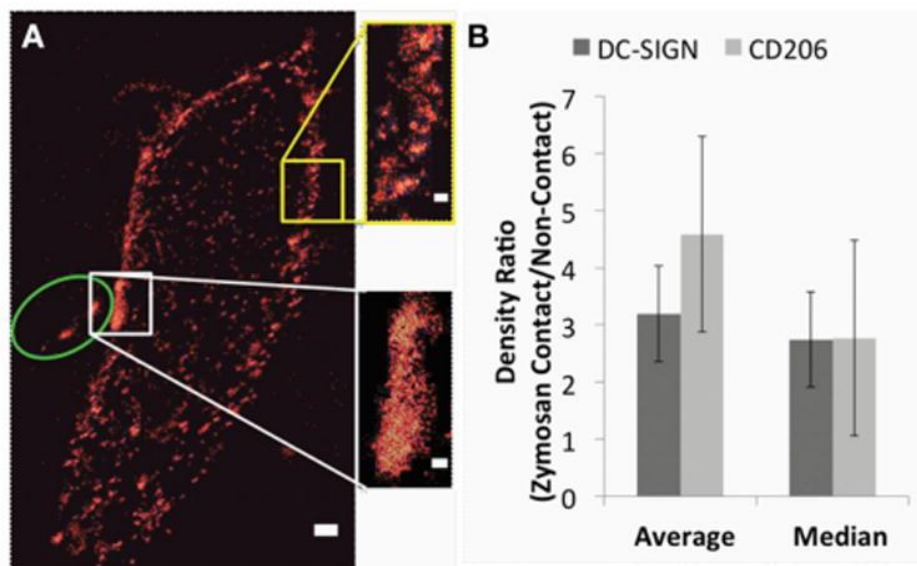
The calcium-dependent interaction between this receptor and its sugar-bearing ligands is mainly oriented thanks to its **EPN motif** (Glu-Pro-Asn) that is, as said, specific for binding to mannose, fucose, N-acetyl-glucosamine and glucose residues. The binding of high mannose (Man) glycan to DC-SIGN (CRD) lectin was studied by molecular docking and NMR spectroscopy. The authors assumed that, according to the docking results, DC-SIGN recognizes Man $\alpha$ 1,2Man moieties via two potential binding sites including the Ca<sup>2+</sup> binding as the main interaction area displaying the most stable complex, totally in agreement with crystallographic data<sup>178</sup>. One more, the binding ability of DC-SIGN (ECD) lectin to fluorinated and natural sugars (Gal, Man, Glc and Fuc) has been elucidated by Molecular Docking simulations and <sup>19</sup>F-NMR spectroscopy<sup>179</sup>. The authors suggested a new binding epitope where the Man residue strongly contributes to the binding.

Via specific recognitions, DC-SIGN binds to carbohydrate ligands from different origins (viral, bacterial, and endothelial cell epitopes) and with various natures (glycoproteins and antigens).



As a PRR, DC-SIGN is able to mediate interactions with *M. tuberculosis*, especially for high mannose oligosaccharides from *Mtb* cell wall PAMPs (i.e., Man-LAM) that have been extensively studied as specific targets for this lectin<sup>170,180</sup>.

As an example study, DC-SIGN lectin was investigated by super-resolution microscopy during fungal recognition<sup>181</sup>. Dendritic cells imaging enabled the localisation of this lectin together with another CLR i.e., Mannose Receptor or MR (CD206) in the cell membrane while interacting with zymosan structures (with repeating glucose units) by using fluorophore labelling (figure 30). Interestingly, the high density of DC-SIGN observed at the fungal binding site, 70% of area increase and 38% of decrease in DC-SIGN domains separation, may results from the avidity phenomenon exhibited by the tetrameric lectin following the interaction.



**Figure 30. Super-resolution images and density ratios of DC-SIGN and MR lectins during interaction with microbial PAMPs.** (A) cell surface densities of DC-SIGN and MR CD206 (white box) in proximity of fungal contact site (green circle) and non-contact site (yellow box) and (B) the density ratio of both lectins (labelled with antibodies conjugated with AlexaFluor<sup>647</sup>) bound to glycan motifs of the zymosan particles over their densities at non-contact site. Taken from<sup>181</sup>.

Apart from bacterial DC-SIGN ligands, viral particles from HIV, Ebolavirus, SARS and hepatitis C viruses have been the most studied as binders of this CLR<sup>182–185</sup>. The recognition of these viral pathogens by DC-SIGN is a crucial step for pathogen internalization and signalling pathways activation. However, it has been reported that HIV-1 virus, for instance, could alternatively interact with other dendritic cell receptors through its glycoproteins (gp120) and perhaps impairs dendritic cell activity by exploiting lectins like DC-SIGN<sup>186</sup>. This is thought to be due to the wide intricate relationship between DC-SIGN and its viral ligands because this depends on the amount

of DC-SIGN expressed by DCs and the efficiency of virus replication inside the cell; and more importantly on the susceptibility of DC-SIGN to the viral infection since many other viruses make malignant use of this lectin as is the case for Ebola virus and SARS-Cov virus<sup>184,187</sup>.

Besides its function as pathogen detector, this lectin is also able to interact with endogenous molecules like blood group self-antigens with high fucosylation especially for histo-blood group B over group A antigen<sup>188</sup>. Furthermore, TB. Geijtenbeek research group has demonstrated that DC-SIGN is involved in TLRs (Toll-like Receptors) modulation as a downstream step (through the activation of a kinase-acetylation-dependent signalling pathway) right after pathogen recognition from bacterial, fungal, and viral pathogens<sup>189</sup>.

### **II-1-2-2. Macrophage Galactose-type Lectin (MGL)**

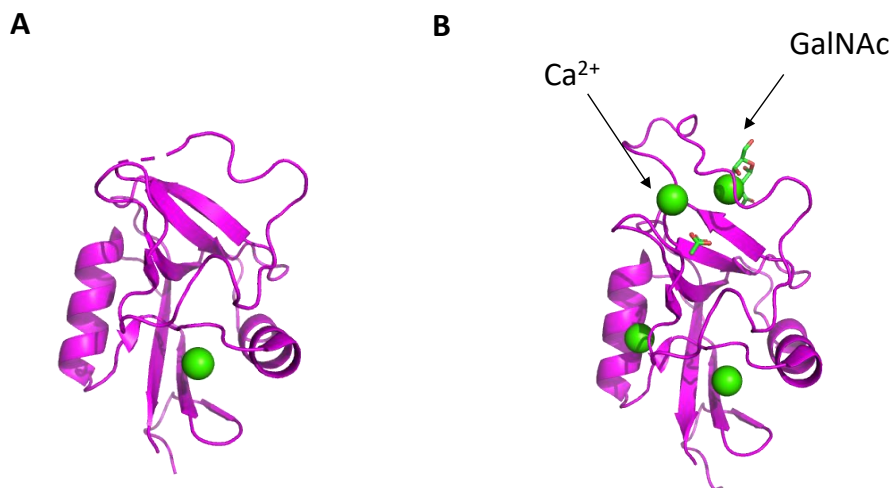
Macrophage Galactose-type Lectin (**MGL**), also called CLEC10A or CD301, is another oligomeric C-type lectin which belongs to group II of CLR family, as DC-SIGN lectin. MGL is expressed on the cell surface of macrophages and DCs from skin and lymphoid organs<sup>190,191</sup>. Although MGL is part of the same group as DC-SIGN, this receptor has single features regarding its function, structure, targeted ligands, and associated diseases. The biological role of MGL has been essentially focused on mediating interactions between endothelial and cancer cells through binding to carbohydrate antigens thus triggering either host-protective or suppressive responses. In antimicrobial immunity, its main role is considered as a PRR that captures pathogens and induces their phagocytosis thus increasing macrophages activation and inducing antigen presentation<sup>165</sup>. MGL is involved in cancer cells development and progression. On the other hand, it acts as dendritic cell activator, thus a dual function between immune tolerance and pathogen uptake is insured by MGL. This lectin could be utilized by pathogens, through interaction with tumor-associated carbohydrate antigens (TACAs), to clear out the immune response. MGL double role also includes prevention from tissue damaging especially when the immune system faces persistent inflammations<sup>192,193</sup>.

#### **II-1-2-2-1. Structural features of MGL**

Human MGL is found in three isoforms originating from immature dendritic cells (isoform-1), macrophages (isoform-2) and mature dendritic cells (isoform-3)<sup>194</sup>. These isoforms possess the same CRDs suggesting similar carbohydrate specificities. The structural differences are represented in a deletion or insertion of a sequence of amino acids in the neck domain<sup>195</sup>. Till the third year of my thesis, only a broad information was found about MGL structure because of the

lack of crystallographic models for describing the architecture of this protein. Chemical crosslinking analyses have shown that the ECD of MGL is organized as an homo-trimer in solution<sup>196</sup>. Each CRD of the three units present in the ECD of MGL harbour a **QPD motif** (Gln-Pro-Asp) at the long loop region that drives MGL recognition toward Galactose (Gal) and N-Acetylated Galactose (GalNAc) residues<sup>162,197</sup>.

In 2020, the crystal structure of CRD MGL was published in the protein data bank (figure 31). However, the related publication is not yet available. According to the PDBs, it seems that GalNAc binding provoked slight conformational changes on the Loop5 secondary structure element of CRD MGL. Collectively, there is a huge lack of molecular details behind MGL binding to bacterial ligands in the literature and this is due to the absence of full knowledge about its 3D structure, especially the ECD construct.



**Figure 31. Human CRD MGL crystallographic structure before (A) and after (B) complex formation with GalNAc residue.** Respective PDBs: 6PUV and 6PY1.

#### II-1-2-2-2. Ligands preferences for MGL

In general, C-type lectin fold is unique but universal since its changeable structure is adjustable for many uses. In humans, there is only one gene that expresses MGL whereas in mice there are two genes responsible for MGL expression (MGL1 and MGL2)<sup>198</sup>. Here again, despite that the specific motif QPD is conserved throughout MGL genes, it is noteworthy mentioning the evolutionary changes in MGL structures since this would interfere with ligands recognition flexibility.

Some studies about the recognition between MGL and its binders, such as Tn Antigen-like epitopes, have considered the strong structural similarity between MGL and its related-protein Asialo-GlycoProtein Receptor (ASGPR), due to the lack of structural informations<sup>199</sup>. Two decades

ago, human carcinoma-associated epitope like Tn antigen or glycopeptides (GalNAc-Ser/Thr), has been found to specifically bind to human MGL, presenting the major example of interaction studies related to this CLR<sup>200</sup>. Since then, other sugar containing antigens like glycopeptides and oligosaccharides progressively have started to show up in the list of human and murine MGL ligands by means of several investigations<sup>201,202</sup>.

Marcelo and co-workers have investigated the binding between human MGL and free Gal, GalNAc and Tn-glycopeptides. According to their analyses, Gal and GalNAc residues both bind to MGL with distinct affinities and binding modes<sup>203</sup>. The highest affinity was estimated for GalNAc with a Kd in the low micromolar range, followed by Tn-glycopeptides that showed strong binding to MGL dependent on amino acids neighbouring. Few years later, the same group has compared the binding profiles between human MGL H259T mutant (from MGL isoform-1) and Tn-glycopeptide ligands, by using glycan micro-arrays. They found out that this mutation affected MGL binding, highlighted by decreased affinities towards Tn-glycopeptides, reflecting the importance of this binding residue i.e., H259 for the interaction<sup>204</sup>. This Histidine residue is missing in the CRD of mice MGL and there is a Threonine instead that binds to the fucose of Lewis<sup>x</sup> structure (Gal-Fuc-GlcNAc) along with QPD motif (GalNAc binding site) binding to Gal moiety.

Clearly, human MGL sugar preferences are exclusively assigned to **galactose derivatives**, especially for **GalNAc** residues since it has been shown that ligands harboring high number of those sugars, exhibit stronger binding with an affinity reaching the nanomolar range<sup>204–206</sup>. However, very few outcomes have been found so far regarding MGL interaction with bacterial ligands and especially from molecular events, that may occur during ligands binding to any possible conformation of MGL (i.e., ECD and CRD).

### **II-2-2. Other members of C-type lectins family**

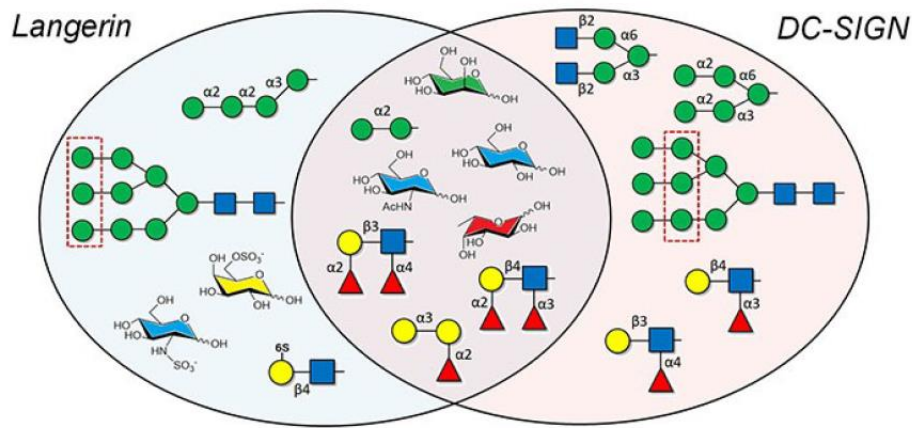
C-type lectins commonly exhibit similarities in terms of ligand specificities, binding scenarios, and immune response contributions. Table 1 summarizes general informations about a number of thoroughly studied human CLRs. The wide variety of the monosaccharides that are exhibited by various ligand structures in nature increases the interest of studying different CLRs functions. In the same line, CLRs structures are adapted to recognize those ligands in a specific manner.

**Table 1. Overview on some examples of transmembrane C-type lectins.**

Lectin	Oligomeric state	Cell type	Ligands	Binding motif		Examples of Pathogens	Ref
				QPD	EPN		
<b>Langerin (CD2017)</b>	Trimer	Langerhans Cells	$\beta$ -glucan*, GlcNAc, Sulfated sugars, GAGs*, Heparin	-	+	<i>Staphylococcus</i>	207 208 209
<b>Dectin-1 (Clec7A)</b>	Tetramer (ligand-induced)	Leukocytes, DCs*	$\beta$ -glucan*, Laminarin	+	-	<i>Fungal species*, Laminaria</i>	210
<b>Mincle (Clec4E)</b>	Dimer monomer	Macrophage DCs, B-cells, Neutrophils	Glycolipids, TDM*, Trehalose	-	+	<i>Mycobacteria Candida</i>	58 211

\* Glycosaminoglycans for GAGs, Dendritic cells for DCs, trehalose-6,6'-dimycolate for TDM, mannose, fucose and glucose containing polysaccharides for  $\beta$ -glucan, *Aspergillus*, *Candida* and *Saccharomyces* for fungal species.

Figure 32 sums up specific and common sugar scaffolds for **Langerin** and **DC-SIGN**. Langerin is another C-type lectin from the same group of DC-SIGN and MGL. The expression of this CLR is restricted to Langerhans cells<sup>212</sup>. As previously discussed, DC-SIGN is partially involved in spreading of the virus by favoring its attachment and internalization whereas Langerin seems to inhibit HIV-1 transmission<sup>213</sup>. It is thought that both CLRs are implicated in HIV-1 infection although if they are believed to have opposed roles. In addition to viral particles, Langerin can anchor antigens from pathogenic bacteria and fungi<sup>214,215</sup>.

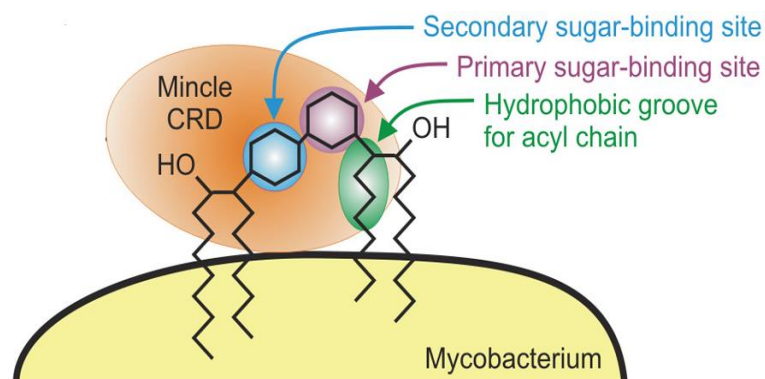


**Figure 32. Similarities in sugar preferences of DC-SIGN and Langerin.** Taken from <sup>216</sup>.

### II-2-2-1. Common binding modes among CLRs, an example of multivalent interactions

As an example of CLRs binding modes, a model of *Mycobacterium bovis* TDM recognition by cow **Mincle** lectin (another CLR) was proposed<sup>57</sup>. The described interaction model supports the existence of an extended surface of binding on the CRD of Mincle (figure 33). In fact, three binding sites with high specificity directed by both sugar and lipid TDM portions, were defined suggesting multivalent interactions that might be needed for disruption of the mycobacterial membrane organization during Mtb infection.

TDM scaffold is constituted of a Glc $\alpha$ 1–1Glc $\alpha$  headgroup linked to two complex branched and hydroxylated acyl chains attached to the 6-OH groups of each of the sugar residues. It has been shown that TDM possess immunostimulatory properties inducing pulmonary granuloma formation and inflammation in mice during the pathogenesis of Mtb infection<sup>217,218</sup>.



**Figure 33. Schematic representation of the proposed model of interaction between cow Mincle and Trehalose TDM from mycobacteria.** Taken from<sup>57</sup>.

Within other studies, it has been found that an Heparin-like hexa-saccharide<sup>208</sup> and deca-saccharide structures<sup>219</sup> bind to **Langerin**, along an extended binding surface, in a calcium

independent manner. Remarkably, a similar interaction scenario has been also observed with **Dectin-1**, the  $\beta$ -glucan receptor, in antifungal immunity<sup>220</sup>.

Commonly, carbohydrate recruitment underneath the CRD of CLR's relies on (i) the local distribution of binding sites and the multivalency of bindings (ii) the length and composition of the sugar-based ligand and (iii) the immunological properties of the two entities which might come from the CLR cellular expression level and/or the ligand pathogenicity.

### **II-3. Applications for the development of sugar-based therapeutics**

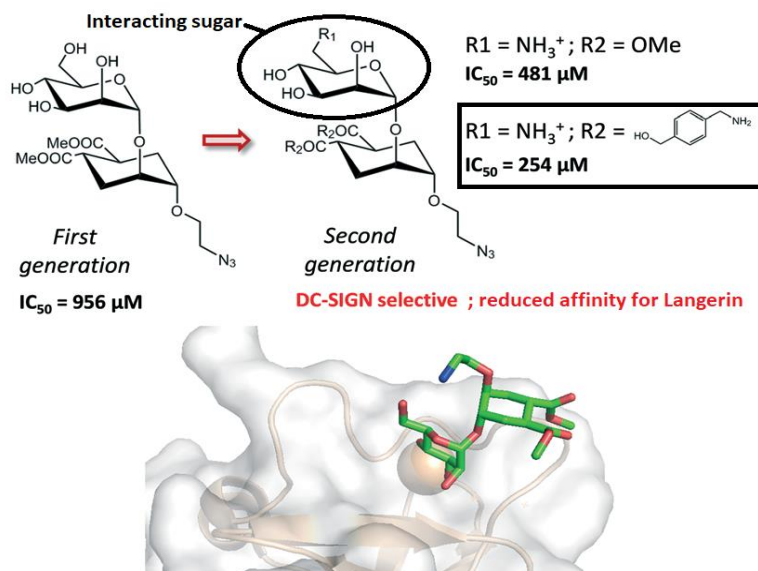
In the previous part, the diversity of C-type lectins functions against pathogens and with partners from immune system during infections was exemplified for some CLR's. It is certainly true that bacterial lectins should be the first resort because of the advantages related to their high production, cheap cost, and straightforward handling, and it is the case. As an example, PA-IL and PA-IIL soluble lectins secreted during *P. aeruginosa* infections, have been used to design and develop glycomimetics<sup>221,222</sup>. For consistency and clarity purposes, this part is meant to cite some examples of promising C-type lectins inhibitors which could be used as therapeutic agents against bacterial, fungal, and viral infections, with cancer besides.

The research period prior to preclinical tests of a certain early discovered treatment for instance, is dedicated to the establishment of laboratory analyses by using several techniques and this stage must be recurrent. It is important to point out the necessity of such period for targeting, understanding every feature, and answering the questions related to the molecules of interest, thus enabling advances in the perception of protein-carbohydrates interactions. To do that, the development of chemically modified carbohydrate structures, the so-called glycomimetics<sup>223</sup>, must be designed for further therapeutic uses against PAMPs related infections. Glycomimetics should be chemically configured in a way that CLR's would be able to recognize them with higher affinities than that for their natural ligands.

Many studies have been established regarding the possibilities of using C-type lectins and related carbohydrate ligands as therapeutics or sugar-based drugs. As an example, **DC-SIGN** tops the first ranks of challenging therapeutic agent for novel drug candidates. As previously mentioned, HIV glycoprotein (gp120) is the famous binder of DC-SIGN mainly by mannose and fucose sugars that decorate the viral glycoprotein surface. Gurney et al<sup>224</sup> have performed in vivo experiments on cells expressing DC-SIGN and the interaction and transfer of HIV-1 particles in human rectal mucosa cells have been assessed at a first stage showing that 90 % of the cells expressing DC-

SIGN bound to HIV-1. They secondly showed that anti-DC-SIGN antibodies successfully inhibited HIV binding up to 50 % in the rectal cells population, meaning that DC-SIGN serves not only as biomarker but it is also a promising functional macromolecule that may indirectly intervene in reducing HIV-1 sexual transmission. Notably, this kind of findings is particularly challenging because it helps in the understanding of cellular dynamics in the host-pathogen interface. In addition, the participation of human CLRs like Dectin-1 and Mincle in immune response to bacterial and fungal chronic infections has been explored, which would eventually help to tailor an efficient therapeutic strategy against pathogens residing in the gut microbiota<sup>156,225</sup>.

Many drawbacks could be overcome by inventing high-affinity and efficient glycomimetics. Importantly, carbohydrate ligand-lectin multivalent interactions are often weak presenting low affinity constants. The binding site is usually extended in that case and so by focusing on those binding regions, the affinities might be eventually improved. There is an interesting study based on structural analyses that helped to identify a modification of first-generation glycomimetics for human DC-SIGN and Langerin that could be potentially efficient well designed therapeutic molecules<sup>226</sup>. The substitution with a positive chemical group (e.g. amino group) at position C6 of the sugar that is considered in the interaction with the two human lectins led to obtaining a specific inhibitor for DC-SIGN (in the micromolar range) whereas the binding was completely destroyed for Langerin (figure 34).

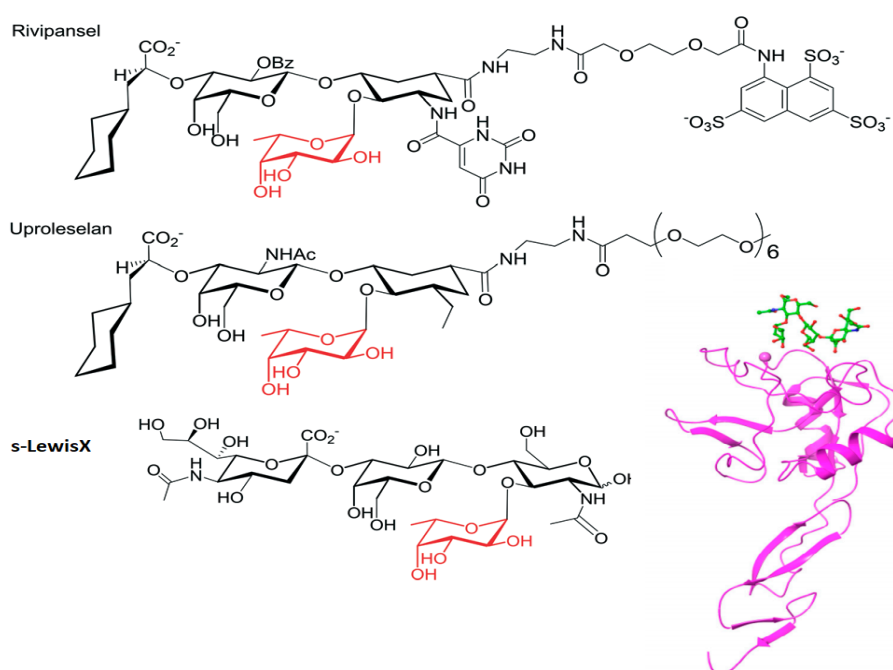


**Figure 34. Strategy used for glycomimetic development as human lectins inhibitors.** The X-ray structure of DC-SIGN (pdb 2XR5) bound to the final second-generation inhibitor is displayed at the bottom. Adapted from<sup>227</sup>.



Still, the new binding affinity was in the micromolar ranges so it must be enhanced. And those findings incited Porkolab and co-workers to actually stick out the idea of investigating multivalent mechanisms<sup>228</sup> where an SPR-based approach was set up to optimize the affinity and the avidity binding mode of surface oriented human CLRs like DC-SIGN, reaching a nanomolar range affinity of DC-SIGN to the studied glycoconjugates (e.g. glycoproteins and glycoclusters).

Selectin (or CD 62), another type of CLRs is a single-chain transmembrane receptor that function as cancer cells adhesion molecule by recognizing sugar moieties through calcium-dependent binding pocket. Sialylated fucosylated glycans like sialyl Lewis<sup>x</sup> (sLe<sup>x</sup>) have been used as templates because they are the main binders of selectins and the trigger of acute inflammation processes. As an example of relevant discoveries that contributed to the success of the development of new therapeutics useful for the treatment of leukemia, late-stage trials for “Rivipansel” and “Uproleselan” selectin inhibitors are ongoing<sup>229,230</sup>. Rational design of glycomimetics bearing sLe<sup>x</sup> like scaffolds (figure 35) and presenting sugars (like fucose and sialic acid) composing Le<sup>x</sup> in a multivalent way, has led to the setup of a library of vast number of sugar-based macromolecules with enhanced drug-like properties. Moreover, patients that have experienced the treatment of asthma and chronic obstructive pulmonary disease (COPD) with “Bimosiamose”, a synthetic inhibitor, have shown encouraging results<sup>231</sup>.



**Figure 35. Chemical structure of selectin (P-selectin, PDB 1G1R) bound to sLe<sup>x</sup> and its inhibitors rivipansel and uproleselan.** Fucose sugar essential for the binding process is depicted in red. Adapted from<sup>227</sup>.

Commonly, the design of such therapeutic molecules revolves around optimizations of approaches and chemical modifications on the desired product (e.g., inhibitors for CLR), thus aiming to reveal a new compound with a comparable avidity and an affinity higher than what we observe at the biological cell surfaces.

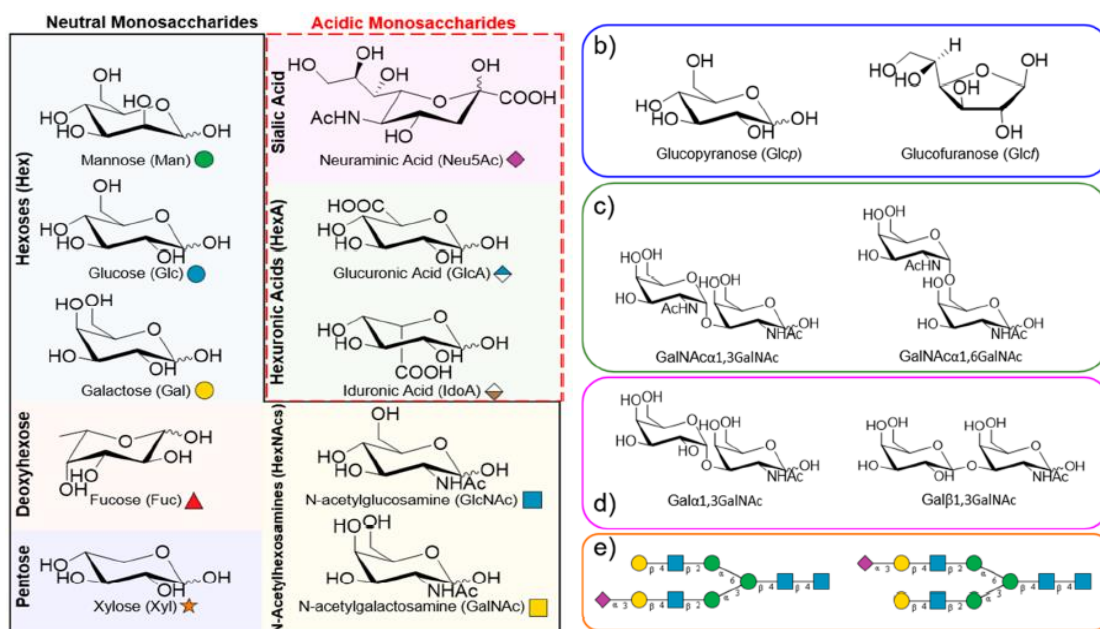
## ***Chapter III.***

### III. Structural and functional analyses of carbohydrates ligands and their interactions with lectins

In this chapter, I review the various technical approaches that have been used for structure elucidation of LPS and carbohydrate-lectins interactions. The aim of this part is not only to add information regarding the structure of LPS and the interactions between carbohydrates and CLR, but also to introduce principles of the related scientific techniques and to assess potential advantages and limitations of some of the methods. This would aid the reader in understanding the rationale behind selecting to employ specific techniques during my PhD.

#### III-1. Structural elucidation of LPS and LOS molecules

As anticipated in chapter I, LPS molecules are naturally produced with different monosaccharides building block structures presenting high diversity in terms of sugar type, nature (figure 36) and branches in the poly- and oligo-saccharide sequence, with the presence of the Lipid-A portion. The later provides a lipophilic character to LPS molecules besides to the hydrophilic property supplied by the poly-saccharide portion. Consequently, the amphiphilic character of LPS/LOS structures renders them hard to study experimentally, especially when it comes to the characterization of molecules with unknown structures. Even so, their elucidation is of utmost importance giving their association to infectious agents being an active research field.



**Figure 36. Examples of carbohydrates that could be found in human and in some LPS molecules.** The figure shows the different isomerism in sugars ring sizes (b), region-isomerism (c), anomeric configuration of the glycosidic linkage (d) and positions (e) that they may adopt in a given polysaccharide sequence. Up to sixteen different isomers could be formed by one monosaccharide.

Taken from<sup>232</sup>.

From our standpoint and based on scientific bibliography, the different biological activities observed from LPS and LOS glycoconjugates might partially result from their different lattices. In the present PhD project, the studied LPS/LOSs were produced by bacteria in native environment. Being in the closest conditions to biological material, their investigations may contribute to having more real outcomes regarding their recognition by C-type lectins, in contrast to synthetic glycoconjugates. Some well-known methodologies for LPS molecules extraction and chemical analyses are discussed in this part of the chapter.

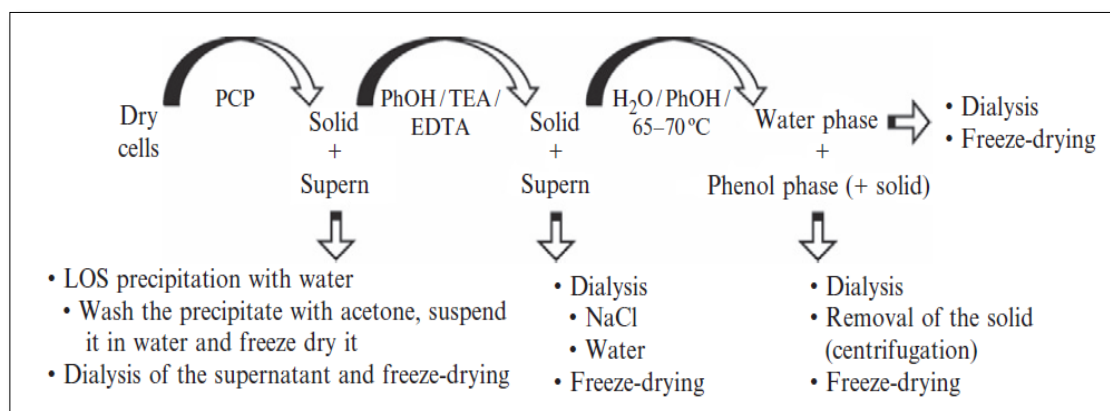
### **III-1-1. LPS Extraction methodologies**

The isolation of bacterial cell wall components is the very first step required for structural elucidation of glycoconjugates. Extraction methods have been widely described since the discovery of LPS molecules in the 1940s. They include extractions with various chemical agents which have been mainly described by Wang and colleagues<sup>233</sup>. Depending on the form in which LPS molecules are produced (LOS or LPS) their extraction is directed to a specific method. LPS molecule possesses the O-Antigen portion which significantly increases its solubility, and therefore LPS tends to be extracted in aqueous solutions. Contrarily, the LOS, which contains varied sizes of core oligosaccharide but no O-Antigen, with higher lipophilic character, hence it is usually extracted into low-polar mixtures. For example, the **Phenol-Chloroform-Petroleum ether method (PCP)**<sup>234</sup> is efficient for LOS isolation by separating the later from LPS and nucleic acid as well as proteins.

In some cases when the LOS presents changeable sizes of core oligosaccharides, which would be related to bacterium growth conditions and/or LPS modifications, it is recommended to complementarily mix the reported methods to obtain the desired LOS molecules. For example, LOS from *Bulkholderia* strains, generally acts with a particular partition behaviour during its extraction, thus yielding various amounts. In this view, single extraction methods are rarely used for the extraction of new polysaccharides, hence their combination is more suited (figure 37). For instance, when the phenol solution (obtained from PCP) is mixed with triethylamine (TEA) and EDTA, this extraction mixture operates likewise, leading to LOS isolation<sup>235</sup>.

In case of LPS extraction, another method based on the use of Hot **phenol/water extraction**<sup>236</sup> mixture which favours LPS precipitation in water phase. This mixture of phenol and water, when heated above 65°C, forms a single phase that is separated into two phases below this temperature. Total proteins are mainly contained into phenol phase, while LPS and nucleic acids are found in the water phase. This method has been used for the extraction of poorly

contaminated and highly pure LPS from *E. coli* and *S. typhi*<sup>237</sup>. Enzymatic treatment followed by dialysis and chromatographic separation are additionally established right after LPS sedimentation in water phase to remove important amounts of soluble contaminants.



**Figure 37. Combined extraction methods strategy for LOS/LPS molecules purification.** Phenol/EDTA/TEA is another extraction protocol that can be inserted among the steps. Taken from<sup>235</sup>.

The above-mentioned methods were employed during my PhD since we were interested in the isolation of both LPS and LOS. As an example of other methods described in the literature, the Tri-Reagent extraction that is dedicated for LPS and Lipid-A isolation from dried bacteria<sup>238</sup>.

### III-1-1-1. Detection of LPS molecules by electrophoresis methods

LPS molecules are investigated, after their extraction, by **gel electrophoresis migration** followed by **silver nitrate staining**<sup>239</sup>. To obtain a well resolved ladder of LPS bands ranging from smooth type LPS to compact LOS, sodium deoxycholate (DOC) is occasionally used instead of sodium dodecyl sulphate (SDS) to dissociate LPS aggregates during its migration in the polyacrylamide gel whence the name DOC-page electrophoresis<sup>240</sup>.

Due to size heterogeneity in terms of sugar number, LPS (50-100 KDa) and LOS (few KDa) can be detected on the gel. LPS having high molecular weight compared to LOS, is detected by the appearance of multiple spread bands along the corresponding gel profile. Counter to LPS, bands from LOS molecules are detected at the bottom of the gel, characteristic for such molecular weight values.

Undoubtedly, the isolated LPS/LOS preparations might be co-extracted with additional bacterial components such as free cationic proteins and membrane proteins and phospholipids. By using

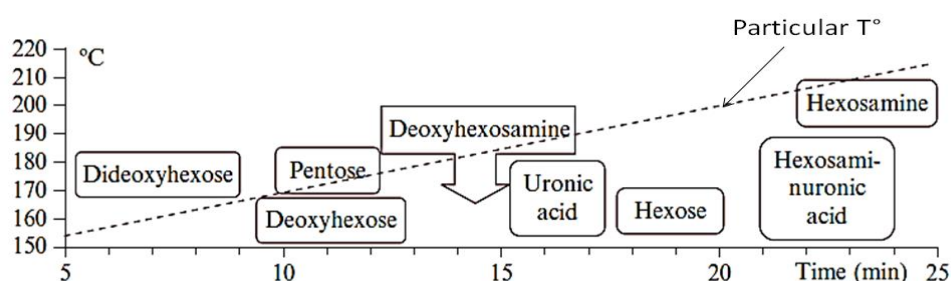
another electrophoretic method by specific staining with **Pro-Q Emerald 300 dye**, LPSs and glycan-linked proteins could be directly distinguished on the gel<sup>241</sup>. This is a very sensitive approach that allows the detection of few nanograms of LPS fluorescent conjugates, formed after dye reactivity towards oxidized carbohydrate groups (i.e., aldehydes).

### III-1-2. Chemical treatments of LPS molecules

Primary structure of the isolated LPS/LOS can be analysed after applying a sequence of chemical reactions to gain insights on sugar composition and nature. This part is meant to describe how the LPS molecule, freshly isolated from bacteria, undergoes different chemical reactions that are selected according to the ensuing investigation. In our case, we mainly used GC-MS analyses to investigate sugars and fatty acids composition of the isolated LPS/LOSs.

#### III-1-2-1. Derivatization of LPS molecules prior to GC-MS analyses

Because monosaccharides are not naturally volatile and do not resist to high temperatures imposed by some analytical approaches, they must be derivatized. In fact, the polysaccharide must be converted into **Acetylated O-Methyl Glycosides (AMG)**. Hexoses, deoxyhexoses, uronic acids, aminosugars and Kdo could be distinguished right after this step by using Gas Chromatography-Mass Spectrometry (GC-MS) (figure 38).

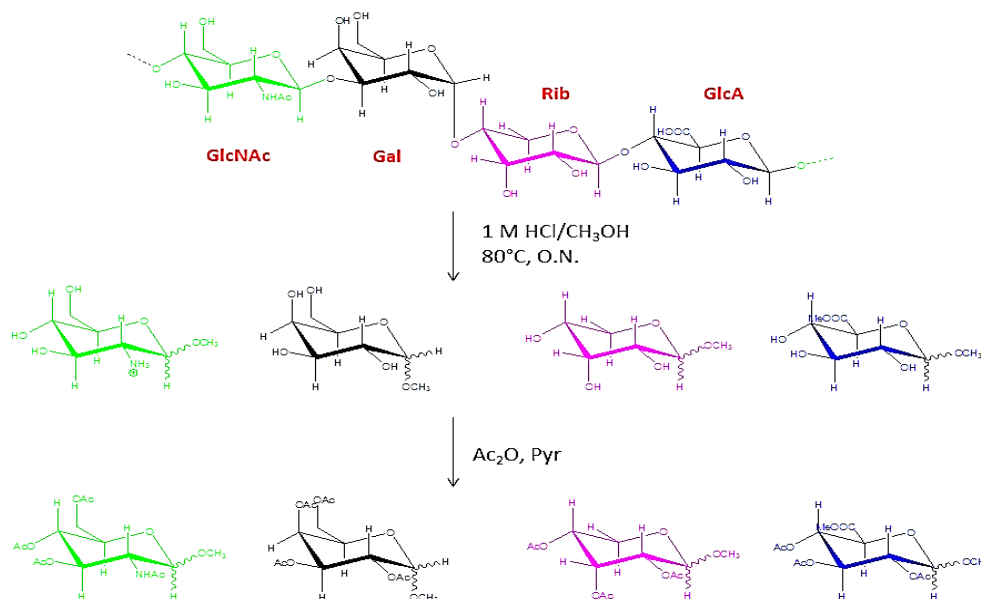


**Figure 38. Elution order corresponding to each type of derivatized sugars in the GC-MS chromatogram.** Adapted from<sup>235</sup>.

The starting molecule undergoes two main reactions (figure 39). Firstly, **methanolysis** or hydrolysis of the glycosidic linkages with anhydrous HCl/MeOH solution is carried out to obtain the corresponding O-methyl glycoside of each monosaccharide. Secondly, **acetylation** reaction is performed with anhydrous acetic anhydride (Ac<sub>2</sub>O) in pyridine (Pyr) which leads to formation of acetylated methyl glycosides<sup>242</sup>, detectable by GC-MS.

This is an efficient way to render each sugar volatile and analysable by GC-MS. In between the two steps, fatty acids are extracted into hexane solution for further spectrometric analyses. A

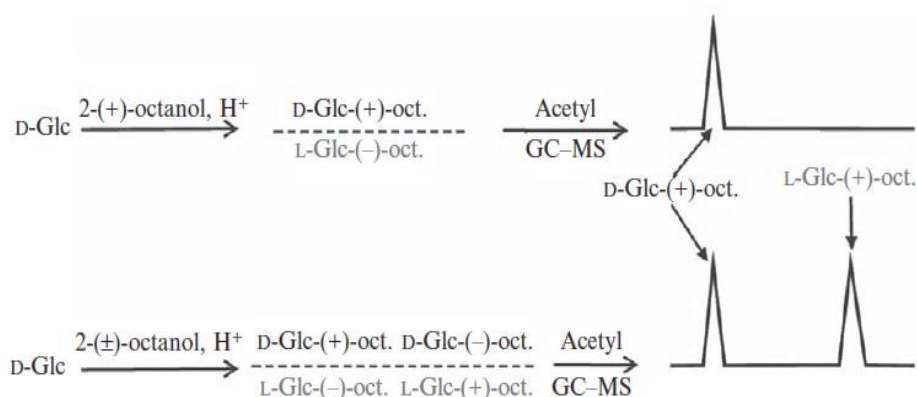
mixture of sugars with different anomeric centers ( $\alpha$  and  $\beta$ ) and ring sizes like pyranose (6-carbon sugar) and furanose (5-carbon sugar) might be generated from this method and those isomers could not be distinguished afterwards in the chromatograms because of their identical fragmentation patterns thus hampering their identification, however this fact has no significant influence on the quality of the analyses, since sugar type and composition could be determined through this method by the analyses of the fragmentation patterns ( $m/z$ ) of each sugar and by comparing them to GC-MS standards.



**Figure 39. Overview of the two main reactions for LPS transformation into acetylated O-methyl glycosides (MGA).** Taken from <http://glycopedia.eu/>

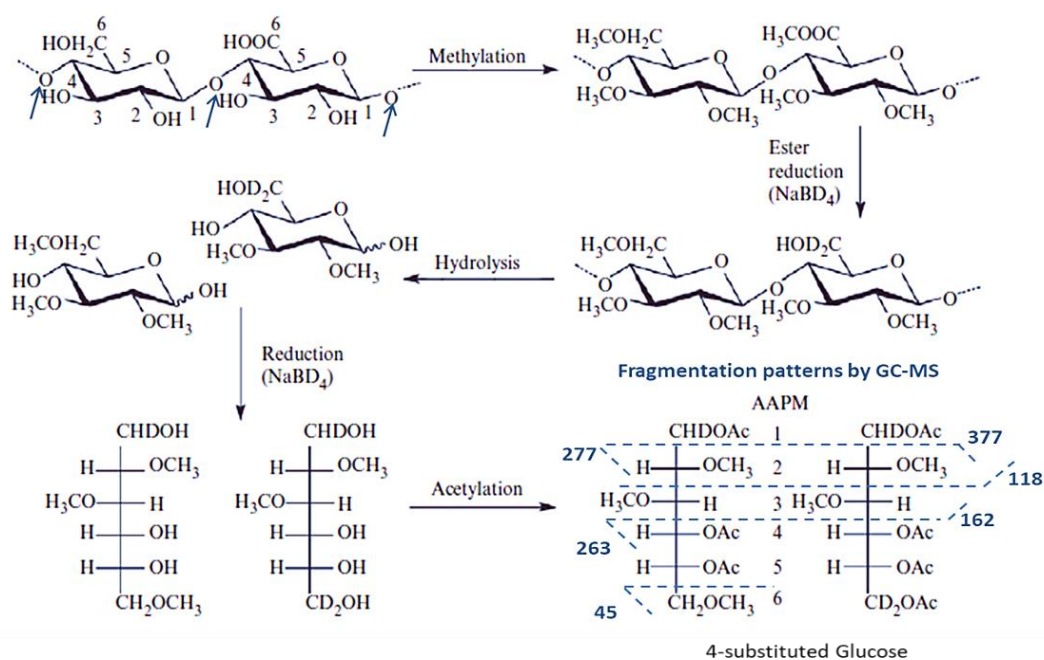
The absolute configuration of sugars present within an LPS/LOS molecule is identifiable by means of the construction of Octyl glycoside standard, a method developed by Leontein et al<sup>243</sup>. The idea is to derivatize the isolated LPS with enantiomerically pure 2-(+)-octanol (or 2-(-)-butanol) to distinguish L from D enantiomers compared with standards of known absolute configuration, according to their retention times (RT) (figure 40).





**Figure 40. Preparation of octyl glycoside standard from D-glucose.** A mixture of diastereoisomers (D) is produced in the second reaction which have the same RT as the enantiomers (L). Taken from<sup>235</sup>.

A succession of chemical reactions producing Partially Methylated Alditol Acetates (AAPM) from LPS/LOS is also used to investigate glycosidic linkages, so branches and ring sizes of sugars composing the LPS molecule<sup>244</sup>. Briefly, a full methylation of the LPS is followed by ester reduction and successive hydrolysis then acetylation to end up with and AAPM analysable by GC-MS. A summary is given in figure 41 where a di-Glucose liked monosaccharide underwent the reaction cascade that gave rise to *O*-Methyl groups (at C2, C3, and C6) derived from free hydroxyl function of the original sugar sequence, and to *O*-acetyl groups (at C4, C1 and C5) which correspond to the substitution point and to the type of sugar ring closure, respectively.



**Figure 41. Example of derivatization protocol used to deduce the monosaccharide sequence of a polysaccharide.** For clarity, the sugars that are shown correspond to a part of the polysaccharide

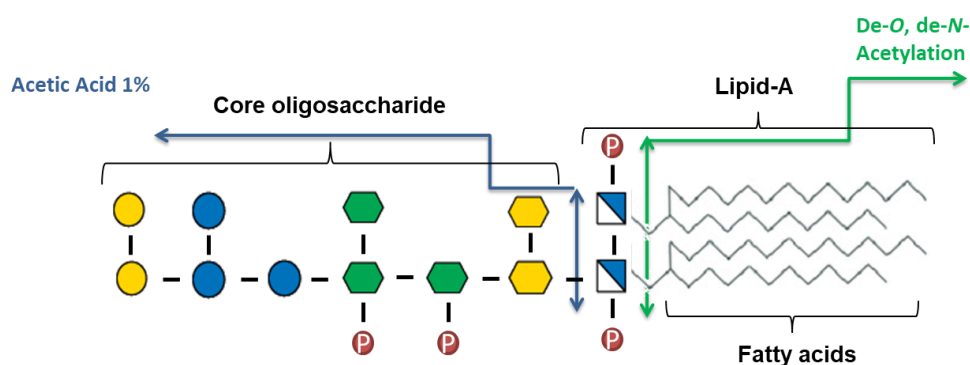
molecule linked through glycosidic linkages (highlighted in blue arrows), and the values are indicative of the fragmentation pattern of each part of the residue (m/z). Adapted from<sup>235</sup>.

### III-1-2-2. LPS delipidation for further NMR analyses

As said several times along this manuscript that the complexity of LPS glycoforms reside in their heterogeneous and variable sized structures. For example, the number of Kdo residues vary in the oligosaccharide sequence and so the presence of phosphorylated groups may vary too, and, in the same way, the Lipid-A composition is variable. Else, the presence of Lipid-A portion complicates the analyses by using “in solution” methods increasing the viscosity of the sample (by forming gel like morphology) due to the hydrophobicity of the molecule. Therefore, the analyses of LOS samples require more attention and specificity. To do that, these molecules must be delipidated, in other words, the lipid-A or the fatty acids should be removed.

LPS possesses various types of chemical linkages including a highly acid-labile ketosidic linkage between the Kdo moiety and non-reducing GlcN of the lipid A. During delipidation, this linkage is cleaved thanks to a commonly used degradation technique that is **hydrolysis** in mild acidic conditions at pH 4.4 or 1% acetic acid solution<sup>245</sup>. Chemically speaking, the starting point for chemical degradation is Kdo sugar where an intermediate carbocation (i.e. oxonium ion) is actually formed during the reaction due to the lack of any electron leaving group in its C3 position which is adjacent to the anomeric carbon C2 in Kdo, thus favouring the passage from chair into half-chair conformation of the Kdo residue. The above-described methodology, based on an acid treatment, has a disadvantage regarding further NMR data analysis (especially for a structural analysis of monosaccharide sequence). In fact, the cleavage of this acid-labile glycosidic linkage between Kdo and glucosamines, usually forms reducing terminal glucosamine residue (about 40% of the total glucosamines)<sup>246</sup> and a terminal kdo with a mixture of anomeric conformations ( $\alpha$  and  $\beta$  anomers).

There is another degradation method that is based on alkaline treatment or mild hydrazinolysis of LPS using **O-deacylation** followed by **N-deacylation** reactions<sup>242</sup>. Ester-linked fatty acids are removed by hydrazine in anhydrous conditions and amide-linked fatty acids are then discarded by the presence of 4M KOH. Therefore, both parts of the LPS or LOS molecule could be recovered separately in two phases by water/chloroform and later analysed.



**Figure 42. Schematic representation of cleavage sites established during the two chemical reactions used for solubilisation of LOSs molecules.** The acetic acid treatment occurs by breaking the glycosidic linkage between Kdo and the GlcN moiety of Lipid-A and De-O-de-N-acetylation reactions release N- and O-linked fatty acids.

The two degradation methods stated above are usually used complementary to guarantee a full delipidation of LPS structures. Importantly, the oligosaccharidic portions (**OS**) that are generated as final product from these reactions are structurally different (figure 42). In particular, the mild alkaline treatment (hydrolysis) produces an OS devoid from these two sugars because of their elimination right after the cleavage of the terminal Kdo residue. On the contrary, the OS obtained from mild hydrazinolysis contain the full sugar composition of the core oligosaccharide part, that is including the two glucosamines from the Lipid-A, since the cleavage is performed between the amino and/or ester groups of these sugars.

### III-1-3. Common methods for LPS structure elucidation

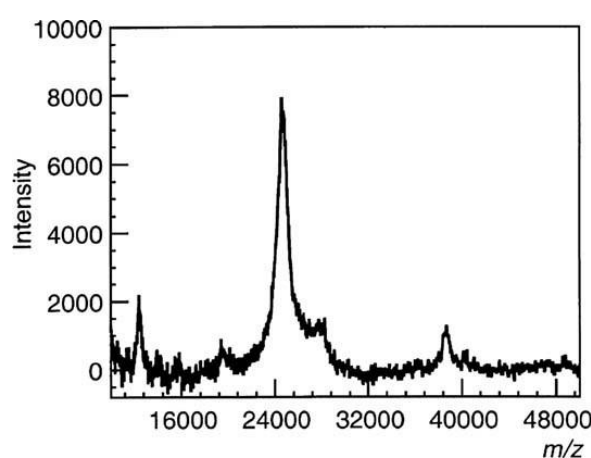
Knowing that fifty to seventy percent of mammalian proteins are glycosylated and that host-pathogen interactions are mediated by glycoconjugates, many spectrometric methods have been directed to understand the fascinating analytical challenge of carbohydrates. Notably, the vast diversity in monosaccharides building blocks constituting LPS/LOS molecules constitutes a big challenge in comprehensive spectrometric analyses. Commonly, the best way to identify glycolipids scaffolds, like LPS/LOS with beforehand known or unknown structure, is to combine different complementary methods that help each other to provide a complete and relevant information. Mass Spectrometry (**MS**) and Matrix-assisted laser desorption/ionization (MALDI) analyses smartly precede Nuclear Magnetic Resonance (**NMR**) spectroscopy to supply with reliable structural information thus facilitating subsequent spectroscopic studies.

### III-1-3-1. Comparison of ionisation techniques for LPS compositional analyses

Mass spectrometry (**MS**) is frequently used for structural investigations of glycoconjugates and especially for obtaining a detailed picture of sugars connectivity<sup>247–250</sup>. It should be noted that tandem MS techniques are preferably chosen for Lipid-A investigations including the acylation profile, the heterogeneity of secondary fatty acids chain lengths and the identification of specific fatty acids positions<sup>251,252</sup>. An excellent review by Kilar et al, describes a combination of analytical methods dedicated to LPS and LOS analyses<sup>253</sup>.

Oligosaccharides sequences are traditionally analysed by MS techniques like **GC-MS** which provides insights into monosaccharides nature and composition together with their absolute configuration, if needed, and for Lipid-A compositional analyses besides. Briefly, carbohydrates (from AMGs) (c.f. fatty acids) experience a passage through the GC under high electric field, so that when they drift thanks to their collision with the gas, fragmented ions are separated and detected after a given time depending on the size of the molecules<sup>254</sup>. Thus, species with identical mass and different sizes could be easily distinguished using this method.

For LPS and LOS analyses, MALDI spectrometry could be selected and experiments are carried out in negative ion mode because of the hydroxyl groups of sugars that experience deprotonation during the experiment and become negatively charged and detectable. The interpretation of LPS/LOS spectra is relatively easy, because MALDI usually generates single charged ions, and especially when the choice of the matrix is well-adapted to the sample. As an example, intact LPS from *E. coli* strain O164 generated a single peak at 24.7 KDa (figure 43), the number of O-Antigen repeating units was also estimated to be 24.



**Figure 43.** An example of negative-ion MALDI-TOF mass spectrum of LPS from *E. coli* O164.

Taken from<sup>255</sup>.

In contrast to GC-MS, soft ionisation techniques like MALDI and ElectroSpray Ionisation (ESI) do not require LPS/LOS derivatization prior to analyses because the energy imparted by those techniques is quite enough to degrade labile bonds like phosphate and acyl groups. Consequently, the interpretation of the outcomes derived from this technique for assessing the biological function of LPS with missing residues becomes complicated. In our group, MALDI technique has been optimized by Sturiale et al<sup>256</sup> to make it more adapted for LOS analyses including molecular weight calculation and identification of the different structural parts without the need of partial sample delipidation. Elegantly coupled to NMR spectroscopy, MALDI and GC-MS have been widely used in our group to characterize isolated LPSs and especially LOSs from unknown or new bacterial strains<sup>257–259</sup>.

The variety of analytical methods within a wide range of spectrometer types absolutely helps in constructing a structural definition of a biological extract like LPS and LOS molecules. Furthermore, this diversity enables scientific researchers to choose a combination of methods that must be adequate to the molecule of interest to guarantee relevant measurements.

### **III-1-3-2. NMR spectroscopy for carbohydrates analyses**

#### **III-1-3-2-1. Principle of NMR spectroscopy**

The utmost way to understand a sample's composition and/or the molecular structures and dynamics of the matter that is contained in it, is to observe a 'response' from those molecules under a certain "treatment". In spectroscopy, the "treatment" could be different type of radiations, depending on the spectroscopic method and on the sample to be studied, that the matter experience by being excited, thus giving off a "response" that is translated into a spectrum, hence the name spectroscopy. Nuclear Magnetic Resonance (**NMR**) spectroscopy is a spectroscopic technique that uses an electromagnetic radiation as a strong and constant magnetic field called ( $B_0$ , to imagine as a big magnet) followed by an oscillating magnetic field dubbed ( $B_1$ ), to interact with the matter which will give rise a resonance, measurable and detectable by NMR spectroscopy.

To do that, at least one type of nuclei or isotopes that compose the studied matter/molecule must have a non-zero nuclear spin ( $i$ ) that describes the intrinsic electromagnetic features of a rotating nucleus with spherical charges distribution, such as  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  or  $^{31}\text{P}$  which have  $i$  equal to  $\frac{1}{2}$ . Nuclei with other spin number values are not measurable (for a minority of nuclei with even number of protons and neutrons, and so  $i=0$ ) or hard to understand by NMR spectroscopy.

Anyhow, the latter case concerns nuclei like deuterium  $^2\text{H}$  with even mass nuclei but odd numbers of protons and neutrons, their  $i$  equals 1, detectable by NMR however their “unshaped” charge distributions make them hard to investigate<sup>260</sup>.

When launching the NMR experiment, the NMR detectable nuclei composing the sample are enclosed by the external magnetic field of the spectrometer ( $B_0$ ), in the NMR tube inserted in it, and they start to behave as little magnets by getting aligned with or against  $B_0$ , while precessing gravitationally (move like spinning gyroscope experiencing gravity) in a direction parallel to the  $B_0$  field. The frequency by which those nuclei precess is called the Larmor frequency ( $\omega_0 = \gamma B_0$ ) and it is proportional to that  $B_0$  field and to the gyromagnetic ratios ( $\gamma$ ) of nuclei, which actually reflects the sensitivity of nuclei in NMR. For example,  $^1\text{H}$  is more sensitive in NMR than  $^{13}\text{C}$  since proton ( $^1\text{H}$ ) has higher  $\gamma$  (26.753). The most isotopes nuclei used for NMR spectroscopy of **sugars** are  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{31}\text{P}$ .

As said, the  $B_0$  field is considered as a big magnet that influence the orientations of those little magnets ( $^1\text{H}$  nuclei for example) which will get the  $B_0$  parallel orientation in two different spin directions that are the  $\alpha$  and  $\beta$  spin states but since the energy level of the  $\beta$  spin state is higher, so the nuclei of the sample will adopt  $\alpha$  spin state orientation. Indeed, nuclei which are shielded from the applied  $B_0$  field assuming that they have a rich electronic environment for instance, would not adopt the same orientation as the others. Commonly, the difference between energy levels that the nuclei may adopt is called the “nuclear Zeeman splitting”<sup>261</sup>. Thence, a kind of force, that is the  $B_1$  field, is needed to make all these equilibrated energy populations enter in resonance, to excite the system and so to generate an NMR signal. Depending on the gyromagnetic ratio of the target/observed nuclei, the energy of  $B_1$  would be different and matches its Larmor frequency. This energy is absorbed by the nuclei making them flip between the energy states and, as soon as they return back to equilibrium, they give the “response” as an electromagnetic oscillating signal that contains all the spectroscopic information of the molecule.

The  $B_1$  field (i.e. a radiofrequency pulse) is introduced to the  $B_0$  field and its energy could be high or low depending on the degree of shielding of nuclei, this would give downfield (from deshielded nuclei) and upfield (from shielded nuclei) signals or peaks in the **1D  $^1\text{H}$ -NMR spectrum** for instance. And here is the particularity of NMR spectroscopy in determining the exact positions and distances between atoms within a molecule solely from 1D spectrum, compared to other spectroscopic methods. To obtain an NMR spectrum, the electromagnetic oscillating signal,

emitted by the excited system, generates an electrical signal detected by coils surrounding the NMR magnet and then translated to a Free Induction Decay (FID). The latter decays exponentially until it gets cancelled when the nuclear spins go back to the initial equilibrium. This FID is then mathematically converted to a frequency in Hertz (Hz) data (which basically reflects the amount of energy necessary to put the corresponding nuclei in resonance), observable in the NMR spectra, by Fourier Transformation (FT).

### III-1-3-2-2. Example of data provided by NMR spectroscopy

In NMR spectroscopy, the **chemical shift ( $\delta$ )** is one of the utmost considered data because it contains the main information about the molecule of interest, and it is obtained from the processed spectrum. In fact, the frequencies generated from the FID are then transformed into part per million (ppm) which is the unit of their corresponding chemical shifts. Not only the nature or the positions of molecules could be deduced from a simple 1D NMR spectrum, but also their molecular size, their quantities and in some cases their interactions with other molecules. For example, the binding activity of a ligand to its receptor is, in theory, measurable in all NMR methods, and it includes  $\Delta \delta$  perturbations or changes in diffusion constants or changes of Nuclear Overhauser Effects (**NOEs**) or saturation transfer in the ligand's or receptor's NMR spectra<sup>262</sup>. NOE is an NMR observable that is detected as changes in NMR resonances intensities (positive or negative) and it is generated from the transfer of spin polarisation between two nuclei or spins, from which one is irradiated by an RF pulse (thus bringing it to resonance as previously explained) at its Larmor frequency and, the other spin has a dipolar coupling with the former spin (i.e., the interesting spin). This transfer happens only when the two nuclei are close in space ( $\sim 1-5 \text{ \AA}$ ) thus providing knowledge about structures and molecular interactions<sup>263</sup>.

**J-coupling** (in Hz) is another NMR data that reflects the magnetic interaction between a pair of protons (c.f. scalar coupling) which influence reciprocally each other (as little magnets) providing a splitted single NMR signal (or multiplet). J-coupling is the difference between sub-peaks that form the splitted NMR signal. J-coupling measurements from 1D-spectra ( $^1\text{H}$ ) for structure investigations (e.g., of non-anomeric protons) is limited in case of line broadening. In this case, line width may be informative about, for instance, the local environment of the corresponding residue.

Interestingly, an expansion to the **multi-dimension** (2D or 3D spectrum) is usually required for NMR analyses, and this widening enables us to observe what is going on in terms of intra- and inter-connections between the different types of nuclei constituting the molecule in a spacial

fashion. These informations are found into NMR **cross-peaks** resulting from correlations between frequencies from  $^1\text{H}$ , in one dimension, and another frequencies from  $^{13}\text{C}$  or  $^{15}\text{N}$  in a second dimension, for example<sup>264</sup>.

The outcomes that we could get from such NMR data are not restricted to **structural investigations** but also to the **dynamics** of the biomolecules. Relevant data could be obtained from NMR spectroscopy with high sensitivity when these biomolecules (e.g., lectin or LPS) are isotopically labelled with nuclei like  $^{13}\text{C}$  and  $^{15}\text{N}$ , by expressing the genes responsible for their biosynthesis into engineered bacteria (like competent *E. coli*) grown in medium supplemented with those isotopes as nutrients for these cells, thus the NMR detection will be directed to those nuclei. This is a way to provide the biomolecules with NMR detectable nuclei that have very low occurrences in nature otherwise (like 1.1% for  $^{13}\text{C}$ ).

### **III-1-3-3. NMR spectroscopy is a good tool to study large molecular assemblies of LPSs**

It is possible with NMR spectroscopy to investigate the isolated biomolecules in a canny way, by choosing the adequate type of experiments to run insure to guarantee the most adapted conditions. This means that depending on the nature and the morphological behaviour of the sample to be studied, **liquid- or solid-state (ss-NMR) spectroscopy** could be selected.

In our case, both ss-NMR and liquid-state NMR methods were used giving that we were interested in different type of biomolecules (membrane proteins and glycoconjugates). The molecular motions of these biomolecules depend on their sizes and their degree of heterogeneity. In case of relatively small biomolecules with high heterogeneity like LPS, ss-NMR spectroscopy would be the best choice because LPS molecules form big micelles in solution with gel like structures (due to the presence of Lipid-A), so LPS molecules will tumble very slowly in solution (liquid-state NMR) and this will generate bad quality NMR spectra with broad peaks. In ss-NMR, this drawback is resolved because anisotropic interactions including Chemical Shift Anisotropy (CSA) and dipolar coupling, are not averaged to zero and thus those interactions could be measured<sup>265</sup> (not like in solution state NMR where the molecules tumble very fast so that those anisotropic interactions could not be “heard” because they are cancelled out).

However, in these large biological systems including native LPS micelles with little mobility, the spins rotation is affected and so an artificial “force” must be introduced to make these spins rotating (tumbling) with respect to the  $B_0$  direction as they do in solution, thus mimicking their



molecular or Brownian motions. Conventionally, **Magic Angle Spinning (MAS)** hardware is used to provide this artificial rotation which extremely helps in orientation averaging, thus generating narrower NMR signals from bilayer structures<sup>266</sup>.

Not long ago, ss-NMR spectroscopy is used as a fascinating approach for establishing scientific studies that consider the biological interplays within the cell membrane like pore forming by AMPs during membrane permeabilisation<sup>267</sup>, LPSs interactions with antibiotics<sup>268</sup> and even atomic level investigations of intact bacterial peptidoglycans-enzymes complexes<sup>269</sup> and fungal cell walls organization<sup>270</sup>. Additionally, intrinsically disordered and unfolded proteins have been analysed by ss-NMR<sup>271</sup>, although these biomolecules could be investigated by solution state NMR spectroscopy under certain labelling conditions<sup>272</sup>. The only problem that researchers may encounter, and this might be the reason why ss-NMR has not been discovered in parallel with solution NMR, is that the use of this method requires large amount of sample. Fortunately, this does not present a real drawback in our system since LPS molecules could be purified in large quantities when using the appropriate extraction method.

#### **III-1-3-3-1. Polysaccharides LPS analyses by liquid-state NMR**

NMR spectroscopy has become a powerful tool to foster carbohydrates analyses thanks to the development of many NMR techniques progressively adapted to proteins and nucleic acids investigations<sup>273</sup>. Generally, the sole use of NMR spectroscopy for structural characterisation of LOS/LPS structure is a tricky and time-consuming, for this reason spectrometric studies, described in III-1-3-1, are formerly employed to quickly identify fatty acids species and sugar residues present in the LPS molecules.

There are several methods that have been exploited for structural and functional investigations of glycoconjugates including various approaches using liquid state conditions<sup>274</sup>. In liquid-state NMR spectroscopy, the whole LPS molecule is usually adapted for this approach because of its high solubility due to the presence of vast number of sugars, yet the O-antigen part is considered as a loosely structured part that governs and hides somehow NMR signals from the core LPS region, as a result, it gives rise to many ambiguities in the elucidation of the core LPS structure from the NMR spectra. On the other hand, removal of fatty acids from LOS, by employing the strategies described in II-1-2-2, facilitates NMR analyses on the core OS. Besides, detergents and organic solvent could be used as an alternative treatment for LPS/LOS solubilisation.

The investigation of O-Antigen or core OS portions by NMR spectroscopy yields important insights regarding host-pathogen interactions because these LPS part is believed to be crucial for LPS recognition and antigenicity. For rough LPS versions, dealing with relatively short sequences becomes affordable in solution, provided they are solubilized.

### III-1-3-3-1-1. Classical strategy for LPS sugars assignments by NMR spectroscopy

In the analysis of the “primary” structure of a carbohydrate sequence, one may identify three distinguishable regions in mono-dimensional 1D and bi-dimensional 2D NMR spectrum, that are summarized in the following table. The chemical conformation (sugar structure viewed 3D coordination), the flexibility in space and the dynamics that are presented by sugars flexibility in relation to time (e.g. while binding to a certain receptor), are all descriptors of 3D structure of the residue. It should be noted that the large chemical shift ranges announced in the table above are the result of the various information that a single sugar may provide by NMR spectroscopy.

**Table 2. Chemical shift regions for  $^{13}\text{C}$  and  $^1\text{H}$  NMR spectroscopy of carbohydrates.** Inspired from<sup>273</sup>.

Assignments	Chemical shift region for $^{13}\text{C}$	Chemical shift region for $^1\text{H}$
Anomeric signals	110-90 ppm	5.5-4.5 ppm
Sugar ring resonances	70-60 ppm	4.2-3.2 ppm
Nitrogen linked carbon signals	60-50 ppm	3.4-2.8 ppm
Aliphatic methylene signals	25-35 ppm	2.2-1.7 ppm
Deoxy-sugars (Methylene carbon and deoxy and acetyl protons)	20-15 ppm	2.5-1 ppm

For spin correlation methods, spectroscopists consider that monosaccharides are constituted of aliphatic chains of coupled spins presenting one of the good NMR properties of carbohydrates. Once soluble LPS derivatives are obtained through the chemical degradation methods introduced previously in this chapter, NMR spectroscopy can be performed provided that these sample are isotopically labelled, or they are available in sufficient quantity that enables the detection of nuclei with natural abundance.

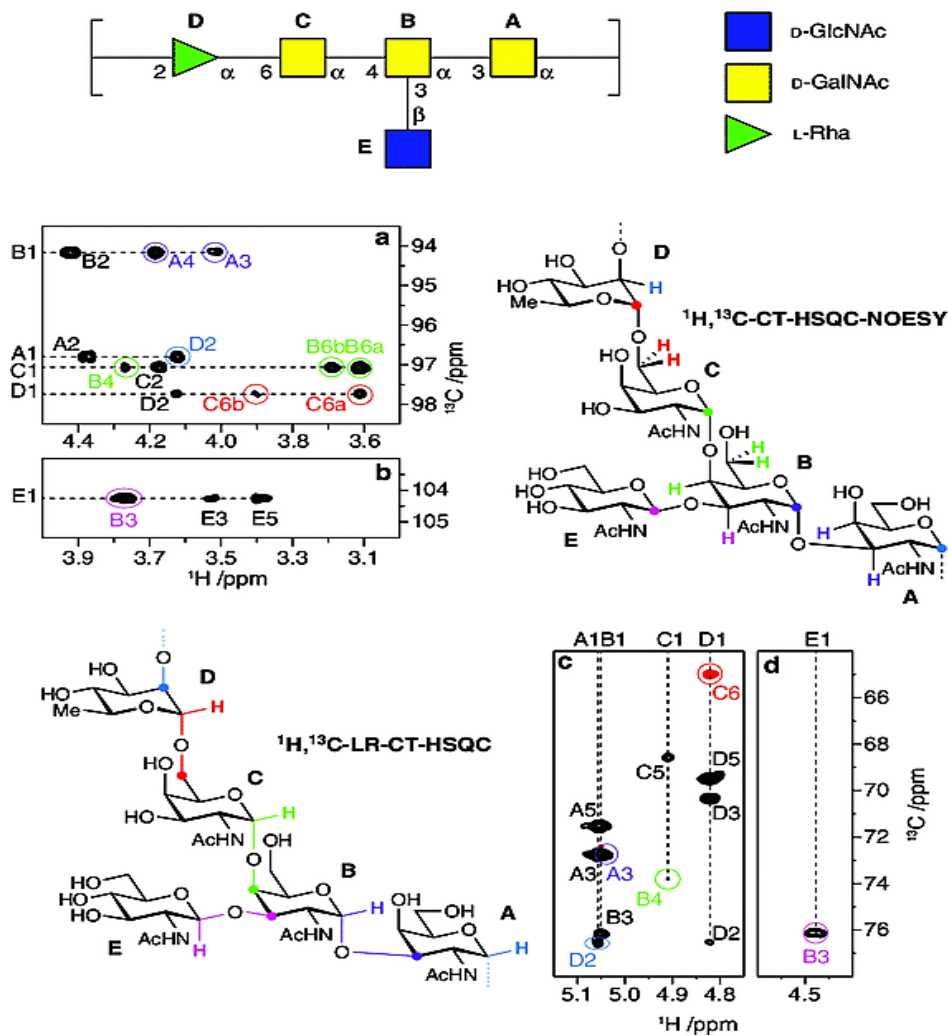
### III-1-3-3-1-1-1. 1D- NMR analyses of monosaccharides

The simplest strategy that could be used to assign monosaccharides is to start with the analyses 1D proton ( $^1\text{H}$ ) spectrum and then to move to bi-dimensional spectra. Those data allow us to map the different type of residues according to their **chemical shift** and J-coupling values as following:  $\alpha$ -configured pyranose sugar residues resonate up-field at 5.6-4.9 ppm, whereas those from  $\beta$ -glycosidic linkage at 4.7-4.3 ppm, for chemical shift values.

In addition to  $^1\text{H}$ -1D NMR spectrum,  $^{13}\text{C}$ -1D NMR signals are used to confirm the number of anomeric sugars. An additional way to facilitate the assignments at the beginning of analyses, is to start with the attribution of Kdo or NeuAc characteristic signals as the  $\text{H}_{3\text{eq}}$  and  $\text{H}_{3\text{ax}}$  from the moment that these sugars have no anomeric carbons.

### III-1-3-3-1-1-2. 2D- NMR analyses for monosaccharides structures assignments

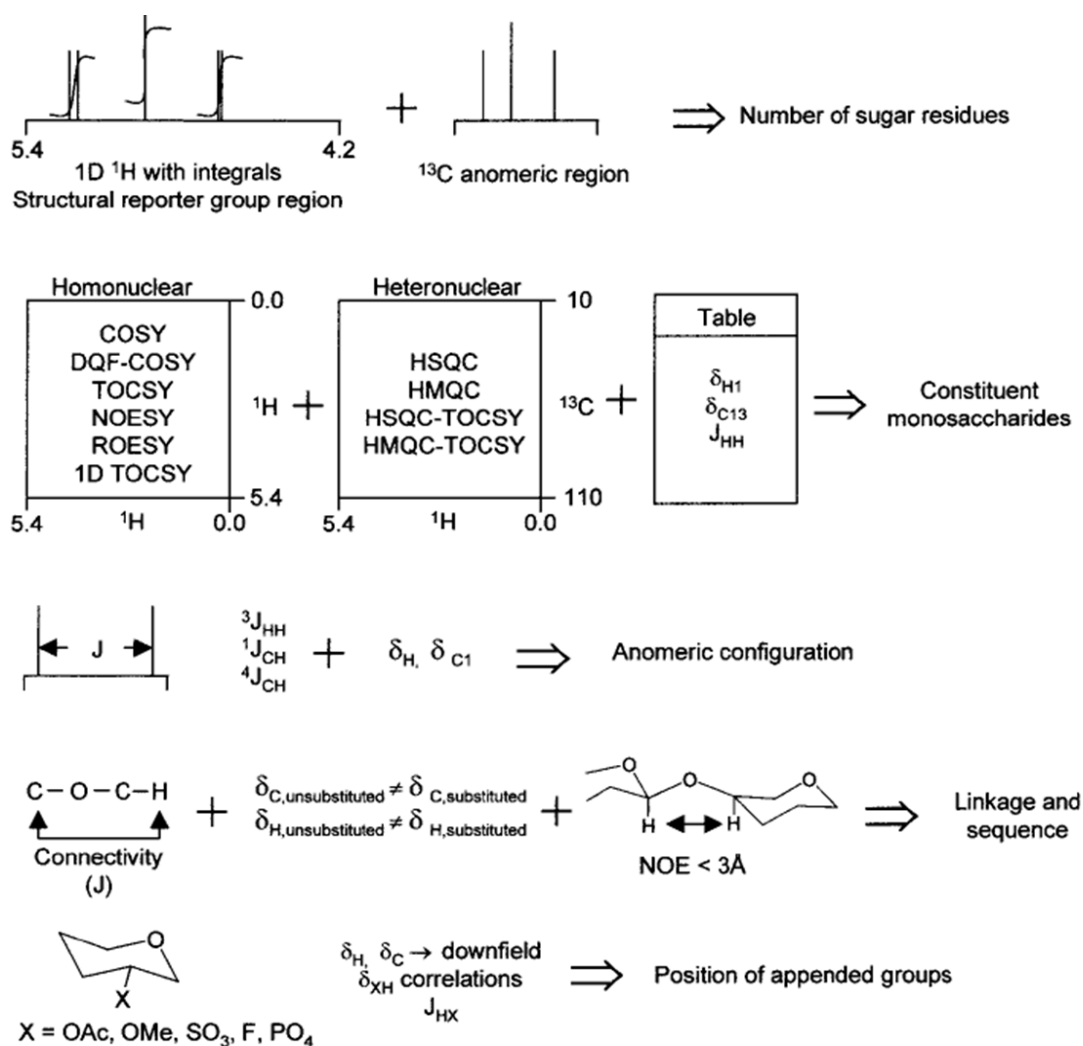
Carbohydrates analyses from NMR data could be obtained from an additional dispersion to a second dimension, by investigating bi-dimensional spectra to gain insights about connections between atoms (i.e. correlations between spins). Notably, Total Correlation Spectroscopy (**TOCSY**) and Correlation Spectroscopy (**COSY**) are analysed for proton-proton [ $^1\text{H}$ - $^1\text{H}$ ] correlations assignment, thus allowing partial estimations of sugar type and number. While the former provides NMR data from all coupled spins in each sugar residue with some exceptions including galactose which has an axial hydroxyl group at C4 that breaks up magnetization transfer; the later experiment allows the detection of individually coupled spins. Besides, another NMR experiment is usually recorded for the quick identification of a given polysaccharide the so called Heteronuclear Single Quantum Coherence (**HSQC**) which enables  $^{13}\text{C}$ - $^1\text{H}$  individual spin systems assignment (figure 44).



**Figure 44. Spin systems (A-E) of intra and inter-residues assignment of anomeric signals of a  $^{13}\text{C}$ -enriched O-Antigen polysaccharides of *E. coli* O142 by HSQC and NOESY NMR spectra.**

Taken from<sup>275</sup>.

$^{13}\text{C}$  resonances exhibit high **chemical shifts** for  $\beta$  anomers (105-103 ppm) and low chemical shift in case of  $\alpha$ -anomers (101-95 ppm). The glycosylation shift indicates glycosidic linkage positions and may be observed in such spectra by larger chemical shifts at the glycosylation site. Additionally, the exact position of appended groups (e.g., sulphate, phosphate groups) is described by a downfield shift (~0.2-0.5 ppm) for  $^1\text{H}$  and higher chemical shift for  $^{13}\text{C}$ .



**Figure 45. General descriptions about the NMR parameters and inputs related to each of the classical experiments recorded for carbohydrates analyses.** Taken from<sup>273</sup>.

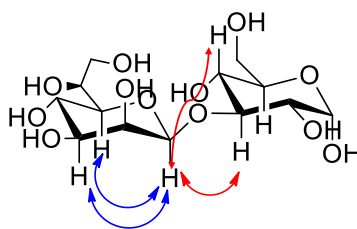
**J-coupling** constants, in addition to chemical shifts, are used to gain information about protons orientations in each monosaccharide (figure 45). J-coupling between vicinal protons is correlated to the dihedral angle perhaps high  $3J_{\text{H,H}}$  (8-10 Hz) diagnostic for trans-diaxial orientation of the protons whereas low  $3J_{\text{H,H}}$  (<4 Hz) is characteristic for equatorial/axial orientation of hydroxyl protons. Heteronuclear coupling constants between anomeric proton and carbon  $1J_{\text{C}_1,\text{H}_1}$  are also elucidated and fundamentally define the anomeric configuration. In 6-carbon membered rings,  $1J_{\text{C}_1,\text{H}_1}$  below 172 Hz is specific for  $\beta$ -anomers and  $1J_{\text{C}_1,\text{H}_1}$  above 172 Hz corresponds to  $\alpha$ -anomers.

### III-1-3-3-1-1-3. Additional 2D-NMR experiments

Nuclear Overhauser Enhancement Spectroscopy (**NOESY**) and Rotating Frame Overhauser Enhancement Spectroscopy (**ROESY**) are carried out to correlate protons that are close in space (Figure 46). Traditionally, they are coupled to homonuclear  $^1\text{H}$ -NMR experiments such as a 2D-

TOCSY-ROESY NMR spectrum to facilitate their analyses by double checking and linking previous assignments.

These heteronuclear NMR experiments are informative about the slow chemical exchange processes between nuclei owing to NOEs observations. NOEs calculations are informative about  $H_1H_5$  and  $H_1-H_3$  correlations and thus providing knowledge to the configuration of the sugar<sup>273</sup>. NOEs could be used to observe and design monosaccharide and peptide 3D structures during NMR spectroscopy investigations<sup>263</sup>. These data are regularly combined to scalar coupling constants (c.f. J-coupling) as their numbering and distribution insufficiently estimate the conformations of sugars in solution which result from their high flexibility. The scalar coupling in this case facilitates the identification of sugar sequence by averaging values over an ensemble of conformations of a given sugar residue.



**Figure 46. Example of detectable NOEs connectivities in a disaccharide by NMR spectroscopy.**  
Inter-NOEs contacts are colored in red and intra-NOEs in blue.

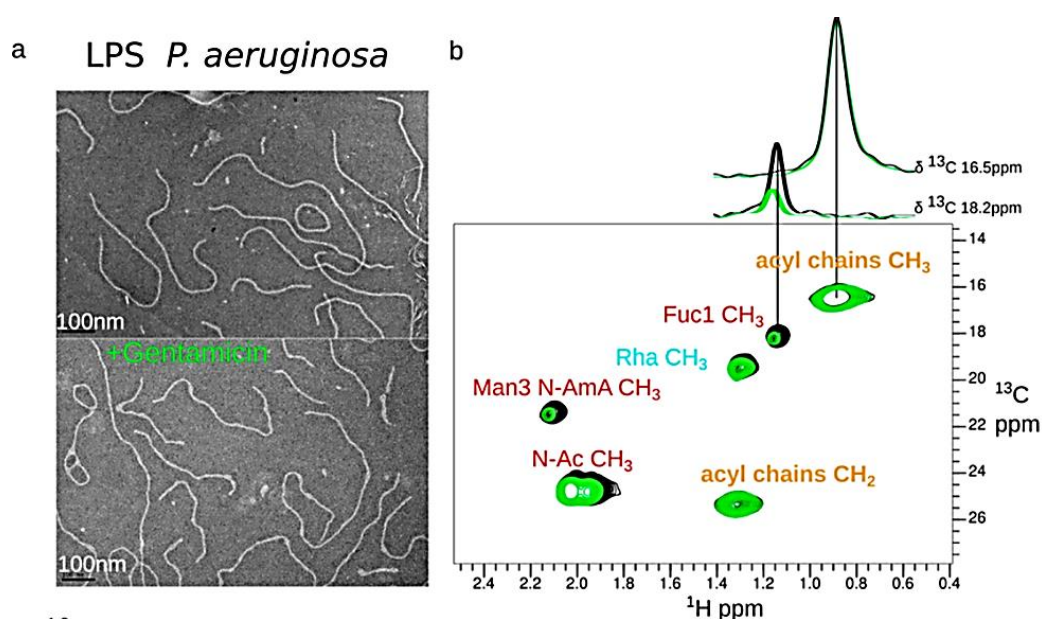
### III-1-3-3-1-1-4. 2D-ss-NMR spectroscopy experiments

As said previously, in an *in vitro* environment, large assemblies cause LPS to tumble very slowly under the magnetic field generating a long correlation time which gives rise to broad peaks in the NMR spectrum, thus hampering the measurements of spectroscopic parameters like NOEs and scalar coupling of vicinal protons. The susceptibility of signals broadening is eliminated by sample spinning by introducing MAS rotation. Thus, flexible (O-antigen) and rigid parts of the intact LPS molecule could be identified by using this method according on the experiments recorded<sup>268,276,277</sup>.

As in solution NMR experiments, there are similar sets of homonuclear experiments for [ $^{13}C$ - $^{13}C$ ] correlations within certain distances, by exploiting dipolar coupling (spin-spin coupling) such as Dipolar Assisted Rotational Resonance (**DARR**) or Proton Driven Spin Diffusion (**PDSD**) experiments. Heteronuclear experiments like [ $^1H$ - $^{13}C$ ] **INEPT** for “Insensitive Nuclei Enhanced by Polarization” experiment that is identical to HSQC in solution in terms of outcomes. By exploiting scalar coupling (J-coupling), INEPT based experiment provides high resolution signals from

flexible portions of the molecule whereas DARR and PDSO type experiments are more sensitive for rigid parts of LPS.

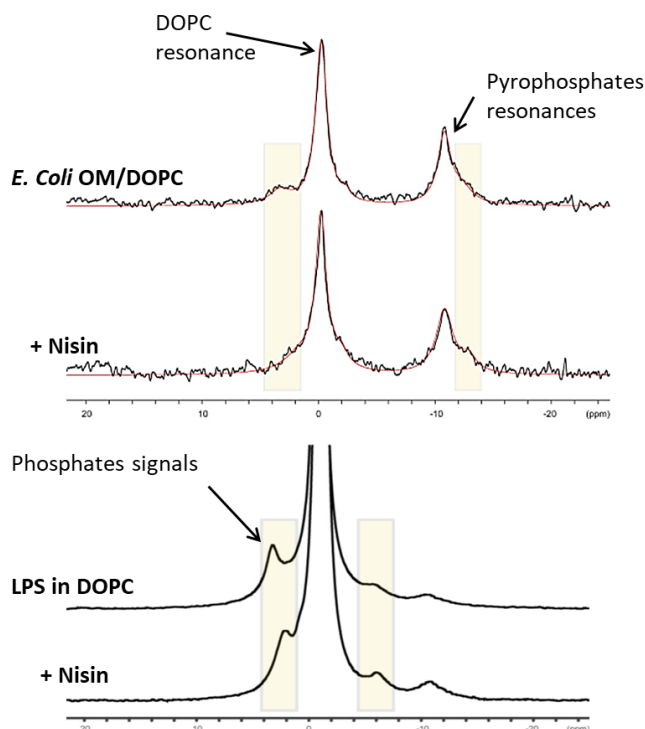
In addition to structure elucidation, those experiments could be recorded, for studying molecular interactions (figure 47). In our group, the interaction between Gentamicin (antibiotic produced by gram positive bacteria) and two LPSs with different O-Antigen portions from two strains of *P. aeruginosa* (A-and B-bands O-Antigen producers) has been investigated by means of ss-NMR spectroscopy combined with high resolution imaging (i.e. EM)<sup>268</sup>. Taken together, the results have suggested a specific binding involving mostly the trisaccharidic O-Antigen portion (Fucose residues mostly).



**Figure 47. The interaction between LPS O-Antigen by EM and ss-NMR INEPT experiments.** (a) EM images of LPS before and after addition of Gentamicin. (b) <sup>1</sup>H-<sup>13</sup>C NMR signals (i.e. O-Antigen B-Band, the most perturbed) are shown. Taken from<sup>268</sup>.

Solid state <sup>31</sup>P NMR experiments are also useful to observe signals of LPS molecules which usually have several phosphorylation sites either linked to the Lipid-A through the sugars or attached to Kdo or other residues. These experiments were used in a study realized by Lanne et al, about understanding the interaction of LOS from *E. coli* K12 strain and LPS forms from pathogenic bacterial strains with antimicrobial peptide. The results were compared based on the <sup>31</sup>P recorded spectra suggesting a potential role from Lipid-A together with contributions from the O-antigen portion during the interaction with the studied peptide (Nisin)<sup>278</sup>. Phosphorylation sites that experienced changes in intensity or little chemical shift perturbations following Nisin

binding (figure 48), were assumed to be indirectly implicated and dynamic in the interaction being proximal to the LPS binding sites.



**Figure 48. Solid state  $^{31}\text{P}$  MAS NMR spectra from (top) the Outer Membrane of commensal *E. coli* and (bottom) LPS/DOPC mixtures, in presence and absence of nisin. Adapted from<sup>278</sup>.**

### III-2. Methodologies for studying carbohydrate-lectin interactions

In this part of the chapter, I focus on introducing various techniques that have been employed mainly for **carbohydrate-lectin interactions** and, some examples of LPS interactions studies are occasionally cited as well. Specific interactions between lectins and carbohydrate ligands are part of initiation of an infection, immune response, cell adhesion, and many other biological functions as presented in the previous chapters. Thus, exploiting **binding specificity** and **selectivity** together with **molecular dynamics** during association and dissociation steps, would address pertinent challenges by obtaining insights about the interaction.

Here, some analytical techniques that have been used in a combined manner, by many researchers working on the field of molecular recognition between lectins and carbohydrates, are illustrated. Most of them considering various thermodynamics relationships (table 3), particularly focusing on binding kinetics and binding energies. For clarity's sake, technical principles details are not introduced for all the techniques cited herein.



**Table 3. Examples of fundamental equations associated to protein-ligand binding kinetics.**  
Adapted form<sup>279</sup>.

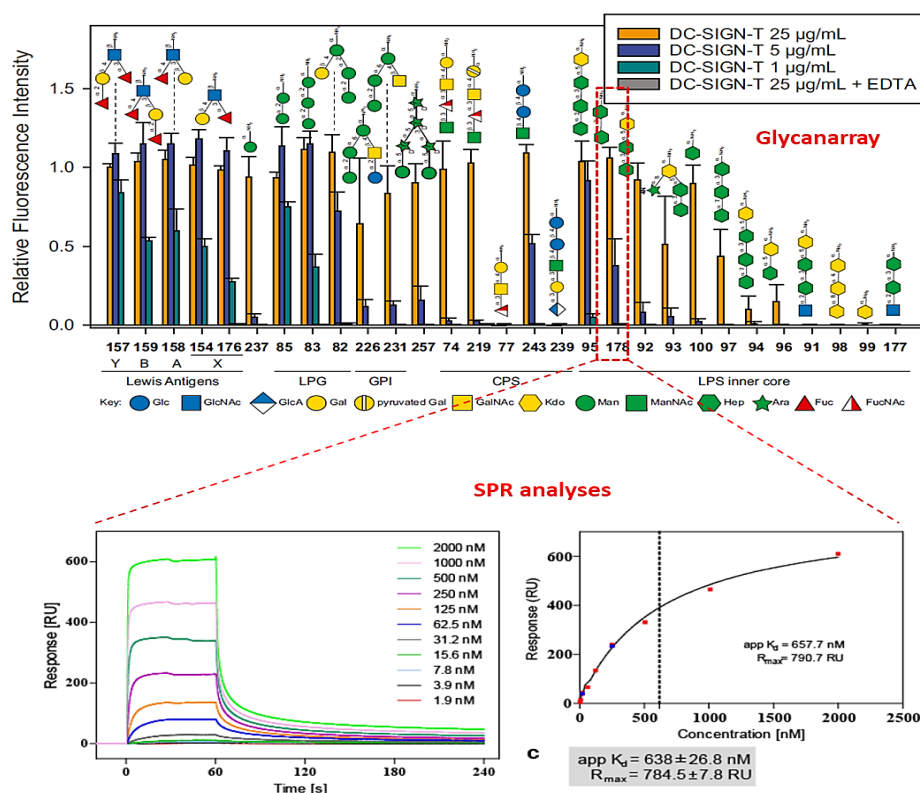
Molecular concept	Formula	Related information
<b>Binding at equilibrium</b>	$k_{on} [P] [L] = k_{off} [PL]$	P= protein, L=ligand  $k_{on}$ and $k_{off}$ = kinetic rate constants for forward ( $M^{-1}\cdot s^{-1}$ ) and reverse ( $s^{-1}$ ) binding, respectively
<b>Binding constant (<math>M^{-1}</math>)</b>	$K_b = K_{on}/K_{off} = [PL]/[P][L] = 1/K_d$	$K_b$ = binding constant ( $M^{-1}$ )  $K_d$ = dissociation constant (M)
<b>Energy from Gibbs relationship</b>	$\Delta G^\circ = -R T \ln K_b$	$\Delta G^\circ$ = standard binding free energy (at 1atm and 298k)  R = universal gas constant ( $1.987 \text{ cal}\cdot\text{K}^{-1}\cdot\text{mol}^{-1}$ )  T= temperature (K)
<b>Total binding entropy</b>	$\Delta S = \Delta S_{solv} + \Delta S_{conf} + \Delta S_{r/t}$	$\Delta S$ = entropy changes from solvent ( $\Delta S_{solv}$ ), from conformational freedom of P or L ( $\Delta S_{conf}$ ) and from loss of translational and rotational degrees of freedom ( $\Delta S_{r/t}$ ).

### III-2-1. Binding essays as useful tools to derive interaction information

Certainly, understanding structure-function relationship of either the protein or the carbohydrate ligand is a crucial step usually described when studying receptor-ligand complexes, nevertheless information about molecular dynamics on bio-surfaces should be enlisted. There is a vast catalogue of available binding assays designated for that objective. For example, Bio-Layer Interferometry (**BLI**), Surface Plasmon Resonance (**SPR**) and Isothermal Titration Calorimetry (**ITC**) are usually employed for determining affinity parameters of the interactions. **Glycan-array** could be added to this technical inventory to screen ligands that bind selectively to the lectin with known specificity. Technically, the common feature between the analytical techniques stated above is the fact that there is no need to isotopically label the molecules of interest. The sole need may be coupling of the interaction partner, either the lectin or the ligand (e.g., LPS), with a specific molecule that permits its covalent attachment to the surface, as for BLI experiments.

### III-2-1-1. Combining surface-based approaches provides comprehensive findings

BLI and SPR methods are based on a common principle which is about immobilizing one biomolecule (e.g., lectin) on a sensor surface and projecting an optical signal to detect a binding molecule fluid solution as it associates with the immobilized partner. As for ITC, it has a different experimental apparatus constituted of two cells dedicated for reference (usually filled with buffer or distilled water) and for receptor sample, separately. The temperature of the microcalorimeter must be equal between the two cells to avoid faking thermodynamics of the binding. Right away titration of ligand, the binding between ligand and the receptor evaluated at equilibrium, by measuring heat changes directly proportional to the amount of binding<sup>280</sup>. In the same line, it is reasonable to combine glycan-array with SPR to refine lectin-sugars binding depiction<sup>281</sup>. An example of such techniques coupling has been used for studying DC-SIGN lectin interaction with LPS inner core heptosides (terminal Di-heptose) (figure 49).



**Figure 49. DC-SIGN binding profiles from Glycan array and SPR experiments.** Top: Screening of DC-SIGN binding to various glycoconjugates ligands, at different concentrations of the lectin. Bottom: SPR results i.e., sensorgram (left) and binding curve (right) of LPS core terminal heptoside following glycan array analyses. Adapted from<sup>281</sup>.

By calculating binding kinetic parameters like the apparent  $K_d$  (Dissociation constant) of DC-SIGN interaction with heptoside, the authors found out that the binding is in the high nanomolar range.

Although monosaccharides are known to exhibit a specific weak binding toward lectins, their corresponding bacterial glycoconjugates binding epitopes (terminal residues without the rest of the molecule) typically show higher affinities through **multimeric interactions**. Such complex binding mechanisms between multivalent glycoconjugates and lectins, with increased valency, entitled as “cluster glycoside effect” usually lead to high affinity complexes<sup>282</sup>.

However, binding affinity outcomes might variate from one technique to the other for the same interaction system hence the necessity of employing combined approaches. In a study performed by Laigre et al, hexadecavalent GalNAc containing glycoconjugates were subjected to *Helix pomatia* agglutinin (HPA) lectin recognition, specific for  $\alpha$ GalNAc containing ligands, by using BLI as the main analytical technique<sup>283</sup>. The authors have suggested different binding affinities between this lectin and various structures ranging from nanomolar (observed with BLI) to millimolar ranges (studied by BLI coupled to SPR and ITC).

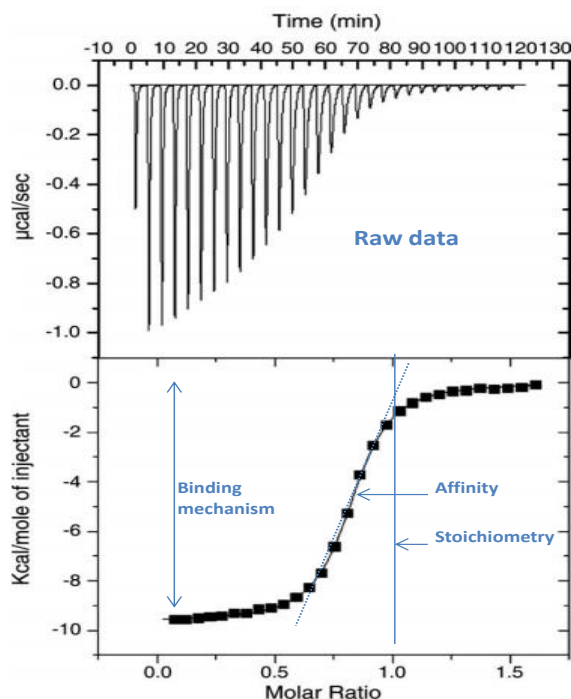
### III-2-1-2. Interaction energies estimation by ITC

SPR approach is usually coupled to BLI and/or ITC as useful complementary methods. As said, the latter approach offers the great possibility of providing access not only to thermodynamic parameters but also to the stoichiometry of interaction. However, typical aggregation behaviour of multivalent partners may generate unexploitable data by ITC which could be analyzed instead by SPR and BLI.

ITC measurements of molecular interactions originate from the released and absorbed heat (i.e. energy) per mole of ligand and could be analysed as binding isotherm providing thermodynamics parameters such as the change in enthalpy  $\Delta H^\circ$  (kcal/mol) upon binding and the Association constant  $K_a$  ( $M^{-1}$ ), and the number of binding sites per monomer of protein<sup>284</sup>. When both the lectin and the carbohydrate ligand (usually complex glycoconjugates) display **multivalency** in soluble cross-linked complexes, their affinity becomes anyway higher (due to **avidity**) approaching nanomolar range<sup>285</sup> whereas the affinity for oligosaccharides is frequently located in the millimolar range.

Binding entropies may variate from one ligand to another depending on conformational degrees of freedom adopted by either lectin or sugars that is also related to conformational changes and to water molecules contributions. Accordingly, Sabin et al<sup>286</sup> have found that galactoside and mannoside ligands exhibit favourable entropy ( $\Delta S_{conf}$ ) probably linked to the conformation of C6-

OH group that seems to be flexible during the binding to a bacterial lectin. In fact, they have investigated the ITC thermodynamics results from monosaccharides binding to PA-IIL lectin from *P. aeruginosa* human pathogen (figure 50) by comparing them to available crystallographic data (as reliable findings).



**Figure 50. Typical titration curve from Me- $\alpha$ -Fuc binding to PA-IIL showing thermodynamics parameters extracted from ITC data. Adapted from<sup>286</sup>.**

Given that each of these techniques has its proper advantages and disadvantages list, like when comparing (i) sensitivity detection of BLI that is hundred times lower than for SPR and (ii) quantity of sample required for ITC that is higher than for SPR, it turns out that by combining those approach in a clever way, it would be possible to overcome many drawbacks.

### III-2-2. Imaging approaches

As binding data should be well investigated to accurately understand the interaction system, single complexes that are formed following the interaction must be considered as well. Individual lectin-carbohydrate complex could be investigated by using **Atomic Force Microscopy (AFM)** which is designated to the study of single molecular recognition events, under physiological *in vitro* conditions. By using this technique, unbinding forces (that break the connections between the interaction partners) could be measured among which forces of biotin-avidin complex<sup>287</sup> as well as DNA mechanics<sup>288</sup>, with an estimated force sensitivity on the order of piconewton (pN) for the studied single molecules. Remarkably, this technique is gradually appreciated by researchers because it possesses unique features presented by its capability to

measure interaction forces at high lateral resolution for single molecule binding and to study biological sample without surface immobilization neither labelling<sup>289</sup>.

Interestingly, Touhami and co-workers have performed AFM force measurements to study lectin-carbohydrates interactions at cells surfaces to investigate aggregation events of yeast cells<sup>290</sup>. Adhesion forces between ConA-functionalized AFM probe and yeast cells (with surface exposed mannose carbohydrates) have been estimated to be of 121 and 117 pN for the interactions of cell surface oligoglucose with lectins and for mannose sugars binding to lectin-functionalized AFM probes, respectively, reflecting strong bindings.

As an additional example considering LPS as ligand, AFM has been wielded together with BLI in order to study LPS from *E. coli O111:B* and LOS from *E. coli B* interaction with an adhesin (gp37) from T4 bacteriophage<sup>291</sup>. The results highlighted a specific binding between LOS from *E. coli B* with an estimated  $K_d$  equal to  $1.05 \cdot 10^{-7}$  M. Unbinding AFM forces were measured too (70 +/- 29pN), confirming the strong interaction between the LOS and gp37 adhesin. However, LPS binding showed lower unbinding forces (46 +/- 13 pN) and the authors assumed it to be a non-specific binding of the protein with the solid surface with a low correlation between LPS binding and topography.

### **III-2-2-1. Cell-imaging as a tool to investigate spatial arrangement of lectin-bound membranes**

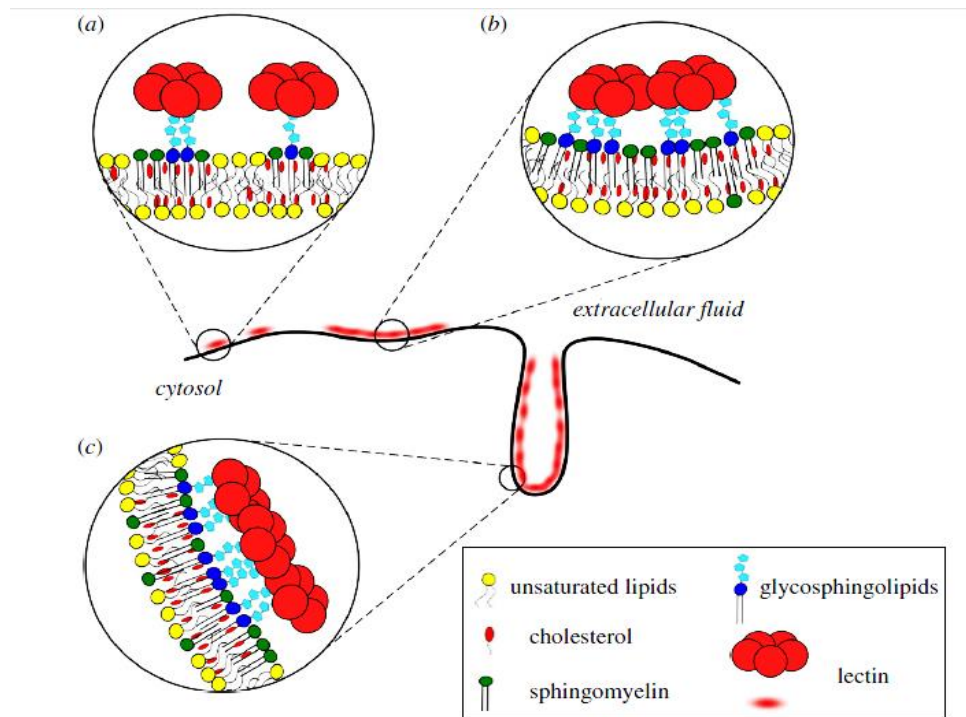
#### **III-2-2-1-1. Microscopic approaches**

When it comes to large interaction specimens, there is an attracting alternative to directly observe binding outcomes on high resolution images. Electron and light microscopies afford that opportunity. The general principle of all these scientific approaches one could be explained as said previously: an observation of a signal or “response” generated following the interaction of the analysed matter with an applied “treatment”. For microscopic analyses, the principle is much simpler than for NMR spectroscopy, but the complexity of this methods may revolve around sample pre-treatments and preparations. For **Transmission Electron Microscopy (TEM)**, a pre-treatment of the sample is essential and includes molecules dehydration and fixation at thin cut membranes to permit the passage of the applied electrons through the specimens. The sample must be chemically treated with heavy metals (more effective in electrons scattering) for TEM thus increasing the level of contrast which helps in ending up with well resolved images.

In both microscopic methods, electrons, or photons (light) act as waves that excite somehow the molecules constituting the sample that could be “seen” on the graph afterwards either following electron scattering by atoms (TEM) or fluorescence emission in a particular wavelength (**Fluorescence microscopy**). The latter requires sample preparation by labelling with a given fluorophore that has a particular excitation wavelength on which the microscope is configured to allow later fluorescence emission capturing.

The two imaging approaches mentioned above have been used during my PhD to visualize morphological behaviour of the different LPS and LOS molecules as well as their producing bacteria, either lonely or with the human lectins, thus gaining double outcomes from the ultra-structure (TEM) and the molecular location (Fluorescence) of these biomolecules. Finally, knowing that those two methods could be crossed technically<sup>292</sup> or by collecting data from both sides, this may help in obtaining a complementary and comprehensive outcome from the molecular interactions whether between LOS/LPS lattices and human lectins or amongst the proteins and the whole bacteria.

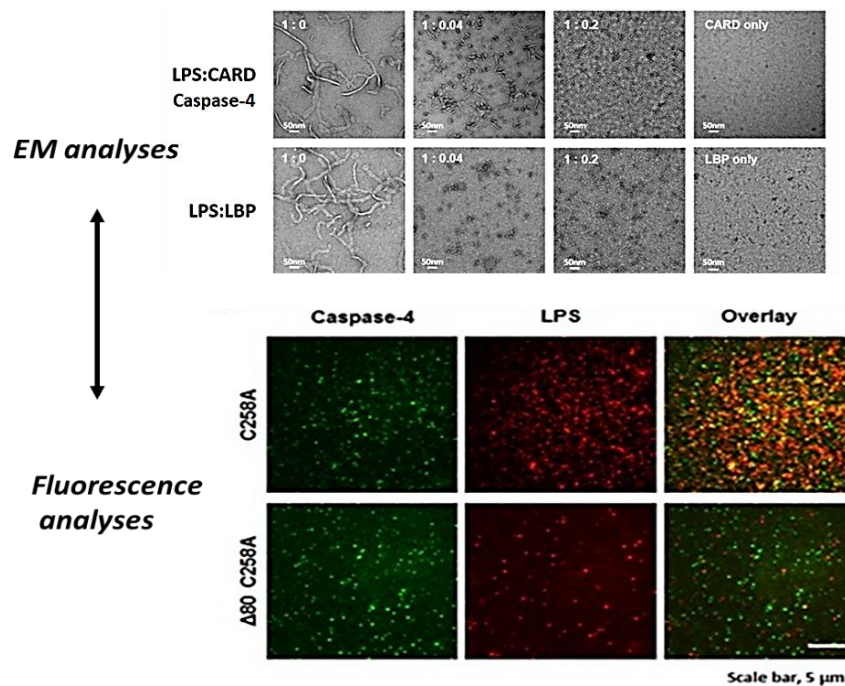
What would be extremely interesting in the case of membrane tethered glycoconjugate recognition by lectins, is to see the resulting morphology of this membrane following the binding by using high resolution imaging approaches like **super-resolution microscopy**. By using the later technique, it is possible to have a graphical vision of real-time biological events although at a microscopic scale, like membrane invagination. Within this regard, Galectin binding to glycosphingolipids (GSLs) was extensively studied<sup>293–297</sup>. A particular interest has been focused on Galectin-3 induced endocytic pathways involving Shiga toxins, such as STxB and serogroups of *E. coli* O157:H7 strain in which we were interested. Glycan binding features of such lectins play an essential role in membrane curvature. Figure 51 shows a graphical view of membrane curvature that occurs after lectin binding to the membrane.



**Figure 51. Snapshot of stepwise membrane curvature formation following lectin binding to carbohydrate structures of GSLs (a) and their cross-link to GSLs (b), generating plasma membrane tubules with many GSLs clusters (c). Taken from<sup>298</sup>.**

### III-2-2-1-1-1. Combining TEM and fluorescence microscopy for the study of LPS interactions

Elongated structure mixed with small spherical particles are the structures that usually represent LPS micelles in TEM and Cryo-EM graphs<sup>299</sup>. Those structures may change (or not) upon protein complex formation following an interaction, this is called the **disaggregation effect** on LPS induced by the protein. TEM approaches are very well designed to allow such observations. The physiological activity as well as the biophysical interactions of cytosolic caspases with an LOS from *E. coli* O55:B5 was revealed by fluorescence microscopy and TEM<sup>147</sup>. The author showed that caspase-4/LPS complex formation is strictly dependent on the molar ratio of caspase-4 to LPS (figure 52). Imaging of LPS aggregates may help in understanding its biological activity as it remains unclear whether who is the immunologically active state: LPS aggregates or monomers, even knowing that LPS aggregates are the most involved in direct initiation of immune cell activation<sup>300</sup>.



**Figure 52. Use of two imaging approaches to study LPS interaction with Caspase4.** Visualisation of LPS/caspase4 and LPS/LBP complex formation at different molar ratios by negative staining using electron microscopy and the effect of C258A mutated caspase4 on the interaction with LPS (both biomolecules labelled with fluorophores) by fluorescence microscopy. They used different versions of mutated caspase4 including a (1-80 amino acids) truncated constructs ( $\Delta 80$  caspase-4 (C258A)). Adapted from<sup>147</sup>.

### III-3-2-2-1-2. Lectin binding quantification by flow cytometry

As a quantitative and qualitative approach, **flow cytometry** has been used since the 1980s for studying lectins-carbohydrates interactions<sup>301</sup>. Briefly this technique relies on the measurements of fluorescent populations at the single cell level through light scattering and multiple fluorescence parameters, measured in two directions (side and forward) thus permitting to have relevant outcomes regarding the relative size of the particles and internal cell complexity. Giving that there are many exploitable lasers in the cytometer, many labels could be tested at the same time, like for example when studying cells interactions and viability, DNA and protein fluorescence are characterized in this case.

### III-2-3. NMR spectroscopy for interaction studies

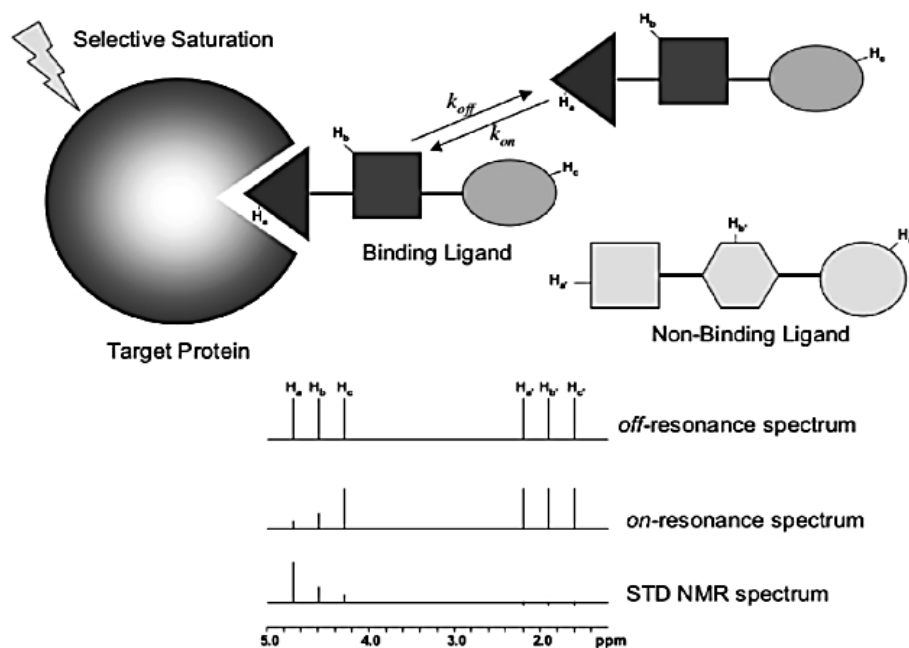
This part of the chapter is organized in such a way that few examples of NMR spectroscopic methods exploiting the ligand side or the protein side, are cited separately: at first ligand-based analyses are described and secondly protein-based methods, are presented.



### III-2-3-1. Saturation Transfer Difference (STD) NMR based approaches

Saturation Transfer Difference **STD-NMR spectroscopy** is a robust technique for transient molecular interaction analyses. This technique is based on the observation of NMR signals from the ligand, being a small molecule analysable by NMR, in large molar excess. In 1999, Meyer and Mayer have discovered STD-NMR spectroscopy when they used it at a first attempt to screen carbohydrates for activity binding towards WGA lectin<sup>302</sup>. The principle of this method is relatively straightforward and relies on recording two proton <sup>1</sup>H-1D NMR experiments: a first 1D spectrum is recorded at thermal equilibrium with an irradiation set at a region where no ligand nor protein signals are present (called off-resonance); a second one is registered where a selective saturation of the protein proton signals by a low-power radio frequency (usually by using successive “Gauss-shaped” pulses) for a specific period (saturation time: typically 1-2 seconds). The second experiment gives rise to the so-called on-resonance NMR spectrum provided that only the protein signals are irradiated and not those from the ligand (i.e., at -1 ppm or 11-12 ppm if the ligand has no aromatic resonances)<sup>262</sup>.

In case of binding between the protein and the ligand, a transfer of this saturation occurs via spin diffusion from the protein to the bound ligand being in proximity of the binding site (figure 53). Consequently, the intermolecular transfer of magnetization recruited by the ligand by rapid exchange with the protein, would be kept and cumulated in solution. In parallel, the transfer occurs between the ligand free and bound states and it is then observed in the difference NMR spectrum: off resonance spectrum– on resonance spectrum, by an increase in the corresponding signals. Certainly, ligands that are present but not interacting with the protein will not exhibit any magnetisation transfer and thus they are cancelled out in the difference spectrum. Observations from the bound ligand signals carry important outcomes such as ligand epitope mapping.



**Figure 53. Illustration of intermolecular saturation transfer during protein-ligand binding in fast exchange regime.** 1D  $^1\text{H}$ -STD experiments applied to the system and the derived STD (difference) spectrum are shown. Taken from<sup>303</sup>.

### III-2-3-1-1. Epitope mapping of the protein-bound ligand

Meyers and Mayer have shown few years later that a qualitative analysis could be done thanks to this method, that is the **epitope mapping** of the ligand. In fact, the transferred saturation to the bound ligand is represented by a relative distribution among the protons of the ligand and is indicative for spatial proximities between the protein and the ligand. Thus, the closer the protons to the protein binding site ( $< 5\text{\AA}$ ), the stronger is the corresponding NMR STD signals.

As a simple qualitative analysis, epitope mapping is based on estimation and attribution of relative percentages of saturation received by ligand protons during the interaction. They are deduced from normalized STD intensities which depend on the saturation time and on the excess of the ligand tested. It could be assumed that relative percentages of protons contributing to the binding may be deduced from the observed STD signals by using this formula:  $100 \times (I_{\text{off-resonance}} - I_{\text{on-resonance}}) / I_{\text{off-resonance}}$ , scaled to 100 % relative to the most intense absolute STD that is attributed to the closest proton ligand to the binding pocket of the protein<sup>304</sup>. Commonly, the use of STD NMR approaches is not restricted to qualitative studies, but the previously mentioned methodologies usually end up with qualitative and quantitative analyses just by combining different methodologies not only NMR related but also one of the methods described in this chapter would be sufficient.

### III-2-3-2. Interaction studies by NMR from the protein perspective

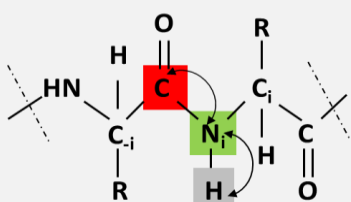
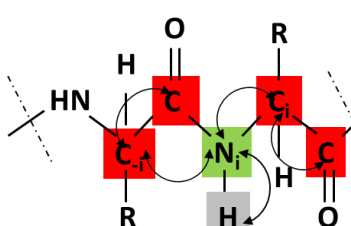
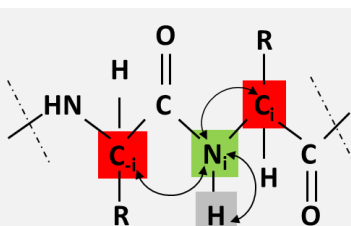
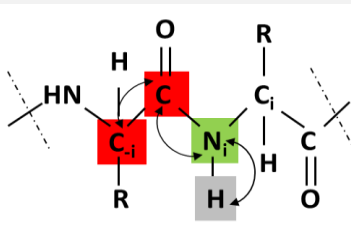
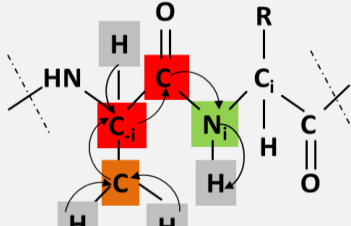
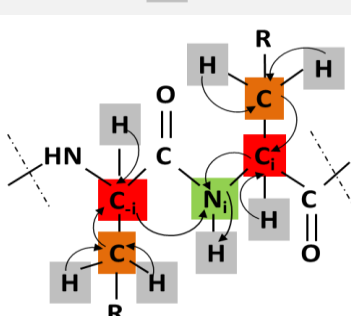
In addition to ligand mapping, NMR spectroscopy is used also to map residues from the protein binding site upon interaction with its ligand and, in case of strong binding, the structure of protein-ligand complexes may be determined. NMR assignment of all residues composing the backbone of the proteins is required to identify each residue and thus facilitate its localisation after ligand binding. Initially, many NMR experiments have to be recorded on the protein sample at different dimensions (from 1D to 3D) in order to get complete information about protein residues nature and sequence.

#### III-2-3-2-1. Backbone assignment of the protein by using NMR facility

Besides typical spectra like  $^1\text{H}$ -spectrum (1D) and  $^1\text{H}$ - $^{15}\text{N}$ -HSQC spectrum (2D), three dimensional spectra (3D) are usually registered for **protein backbone assignment**. This type of spectra implies all three nuclei:  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$  as major atoms constituting proteins and more precisely the backbone. Three dimensional spectra could be imagined as a cube with three axes x, y and z and, each axe represents one dimension that correspond to  $^1\text{H}$  for x-dimension,  $^{15}\text{N}$  for y and  $^{13}\text{C}$  for z. Considering that at any plane inside the cube, like for instance y-plane ( $^{15}\text{N}$ ), contains cross peaks from  $^1\text{H}$  and  $^{13}\text{C}$  and so x- and z-dimensions, it becomes possible to visualize correlations between all three nuclei by “walk-through” these plans flour by flour until visualizing the whole backbone lattice. This “walk-through” is based on the use of strips from 2D spectrum that consists of “fixing” one dimension and looking at the two others at different ppm positions, this enables finding the right connections between the three nuclei.

To do that, several types of 3D NMR experiments (resumed in table 4) that employ different transfers of magnetization between these nuclei are considered, indeed, to obtain resonances of  $i_1$  and  $i_{-1}$  residues (i.e., actual and previous) and correlate them to get the assignment. In general, there is a typical manner with which magnetization is transferred in between nuclei (for instance from  $^1\text{H}$  to  $^{15}\text{N}$  or  $\text{C}_\alpha$  or  $\text{C}_\beta$  and from  $^{15}\text{N}$  to  $^{13}\text{C}=\text{O}$ ) in order that it ends up at the  $^1\text{H}$  for NMR detection. As an example from the HNC0 experiment, magnetization is transferred first from  $^1\text{H}$  to  $^{15}\text{N}$  of backbone amide groups, then it passes to the carbonyl through  $^{13}\text{C}$  via J-coupling, and it goes back to  $^1\text{H}$  through  $^{15}\text{N}$  for detection, the 3D spectrum results from chemical shifts of the three nuclei  $^1\text{H}$ ,  $^{15}\text{N}$  and  $^{13}\text{C}$ .

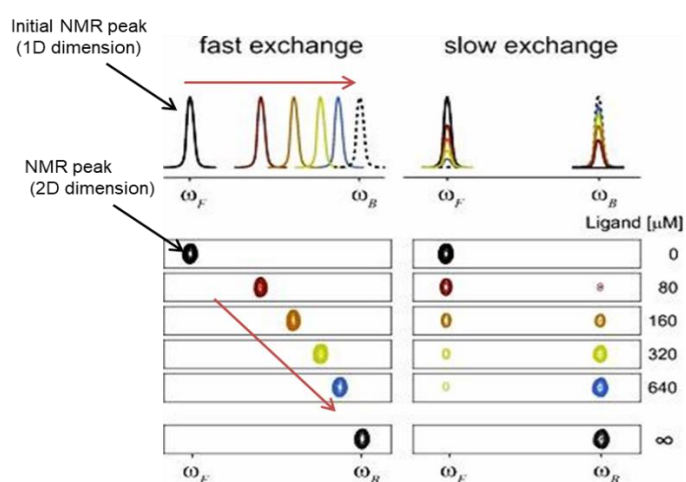
**Table 4. Typical 3D-NMR experiments used for protein backbone assignment.** Colored nuclei correspond to the ones that are experiencing magnetization, and for clarity,  $C_{\alpha}$  are in red and  $C_{\beta}$  in brown; R signifies main chain of residues; and curved arrows indicate the magnetization transfer evolving chemical shift or J-coupling. Depending on the experiment, the chemical shift might be evolved for some nuclei and not for the others. inspired from<sup>305–307</sup>.

3D Experiment	Transferred magnetization	Backbone assignment comments
HNCO		<ul style="list-style-type: none"> <li>- requires conjunction with HN(CA)CO for the analyses</li> <li>- 3D correlations enable to map <b>C=O chemical shifts</b> of the preceding residue (i-1) attached to each NH group of the backbone.</li> </ul>
HN(CA)CO		<ul style="list-style-type: none"> <li>- in combination with HNCA, HN(CO)CA and HNCO for backbone assignments</li> <li>- generates <b>two observable C=O signals for one NH group strip</b></li> <li>- <math>^{13}\text{C}</math>O cross-peaks from i-1 residues are less intense than those from i residues due to stronger <math>\text{N}_i\text{-C}_{\alpha i}</math> coupling.</li> </ul>
HNCA		<ul style="list-style-type: none"> <li>- for each NH 2D strip, <b>two <math>C_{\alpha}</math> signals</b> are obtained among which a weak <math>C_{\alpha}</math> cross-peak that corresponds to the preceding <math>C_{\alpha i-1}</math> and it is observable in the HN(CO)CA spectrum besides thus facilitating the backbone assignments.</li> </ul>
HN(CO)CA		<ul style="list-style-type: none"> <li>- the resulting spectrum is similar to the HNCA but discriminating for the <math>C_{i-1}\alpha</math>, so only one <math>C_{\alpha}</math> signal is visible for one NH strip</li> <li>- must be overlaid with HNCA spectrum for residues attributions</li> </ul>
HN(CO)CACB		<ul style="list-style-type: none"> <li>- part of concoction of experiments recorded also for backbone assignment, including HSQC and HNCACB</li> <li>- determines <b>the <math>C_{i-1}\alpha</math> and the <math>C_{i-1}\beta</math></b> of the previous residue</li> </ul>
HNCACB		<ul style="list-style-type: none"> <li>- provides specific outcomes for <math>C_{\alpha}</math> and the <math>C_{\beta}</math> of i and i-1 residues</li> <li>- four cross-peaks are detected for each corresponding amide backbone group NH: <b>two <math>C_{\alpha}</math> and two <math>C_{\beta}</math></b>, with opposite signs (negative or positive)</li> <li>- <math>C_{\alpha}</math> and <math>C_{\beta}</math> from preceding residue have weak intensity</li> </ul>

### III-2-3-2-2. NMR Titration experiments as a tool for studying protein-ligand interaction

For the investigation of protein-ligand interactions, looking at the protein contributions presents useful tool for a comprehensive study. Titration of increasing ligand concentrations to a given protein is the basic protocol used within **NMR titration** experiments. To do that, the protein has to be isotopically labelled with  $^{15}\text{N}$  isotope, which has a high sensitivity and signal dispersion in NMR. By recording  $^{15}\text{N}$ -heteronuclear NMR related experiments including HSQC experiment, some analyses could be done on the protein in its Apo state (native, without ligand). The signals observed from such NMR experiments correspond to every residue constituting the protein.

After stepwise addition of the ligand, signals from the protein residues experiencing changes in their chemical environment would be perturbed following their interaction with the ligand. Those perturbations are highlighted by chemical shifts and/or intensity changes and they are observed in the 2D overlapped spectra resulting from each titration point. Cross-peaks (H-N pairs) displacement or intensity changes in the HSQC spectrum, reflects the protein-ligand exchange regimes (figure 54). For example, when the exchange is fast, the **chemical shift perturbations (CSPs)** of the altered residues would move gradually towards the “new” protein peaks which result from protein saturation with the added ligand, describing both free and bound states. In this case, the  $K_{\text{off}}$  is significantly larger than the chemical shift difference between free and bound protein residues. As for slow exchange regime, free and bound protein peaks change of relative intensities by the disappearance of the peak corresponding to the free protein residue and the simultaneous appearance of resonances of the complex, during titrations with the ligand<sup>308,309</sup>.



**Figure 54. Dependency of NMR cross-peaks shapes and trajectories to slow and fast exchange regimes during protein binding to the ligand. Adapted from<sup>309</sup>.**

Apparent binding affinity values could be estimated from chemical shifts. Indeed, estimation of  $K_d$  values is possible for weak-binding ligands by using the following formula<sup>309</sup> :

$$\Delta\delta_{\text{obs}} = \Delta\delta_{\text{max}} \left[ \frac{([A]_t + [B]_t + K_d)}{2[A]_t} - \frac{\sqrt{([A]_t + [B]_t + K_d)^2 - 4[A]_t[B]_t}}{2[A]_t} \right]$$

where  $\Delta\delta_{\text{obs}}$  is the chemical shift change comparable with the free state,  $\delta_{\text{max}}$  corresponds to the maximum  $\delta$  when the protein is saturated with the ligand, and  $[A]_t$  and  $[B]_t$ , are total protein and ligand concentrations, respectively.

In addition to HSQC NMR experiments, transverse relaxation-optimized spectroscopy (TROSY)<sup>310</sup> is usually applied to increase the resolution of NMR signals, especially for large sized proteins which have the tendency to give rise to broad 2D peaks due to their fast transverse relaxation (a complex NMR phenomenon called  $T_2$  relaxation, that strictly depends on the correlation time that is, in turn, dependent on the protein size)<sup>311</sup>.

### III-2-4 Other methods to analyze molecular interactions

There are additional ways to make structural and functional studies by looking for example at the electronic distribution of biomolecules in **crystal structures** like it has been most frequently the case for big molecules like membrane proteins but also complexes like carbohydrates-lectins. As an example, the interaction between a galactose binding lectin (dubbed AJLec from sea plants, with a C-type CRD domain) to lactose (Gal-Glc) has been analysed by solving the crystal structure of the complex using X-ray crystallography, in parallel with ITC measurements and glycan array screening that confirmed: its specificity for Gal and  $\beta$ -linked Gal in complex structures and, its inability to bind to glycans containing N-Acetylated galactose (GalNAc) despite the high affinity of this lectin to that residue<sup>312</sup>.

Unfortunately, it is not feasible so far to perform X-ray crystallography on LPS and LOS structures because of the elevated flexibility of the saccharides cores generating “false” models, in that case this methods becomes insensitive for structural and dynamic investigations for our biomolecules.

It is implausible to introduce several *in vitro* methodologies, used for lectin-carbohydrate interaction studies, without referring to **computational techniques** which have often been called for to complement or double check binding mode and properties. These computer studies are

performed using specific software with various procedures (e.g. FlexiDock, AutoDock) considering the crystallographic structure of the receptor (lectin) if available or an homology model if not. In general, one starts **Molecular Docking** by testing several versions of the ligands with different sizes, affinities (if known) and chemical properties and synchronously eliminating those who do not show favoured binding. Contrariwise, those who have the best outcomes in terms of energy of binding and ligand orientations, are the ones selected.

As presented along with chapter II, in almost all C-type lectins, there is at least one  $\text{Ca}^{2+}$  binding site in their CRDs which might increase the chances in chemically attracting non reducing ends of polysaccharides that usually bear phosphate groups, for this reason, biological relevance of such findings has to be demonstrated. One may use **Molecular Dynamics simulations** for further refinement after docking<sup>313</sup>. MD simulations enables following the movements of atoms (dynamics) in an interaction system over time. For structure confirmation, this method is also used complementary to X-ray crystallography that usually ends up with averaged/frozen structures.

Decisively, combining at least three type of experiments counting in vitro analyses like binding essays, spectroscopic investigations, and computational studies, for example, is crucial for the analyses of lectin-carbohydrate interactions including the LPS-lectin system.

## ***Chapter IV.***



## IV. LPS-C-type lectins interactions

Here, I discourse the available studies about the interaction between LPS and mammalian CLRs that have been published so far, and I introduce the origins and the structures of different LPS and LOS molecules that we concerned during the present PhD work.

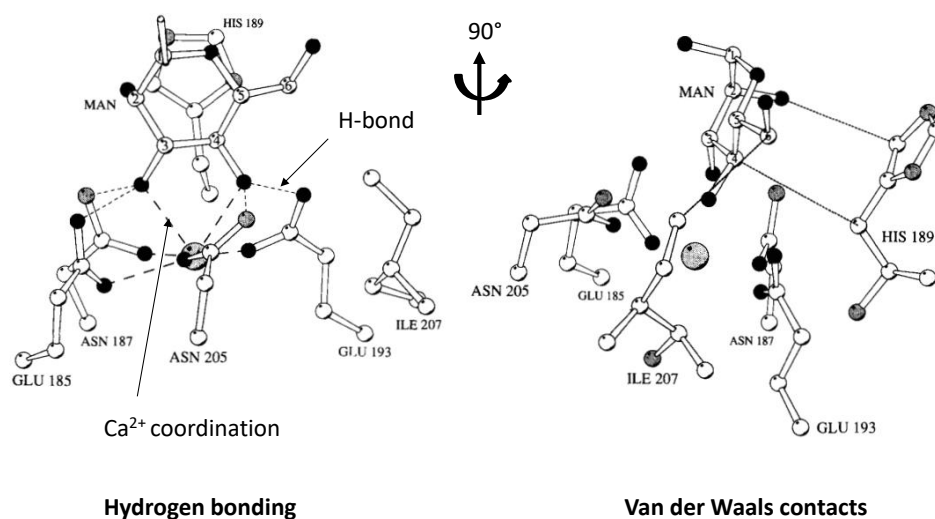
### IV-1. Examples of up-to-date studies of LPS-lectins interactions

#### IV-1-1. Mannose Binding Lectin-LPS interaction

MBL or mannose-binding lectin is a C-type lectin that can recruit ligands underneath its trimeric structure which may adopt many oligomer conformations up to hexamers of trimers<sup>314,315</sup>. MBL oligomers are comprised of CRD domain at the C-terminal region linked to a hydrophobic neck region and a collagenous part and N-terminal cysteine rich domain. Its carbohydrate preferences are relatively broad including D-mannose, N-acetylglucosamine GlcNAc and L-fucose from various types of pathogens like bacteria, viruses, and yeast. MBL circulates freely in serum and, when binding to microbial surface polysaccharides, MBL indirectly activates the complement pathway which results in microbe clearance. This indirect activation occurs mainly via inducing the “lectin pathway” right after pathogen recognition, with the participation of other receptors like collectin and ficolin and, since then, a cascade of interactions including complexes formation and enzymatic reactions (convertase C3b) take place in an orchestrated manner<sup>316</sup>.

This lectin binds specifically to terminal GlcN of *Salmonella* LOS but not LPS<sup>317</sup>. O-Antigen recognition by MBL was demonstrated to be selective for some other strain mutations of *Salmonella* through ELISA tests and to induce afterwards complement pathway (convertase C4c production)<sup>318</sup>. In vitro analyses (e.g., ELISA, flow cytometry) of *Yersinia enterocolitica* LPS mutants -truncated at different levels giving rise to rough and deep-rough variants- targeting by MBL have shown that this C-type lectin is able to bind to a variety of LPS glycoconjugates containing O-Antigen (homo-polymannoses) but not to LPS variants lacking the outer core portion<sup>319</sup>.

Despite its low affinity to carbohydrates, it appears that this lectin has a strain-specificity that could be strongly dependent on repetitive sugar residues in one hand and, on multivalent interactions resulting in avidity-based contacts within the LPS sugars sequence and portions, on the other hand<sup>320</sup>. Still, there is the only atomic level investigation of MBL-carbohydrate binding which dates from 1994, by NMR spectroscopy and site directed mutagenesis (figure 55).



**Figure 55. Binding interactions of terminal mannose oligosaccharide at the canonical binding site of rat MBL CRD.** For clarity, atoms are colored as following: Oxygen in black, Carbon in white, Nitrogen in grey and calcium in light grey. Adapted from<sup>320</sup>.

Conversely, MBL binding to MASPs (associated proteins) has been much considered<sup>321</sup>. This lectin has a multifunctionality feature that includes protein-carbohydrates usually with low affinity and protein-protein nanomolar range interactions, needed for lectin pathway stimulation. This could be due, at least in part, to its heterogenous and rich structure in terms of domains.

#### IV-1-2. DC-SIGN lectin targets bacterial LPS

As described in chapter I-1-2-1, DC-SIGN is a tetrameric C-type lectin that interacts with ligands from many origins. Having a very broad pathogen recognition outline, DC-SIGN is designed to mediate interactions with mycobacterial PAMPs, and sense bacterial infections besides<sup>322,323</sup>. It appears that an *in vitro* interaction between DC-SIGN and LPS variants from *Helicobacter pylori* makes the T helper cell (Th)1 development occluded, which might contribute to a long residence of *H. pylori* pathogen in the stomach<sup>322</sup>. Using phagocytosis assays on several mutants, it has been shown that GlcNAc residues from core LPS produced by many Gram-negative pathogens, including *Salmonella* and *E. coli* are recognized by DC-SIGN<sup>324</sup>.

Mutants of *E. coli* producing different variants of LPS have been studied as ligands for human DC-SIGN. Klena and colleagues have demonstrated the crucial role of the O-antigen portion in protecting enteric *E. coli* bacteria from phagocytosis<sup>325</sup>. In fact, the interaction between dendritic cells (DCs) via DC-SIGN and *E. coli* K12 mutants expressing LOS glycoconjugates led to phagocytosis following DC-SIGN binding to the core of LOS. Conversely, *E. coli* expressing the whole LPS molecules escaped from DCs action and the authors supposed that this is due to the

presence of the O-Antigen which hid the core exposition to DC-SIGN. Nonetheless, deeper understanding of molecular mechanisms behind these interactions is awaiting further studies.

DC-SIGN has been studied in another context (physiopathology of kidney injury) where it has been demonstrated that this lectin interacts with TLR4 activated by the presence of LPS. Using immunoblotting and immunofluorescence essays together with confocal microscopy (figure 56), the authors have employed epithelial cells from humans suffering from kidney diseases (HK-2 cells) to search for possible LPS-induced interactions between DC-SIGN and TLR4<sup>326</sup>. In this study, they showed that LPS induces the expression of DC-SIGN by those cells favouring endogenous interactions between DC-SIGN and TLR-4, without questioning about potential LPS-DC-SIGN interaction, prior to TLR4 activation for example. It seems that DC-SIGN was considered only as an adhesion molecule in this study.

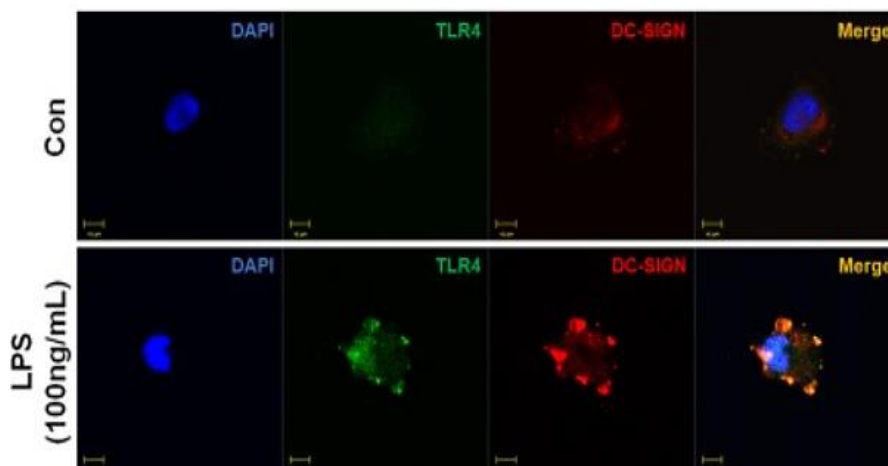


Figure 56. **Confocal microscopy images with fluorescently conjugated antibodies specific for TLR4 and DC-SIGN produced by HK-2 cells, before and after LPS addition.** Taken from<sup>326</sup>.

In parallel, they evaluated cytokines production to be increased following LPS-induced DC-SIGN expression. However, it is not clear whether DC-SIGN signalling induced by LPS can happen independently from TLR4. Thus, further studies should be addressed about DC-SIGN-LPS interactions and the ensuing mechanisms of the inflammatory responses (e.g., Cytokine production).

#### **IV-1-3. MGL lectin recognizes *Campylobacter jejuni* LOS**

The interaction between the ECD of human MGL (MGL-Fc chimera) and glycoconjugates from *C. Jejuni* (expressing several potential MGL ligands including N-glycosylated proteins and LOS) has been proved through the analyses of *C. Jejuni* strain Wild-type and mutants of relevant genes (such as *pgl*) implicated in LOS biosynthesis. In fact, by using HPA lectin (GalNAc binder) as

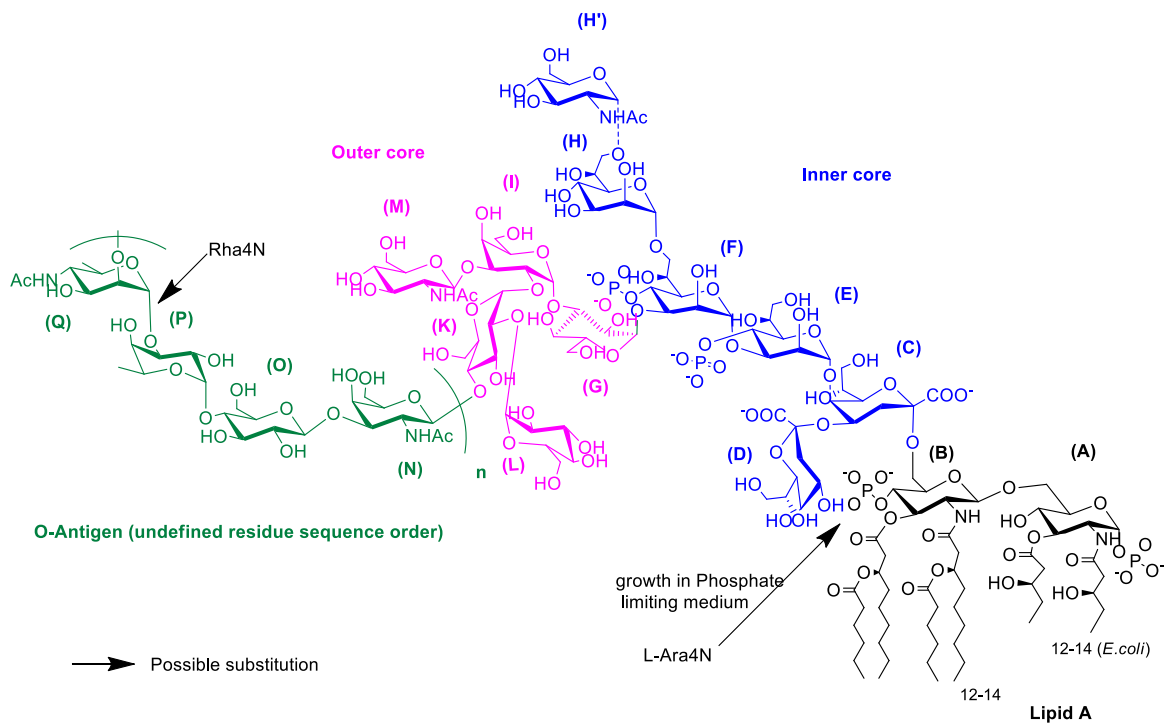
template, the authors showed that the deletion of that gene led to the loss of the incorporation of terminal GalNAc sugars in *C. Jejuni* LOS which resulted in total absence of binding with MGL. As a negative control, *C. Jejuni* mutants lacking a Glc branch (*pgII*) were considered and showed that MGL binding to the expressed LOSs was not affected. In this study, the authors used Western blotting, lectin binding essays and silver staining, which indeed provided insights about MGL specific binding to active GalNAc-containing glycoconjugates not only *C. Jejuni* LOS but also LPS from *E. coli* with a DH5 $\alpha$ -*pgI* machinery<sup>201</sup>. To evidence that these interactions are mediated solely by MGL construct and not Fc (IgG), the authors carried out inhibition assays by adding excess of GalNAc or MGL-blocking antibody. In addition, their microscopic analyses showed a high density of MGL at the surface of *E. coli* DH5 $\alpha$ -*pgI* thus confirming MGL specificity towards GalNAc containing glycoconjugates. However, molecular details of bacterial LPS-MGL interactions are missing, and this is the sole published study that describes glycan motifs recognition by human MGL.

## **IV-2. LPS glycoconjugates of interest**

Within the present project, three LPSs were selected and studied for their importance in triggering innate immune response, pathogenesis, and for their distinct carbohydrate scaffolds. LPS from *E. coli* **O157:H7** strain, **LOSs** from *E. coli* **R1** and *E. coli* **R3** strains were considered in this PhD work. Another LOS from *B. cenocepacia* *J2315* pathogen was isolated and chemically treated and analysed besides.

### **IV-2-1. *E. coli* Sakai LPS**

*E. coli* pathogens could be found everywhere hence increasing the probability of being infected. Enterohemorrhagic and verotoxigenic *E. coli* **O157:H7** dubbed as “*Sakai*” strain is responsible for severe collective food-borne infections and produces smooth type **LPS** bearing an O-Antigen O157 (and a flagellar antigen namely H7). The structural studies that have been established about LPS O157:H7 so far, discussed one possible core oligosaccharide structure and potential structures of the O-Antigen part similar to other possibly pathogenic strains (figure 57).



**Figure 57. Summarized *E. coli* O:157 LPS structures that have been resolved to date.**

The core OS of LPS Sakai shares identical structure with the LOS from *E. coli* R3 strain, supporting the hypothesis that *E. coli* O157:H7 strain is a mutant of *E. coli* R3 since the latter have been found in verotoxigenic isolates belonging to the common enterohemorrhagic *E. coli* (EHEC) serogroups O157<sup>327,328</sup>. Moreover, the O-antigen of LPS from *E. coli* O157:H7 strain shares very similar structure with the one from *Citrobacter* strain and its serological reactivity has been tested showing strong cross-reactivity with that strain<sup>329</sup> and identical structure with *Salmonella* serogroup O30 O-Antigen<sup>330</sup>. Interestingly, several strains including *Salmonella*, *Citrobacter* and *Pseudomonas*, have been studied for their serological reactivity in mouse models because they produce similar epitopes to *E. coli* O157 O-antigen<sup>331</sup>.

The O-Antigen of LPS Sakai is believed to contain a number of chains bearing four residues : [4)- $\beta$ -D-Glc-(1-3)- $\alpha$ -D-PerNAc-(1-4)- $\alpha$ -D-GalNAc-(1-3)- $\alpha$ -L-Fuc-(1) $_n$  which have been identified by mass spectrometry and in solution NMR spectroscopy<sup>332</sup> and by gene sequencing of its biosynthetic enzymes<sup>333</sup>. Nevertheless, it is still under investigation whether if LPS is involved in *E. coli* O157:H7 uptake by immune cell receptors, as a strain of important significance for public health and food industry.

#### IV-2-2. LOSs from *E. coli* R1 and R3 strains

Apart from studying smooth-type LPS, we are attentive about rough-type LPSs (LOS) from *E. coli* R1 and R3, which are of particular interest. In fact, these two strains have been most

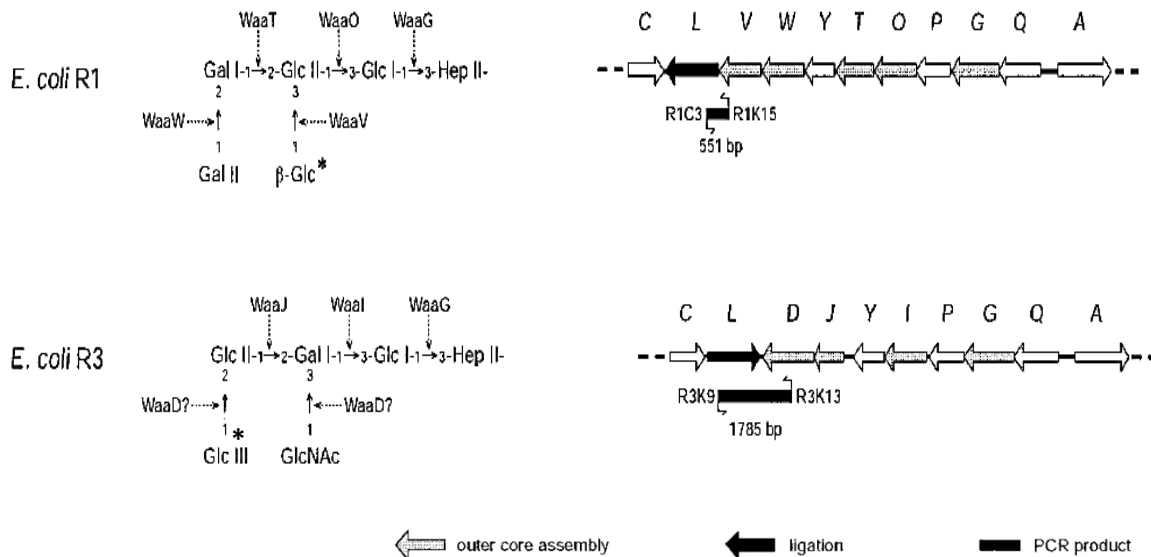
frequently found in *E. coli* clinical isolates and they are believed to be the causative agents for gastro-intestinal and urinary tract infections.

Characteristically, the LOSs produced by commensal *E. coli* bacteria exhibit significant shared structures, certified by an identical inner core-Lipid-A architecture<sup>334</sup> and a highly similar but not identical outer core composition. R1 and R3 are the most abundant core OS types found in verotoxigenic *E. coli* (VTEC) isolates from humans. The synthesis of these two LOSs is controlled by the chromosomal Waa (formerly rfa) loci which was previously determined using comparative sequence analyses<sup>75</sup>. PCR-based LPS core typing system has been developed for commensal *E. coli* isolates by Amor et al to study core OS distributions in human and animal verotoxigenic *E. coli* (VTEC) isolates representing the predominant enterohemorrhagic *E. coli* (EHEC) serogroups<sup>328</sup>. The authors provided genes determining structural features of all core OSs in *E. coli* (R1, R2, R3, R4 and K-12) where unique DNA sequences were exploited for core OS diagnostics of R1 and R3.

**Table 5. Core OS types distribution in VTEC *E. coli* isolates from humans and animal. Taken from<sup>328</sup>.**

Group	Serotype	No. of isolates tested with core type:					Total no.
		R1	R2	R3	R4	K-12	
VTEC associated with human disease (human and animal isolates)	O26:H11			5			5
	O55:H7			1			1
	O91:H21			4			4
	O103:H2			4			4
	O111:H <sup>-</sup>			12			12
	O111:H8			3			3
	O113:H4			1			1
	O113:H21	3					3
	O145:H <sup>-</sup>	3				1	4
	O145:H8	1					1
	O157:H7			14			14
	O157:H <sup>-</sup>			4			4
	Total (%)	7 (12.5)	0 (0)	48 (85.7)	0 (0)	1 (1.8)	56 (100)

From one side, R1 core OS predominated in serogroups O113 and O145 and from the other side R3 core type was the most abundant in O157:H7 corresponding to LPS from *E. coli* O157:H7 strain and O111 serogroups (Table 5).



**Figure 58. Structures of the R1 and R3 outer core OSs from the LPS/LOSs of *E. coli* and genetic organization of the central *waaQ* operon (core OS biosynthesis) loci.** The asterisks indicate the points of attachment of O-antigen in smooth-type LPS bacteria. Adapted from<sup>328</sup>.

As said, the R1 and R3 core OS structures are slightly different and the majority of the corresponding *waa* locus sequences are globally similar, especially in their specific *WaaL* genes<sup>76</sup>. However, these OS structures differ in the nature and linkage of the  $\beta$ -linked hexose residue attached to GlcII in R1 and GalI in R3 and the terminal disaccharides in the outer core OSs (figure 58), reflecting differences in the activities of the *waaV/waaW* (R1) and *120ly* (R3) gene products. Still, many interrogations need to be clarified regarding the interference between core OS biosynthesis and LOS function as PAMP and how these distinct *E. coli* LPS/LOS structures differentially modulate the strength of the host immune response.

Knowing that these two strains are influenced by or responsible for several diseases related to gut microbiota<sup>225</sup> (same for *E. coli O157:H7*), the study of their molecular recognition by humans' immune cell receptors is urgently needed in order to aid understanding binding specificities by CLRs for an eventual development of sugar-based therapeutics.

#### IV-2-3. LOS from *BC J2315* pathogens

Another LOS was considered during my PhD that is produced by *Burkholderia cenocepacia* (*BC*) which is one of the most important causative agents of septic shock in Cystic Fibrosis patients. Among *BC*, rough and a smooth chemotypes have been observed in clinical isolates<sup>335</sup>. *BC J2315* or LMG 16656, is a strain that has predominantly LOS glycoconjugates in its OM. However, it may differentially express smooth LPS depending on the local environment and growth conditions.

From a structural point of view, LOS from *BC J2315* possess a highly diverse structure in terms of sugars number and nature which have been elucidated by several techniques including compositional analyses, NMR spectroscopy and GC-MS / MALDI spectrometry<sup>259</sup>. The structure agreed from the isolated OS fractions contains several type of sugars, mostly heptoses were located in the inner and outer core, unusual residues like Ko (121lycerol-d-talo-oct-2-ulopyranosonic acid) and QuiN (2-amino-2,6-dideoxyglucose). The investigation of chronic lung infections involving Gram-negative bacteria, including *Burkholderia* represents a challenging task.



## ***Chapter V.***

## V. Objectives and framework of the PhD project

Studying the structural bases beyond LPS-lectin direct interactions is crucial for understanding pathogen recognition by humans' immune systems and this would eventually aid healing infections caused by **LPS** producing bacteria like *E. coli* O157:H7. There are only few available studies on these large interaction systems, and they are enumerated in the previous chapter.

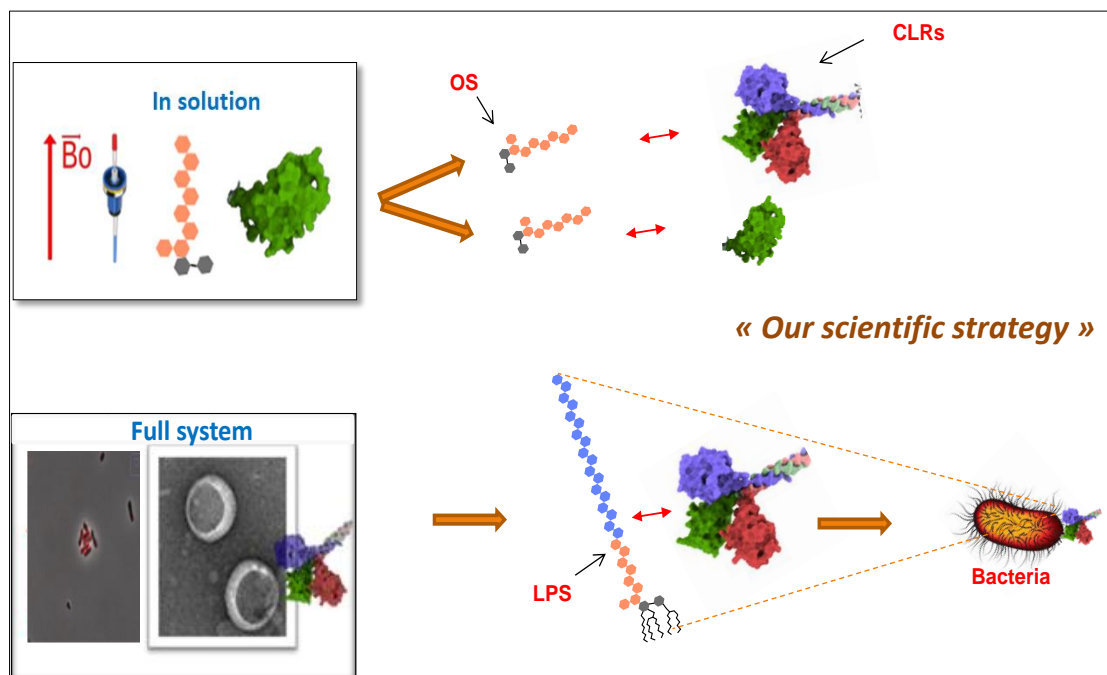
Published studies have suggested diverse binding modes and affinities involving **oligomeric human CLRs**, pointing out monovalent or multivalent interactions, with single sugars or with few LPS ligands that have been little studied. The **ECD** represents the whole lectin extracellular structure that directly encounter pathogens during innate immunity and may somehow tune contacts with LPS epitopes via the **CRD** domains. The latter possess a common conserved overall structure among CLRs but exhibits various binding regions that depend, at least in part, on the ligand being present. Thus, it is hard to forecast the binding scheme between bacterial LPS and human lectins without analysing separately the different parts of the interaction system which was the main purpose of the project.

This PhD project is part of GlycoAlps program (<https://glycoalps.univ-grenoble-alpes.fr/>), a wide community of researchers essentially from Grenoble and Europe. Those collaborations are fostered by GlycoAlps in order to share scientific outcomes principally from carbohydrates and their related conjugates (e.g., LPS) in biotherapeutics and many other subjects linked with structural and functional complexity of carbohydrates. As said several times along this manuscript that structural features of LPS molecules render them hard to explore experimentally and the task would become even more difficult when considering the large interaction system counting lectins. Our collaborative framework would allow us to remedy LPS structural complexity through the use of specific techniques in the first part of the PhD realized at my Italian laboratory in Naples; and using several approaches including NMR spectroscopy and biophysical methods to analyze the interactions at different levels from molecular to cellular scale, during the second period of my PhD with my French group in Grenoble. Besides, the production of the studied human lectins, from our collaborators at IBS in Grenoble (F. Fieschi's group, also member of GlycoAlps community) provided an important support to our PhD work by providing several lectin samples.

Covering the distinct aspects of LPS-lectin interaction is possible if considering the following objectives:

- *Successful isolation, derivatization, and chemical analyses of all **LPS/LOS/OS** from wild-type and mutants,*
- *Setting-up a well-adapted scientific strategy by selecting the adequate experiments to be used,*
- *Rapid screening of interacting systems to be considered and analysed accurately,*
- *Seeking to find missing structural data from human lectin of interest i.e. **MGL**,*
- *Providing new and/or putative insights about the mechanisms of bacterial LPS recognition by human MGL.*

To achieve our final goals, it is of paramount importance to ensure appropriate experimental conditions according to the actual interaction network, and this would permit us to attain high relevance. To do that, we sought to discompose the interaction system into two main networks counting fragmented interactors analysable in solution, and the whole molecular organisations of the interaction partners. This would help us to unveil the interaction between LPS and lectins at many levels. The graphical scheme summarizes the scientific strategy that has been set during these three years of PhD.



The rationale behind these combined approaches is to follow and investigate molecular events at different levels, by starting gradually from the atomic level on soluble ligand fractions

(delipidated LOS) by means of liquid NMR studies, to the molecular and cellular states where intact ligands and bacteria were considered, thus we would be able to puzzle out a global view on the interaction between bacterial LPSs and human CLRs.

In the following chapters, the results obtained during my thesis will be presented and discussed. Chapter VI includes the results obtained when I started my PhD, from the production and analyses of several bacterial LPS/LOS, by means of GC-MS and NMR spectroscopy. The structural complexity and diversity, within the selected *E. coli* LPS/LOS glycoconjugates, is extremely challenging because they could be considered as homologous ligands, originating from mother strains, with versatile contributions during pathogen recognition owing to their fine structural differences.

Chapter VII is meant to present the results from rapid screening of an interacting *E. coli* LPS-lectins counting the three human CLRs in which we firstly directed our attention (DC-SIGN, MGL and Langerin). Those examinations were performed mainly using NMR spectroscopy coupled to EM. In solution NMR analyses of the CRD backbone of human MGL and its interaction with decomposed LPS ligands are thereafter presented in chapter VII.

Ultimately, large scale analyses of the unveiled interaction system are presented in chapter VIII. This molecular network considers the ECD of MGL and intact or fragmented LOSs scaffolds as ligands using STD-NMR spectroscopy, EM and BLI. Moreover, our vision was enlarged reaching the entire bacteria by collecting data from single-cell investigations on fluorescently labelled human MGL.

## ***Chapter VI.***

## **VI. Isolation and characterization of LPS ligands for ensuing interactions studies**

As stated earlier, the present PhD project is shared between two research groups from the university of Naples (Federico II) and the university of Grenoble (UGA, IBS institute). Both teams are interested in investigating bacterial pathogens recognition by innate immune system. We first directed our attention on the production and quality control of glycoconjugates isolated from different Gram-negative bacteria, wild type and mutants strains of *E. coli* (as said in chapter IV that those strains are predominant in clinical isolates) and *B. cenocepacia* J2315 strain, to use them later for assessing CLR/LPS interactions at different levels.

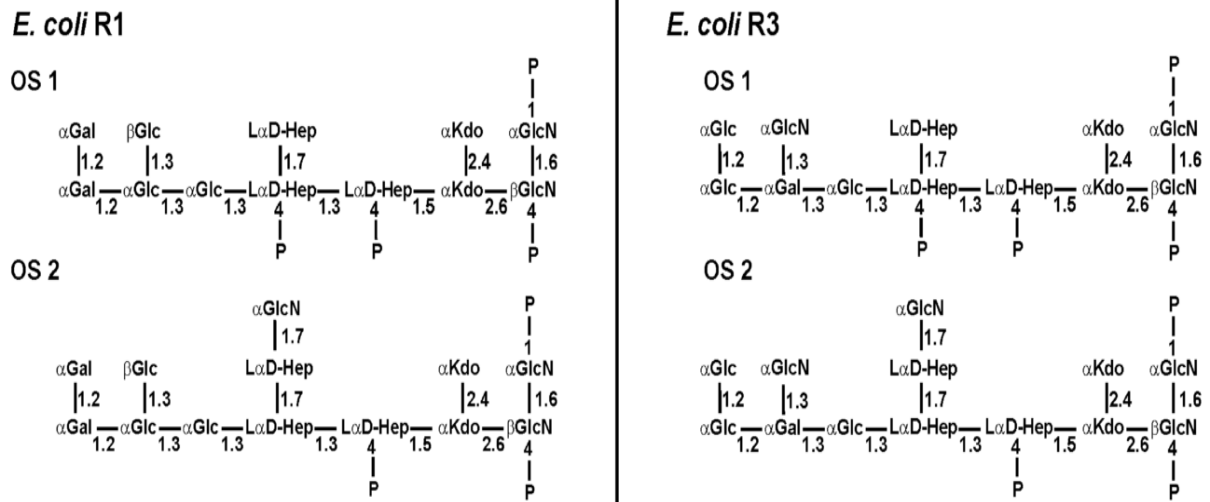
### **VI-1. Results and discussions**

#### **VI-1-1. Extraction and purification of LOSs and LPSs glycoconjugates**

##### **VI-1-1-1. Extraction and detection of LOSs from *E. coli* R1 and R3 strains and *B. cenocepacia* J2315**

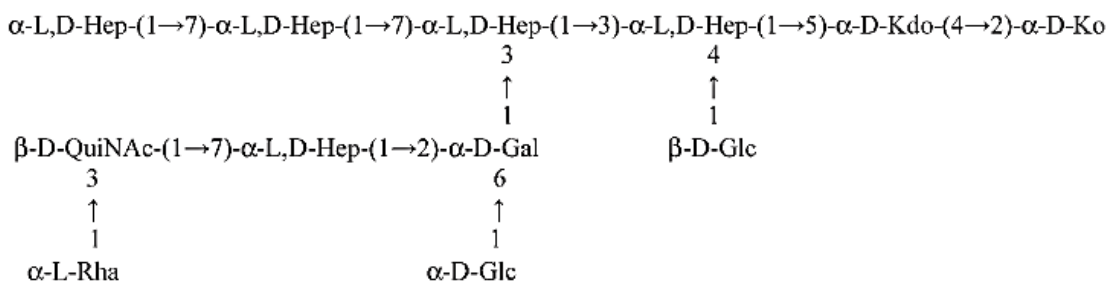
Among all the LPS glycoconjugates previously described we chose to start with the investigation of LOS produced by *E. coli* bacteria i.e. LOS from *E. coli* R1 and R3 strains. As presented in chapter IV-2, those strains are the most abundant in clinical isolates associated with human disorder such as urinary tract infections and gastro-intestinal diseases. *E. coli* strains with core OS of R1 and R3 have been detected at 7 and 48 % in human and animal clinical isolates<sup>328</sup>, respectively, whereas their counterparts (e.g., R2 and R4) were completely absent (0 %). Being the causative agents in gut microbiota related diseases, these strains are of high interest and, understanding if CLR can establish specific molecular interactions with LOSs would help in elucidating the role of CLR in driving immune responses. Importantly, having satisfying knowledges on the composition of these glycoconjugates following their isolation would permit mapping LOS residues that are directly or indirectly involved.

*E. coli* R1 and R3 strains were kindly provided by Muller Loennies's research group, which has already characterised the chemical structures of the core OSs produced by these two strains<sup>106</sup> (figure 59).



**Figure 59. Core oligosaccharide structures after alkaline degradation of *E. coli* LOS R1 and R3 with the corresponding minor and major forms.** Taken from<sup>108</sup>.

With the aim to investigate if the structural variation between R1 and R3 in the outer core region could influence their recognition by CLRs in response to *E. coli* pathogens, the production and isolation of LOS has been performed. We also focused our attention on the LOS produced by *B. cenocepacia* ET-12 type strain LMG 16656 (J2315), the major pro-inflammatory factor in cystic fibrosis patients. Its chemical structure and activity have been elucidated in our group suggesting a particularly rich sugar composition including characteristic residues such as D-quinovosamine QuiN (dideoxy-aminoglucose) and D-glycero-D-talo-oct-2-ulopyranosonic acid ko (Kdo derivative)<sup>259</sup> (figure 60).



**Figure 60. Chemical structure of the core oligosaccharide from *BC J2315* strain.** Taken from<sup>259</sup>.

Commonly, LOS structural variations through gene mutations are key events by which pathogens acquire resistance to antimicrobials by abolishing certain bindings. Therefore, prior knowledge of their structures is crucial for further investigation of their molecular interactions.

As a first step needed for LOS production, bacterial growth was carried out for *E. coli* R1 (F470, mutant derivative from *E. coli* O8:K27, R3 (F653, mutant from *E. coli* O157:H7) and *BC J2315* strains, followed by bacteria recovering by centrifugation. Bacteria growth yielded an averaged mass of 0.3 g of dried bacteria per liter of culture.

Dried cells were first washed with water, ethanol, and acetone to remove cell contaminants and later subjected to two extraction protocols. The isolation of LOS and LPS molecules, which exploit different amphiphilic characters, is achieved through two complementary extractions, that lead to their selective isolation. LPS has higher hydrophilicity, due to the presence of the polysaccharide moiety of the O-Antigen, whereas LOS, which lacks polysaccharide portion (core OS solely), has a higher hydrophobic character indeed. A high degree of purity was guaranteed by using the following selective extractions:

The method that was carried out in the case of **LOS** extraction from *E. coli* **R1 and R3**, consists of the use of Phenol/Chloroform/light petroleum (**PCP**) protocol as introduced in chapter III-1. The obtained yield from LOS R1 and R3 isolation equals 4% of LOS from dried biomass which was in good agreement with the literature<sup>336</sup>. Obtaining such good yields is not always insured since in case of new bacterial strain investigation for example, in which the partition behaviour of its LPS or LOS molecules is not known in advance, applying only PCP method would result in no or highly contaminated LPS fractions. This technique favored the immediate separation of LOS from LPS (if present), proteins and cell membrane components giving that lipophilic LOS is entirely soluble in monophasic PCP mixture whereas LPS and proteins are not<sup>235</sup>.

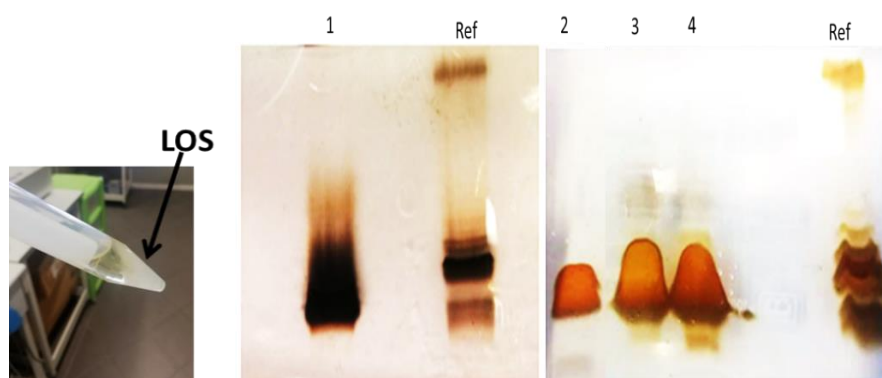
On the other hand, we produced isotopically labelled <sup>15</sup>N-<sup>13</sup>C-LOSs from *E. coli* R1 and R3 strains which were grown in M9 medium (see chapter XI for more details), to facilitate further NMR spectroscopy investigations. Labelled LOSs were extracted using the PCP method followed by four additional water washes. We assessed the purity of the extracted samples by recording <sup>1</sup>H-1D NMR spectra (data not shown) after each washing step and, the latest sample presented no signals from phenol traces. We managed to produce reasonable amounts of <sup>15</sup>N-<sup>13</sup>C-LOS R3 (0.8% from dried biomass) and, the quantity of the isolated <sup>15</sup>N-<sup>13</sup>C-LOS R1 was 10 folds higher.

LOS from *BC J2315* bacteria was isolated by using the **hot Phenol/Water** extraction protocol. The two phases were then extensively dialysed against water and lyophilised. The hot Phenol/Water method requires a further enzymatic treatment by using a typical reaction with nucleases (DNAse and RNAse) and proteinase k, which hydrolyse coextracted proteins and nucleic acids. This step



was followed by extensive dialysis in water which is necessary for removing the digested material. A Sephacryl S-300 column chromatography was performed afterwards to the LOS BC J2315 enzymatically treated sample to purify it from residual nucleases.

The presence of LOS was then assessed by sodium deoxycholate (**DOC-PAGE**) **electrophoresis** followed by silver nitrate staining (figure 61). The latter screening allowed us to define the type of the extracted material (LPS or LOS) and its purification degree. The principle of LPS molecules migration is the same as proteins migration in SDS polyacrylamide electrophoresis gel (SDS-PAGE) and the only difference between the two gels reside into higher resolution for DOC-PAGE over SDS-PAGE for LPS detection<sup>240</sup>.



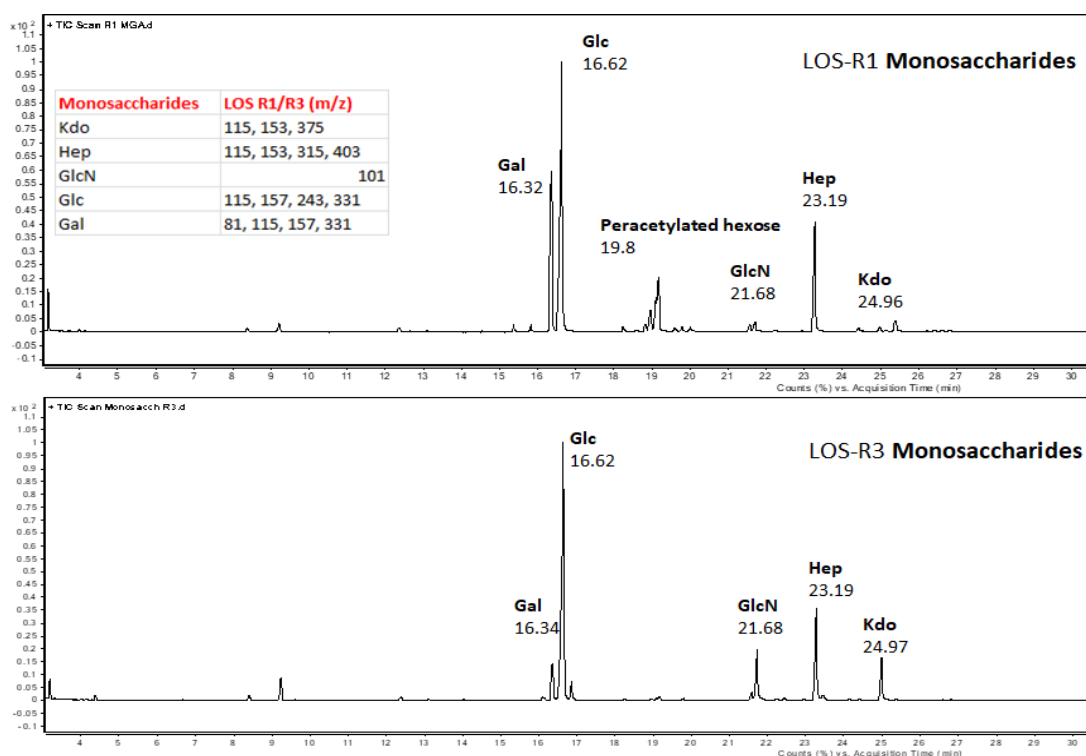
**Figure 61. Right panel: Different electrophoretic profiles of the extracted materials from *E. coli* and *B. cenocepacia* strains using the DOC-PAGE electrophoresis.** Left panel: PCP precipitate formation prior to LOS R1 and R3 extractions. Lanes **1** and **4** were loaded with LOS R1, **3** with LOS R3 and **2** with LOS from BC J2315 strain. Ref corresponds to the reference used as pure LPS isolated from *E. coli* O127:B8.

By comparing the electrophoretic profiles of the isolated LOSs (R1 and R3) with LPS from *E. coli* used as reference, the identification of LOS nature was established. LOS from R1 strain appeared as a medium-sized dark band at the bottom of the gel, denoting the low molecular weight of LOS lacking O-chain domain. Conversely, LPS reference showed a different profile of multiple fine bands on top of the gel, observed due to size heterogenicity of the O-Antigen, corresponding to the number of repeating units composing the O-polysaccharide moiety and a large band in the bottom corresponding to the core oligosaccharide (OS) and sugar residues from the Lipid-A portion.

## VI-1-1-2. Chemical and compositional analyses of the purified LOS fractions

### VI-1-1-2-1. GC-MS analyses of monosaccharides composition

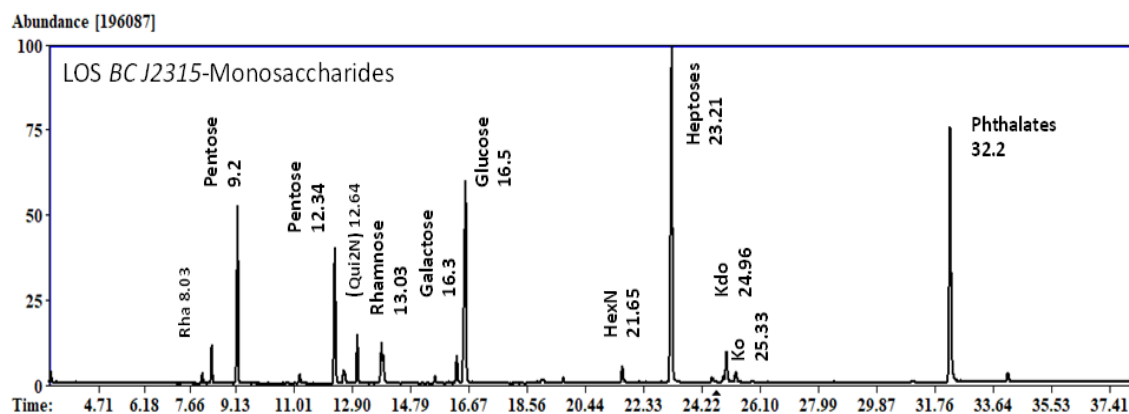
LOS fatty acids and saccharide composition were determined after appropriate derivatization of the isolated LOS samples into Acetylated Methyl Glycosides **AMG** (see chapter XI for experimental details). AMG method is useful for qualitative characterization of the primary structure of monosaccharides and fatty acids (FA) forming the LOS scaffold as introduced in chapter III-1. Once derivatized, AMG underwent GC-MS analysis by referencing to fragmentation patterns of opportune standards for comparison and the analyses retention times the sugar and fatty acids composition were derived using AMDIS analysis<sup>®</sup> software.



**Figure 62.** CG spectra of acetylated methyl glycosides from LOS R1 and R3. The reporter ions masses corresponding to distinctive sugar moieties are shown in the table.

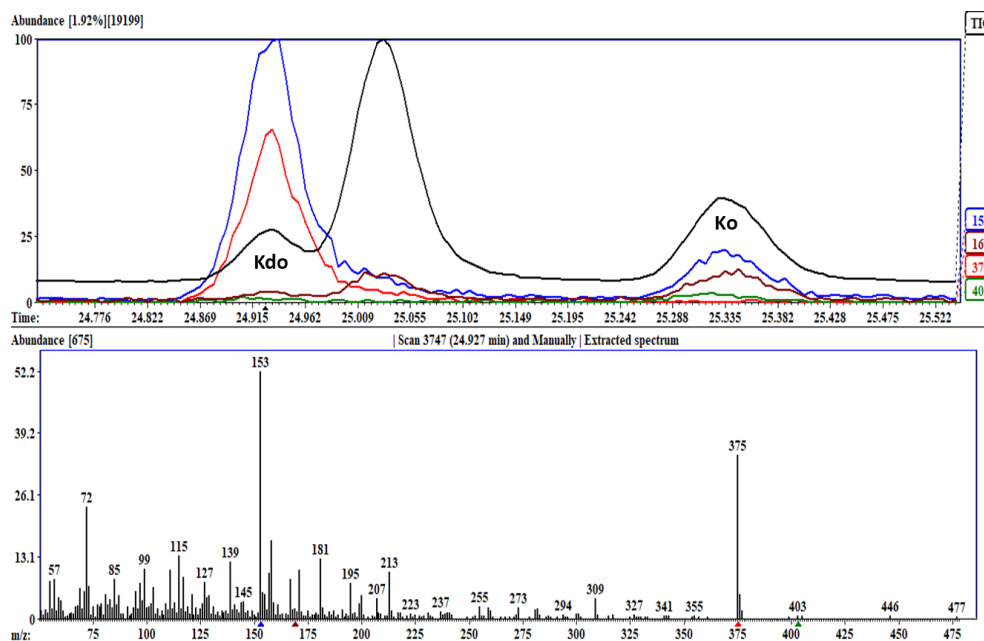
Considering monosaccharides analyses, LOSs from R1 and R3 showed remarkably similar chromatographic profiles regarding sugar types (figure 62). Retention times peaks corresponding to galactose (**Gal**), glucose (**Glc**) and amino glucose (**GlcN**) were observed. In addition to hexoses, acidic monosaccharides such as **Kdo** (3-deoxy-D-manno-oct-2-ulosonic acid) were detected (see annexes figure3). This characteristic sugar residue (and their derivatives) of the inner core of LPS molecules has been investigated as many attempts to analyze the presence of LPS by the use of GC-MS, mainly directed at detection of Kdo as a specific marker<sup>109,337</sup>.

As for the isolated LOS from *BC J2315*, compositional analysis revealed the presence of hexoses like glucose (**Glc**) and Galactose (**Gal**); important amounts of heptoses (**Hep**) -a dominant sugar residue among *BC* LPS core structures- and **pentoses** were detected besides, according to their retention times and fragmentation patterns (figure 63). An unusual sugar (**Qui2N**) (2-amino-2,6-dideoxyglucose) was also detected in the spectrum.



**Figure 63.** GC spectrum of acetylated methyl glycosides from the isolated LOS *BC J2315*.

In addition to all these residues, two other characteristic sugars were present in the LOS *BC J2315* extract, **Kdo** and **Ko** (glycero-d-talo-oct-2-ulopyranosonic acid) (figure 64).



**Figure 64.** GC and MS partial chromatograms of LOS *BC J2315* zoomed region of Kdo and Ko retention time peaks and the occurrences of characteristic fragmentation patterns of Kdo corresponding to 153 and 375 m/z reporter ions.

### VI-1-1-2-2. Fatty acids analyses of LOSs

Fatty acid methyl esters (from Lipid-A LOS) were extracted in the hexane phase during AMG and subsequently injected into GC-MS machine. In the case of LOS R1 and R3, saturated fatty acids as C14:0 and C12:0 and C-3 hydroxylated fatty acid C14:0 (3-OH) was also observed (table 6). These results are in complete agreement with the well-known dominant fatty acid species in the LPS in *Enterobacteriaceae*<sup>338–340</sup>.

**Table 6. Fatty acids components of the isolated LOS from R1 and R3 *E. coli* strains** showing the most relevant fatty acid species (bold) of *E. coli* strain and their corresponding retention times and molecular masses of the fragmented molecules obtained from fatty acid methyl esters.

<b>Fatty acids (R1/R3)</b>	<b>Rt (min) R1</b>	<b>Rt (min) R3</b>	<b>Ions (m/z)</b>
<b>C12:0</b>	6.11	6.13	214,171,143,87,74
<b>C14:0</b>	8.56	8.53	242,199,143,74
<b>C14-3OH</b>	10.19	10.18	258,166,103,83,69
C16:0	10.74	10.74	270,227,143, 87, 74
C18:0	12.69	12.71	298,255,199,87,74

LOS from *BC J2315* contained saturated fatty acid species such as C14:0 and C16:0; and C-3 hydroxylated fatty acids C14:0(3-OH) and C16:0(3-OH) (Table 7).

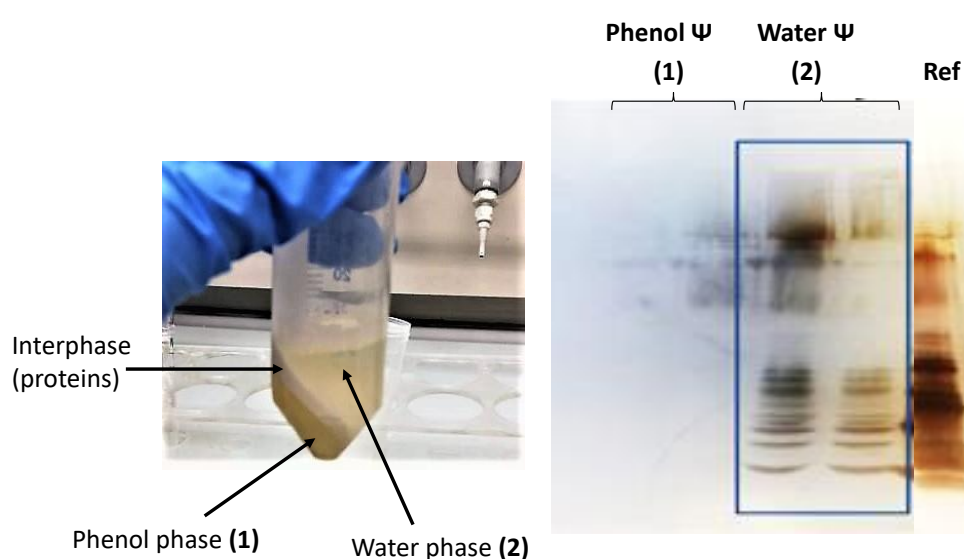
**Table 7. Fatty acids components of the isolated LOS from BC J2315** showing the most relevant fatty acid species (bold) of BC J2315 strain and their corresponding retention times and molecular masses of the fragmented molecules obtained from fatty acid methyl esters.

<b>Fatty acids (LOS J2315)</b>	<b>Rt (min)</b>	<b>Ions (m/z)</b>
C18:0	12.70	298,87, 74
<b>C16:0</b>	10.74	270, 87, 74
<b>C14:0</b>	8.56	242, 87, 74
<b>C14-3OH</b>	10.18	166,103,74
<b>C16-3OH</b>	12.25	194,103,74

### VI-1-1-3. Production of <sup>13</sup>C-LPS and detection of isotopically labelled <sup>13</sup>C LPS from *E. coli* 157:H7

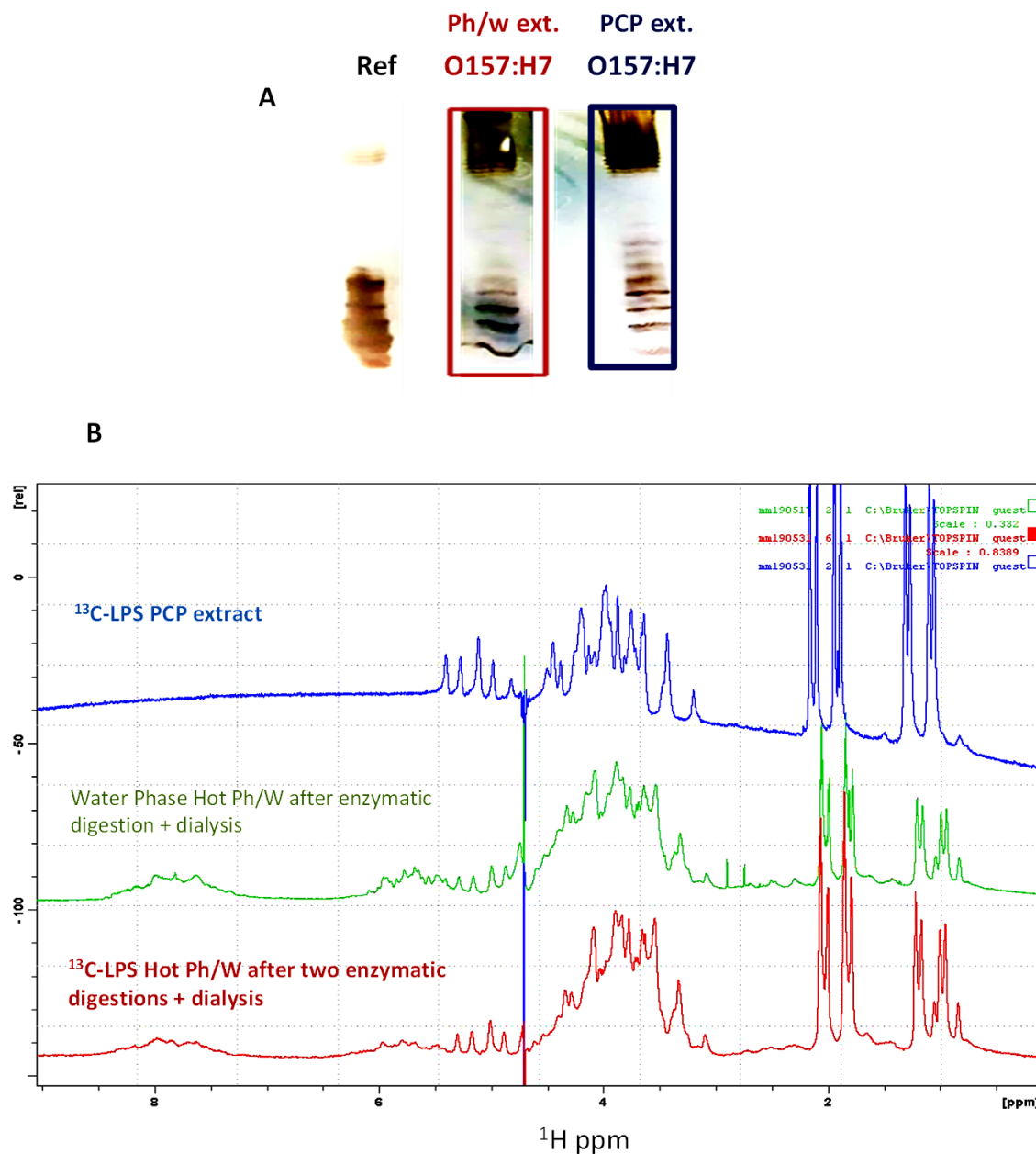
As presented in chapter IV-2, *E. coli* O157:H7 strain expressing LPS with an O-Antigen O157 was detected in human clinical isolates suffering from enteric diseases and was identified as a Shiga-toxigenic pathogen (i.e., verotoxigenic *E. coli*). As a reminder, one of our interests revolves around understanding molecular recognition mechanisms by which human lectins like MGL, for instance, might interact with such highly pathogenic strains through binding to LPS.

To increase the abundance of  $^{13}\text{C}$  isotopes in glycoconjugates for further NMR spectroscopic investigations, we sought to produce isotopically labelled  $^{13}\text{C}$ -LPS. *E. coli* O157:H7 bacteria producing  $^{13}\text{C}$  labelled LPS was kindly provided by Prof. A. Polissi from the department of Pharmacological and Biomolecular Sciences at Università Degli studi di Milano. We then extracted  $^{13}\text{C}$ -LPS fractions by using the PCP method followed by Hot-Phenol/Water extraction. Thus, the isolated material can be detected either in the water phase or in the phenol extract. In our case, the extracted LPS was found in the water phase (figure 65). The PCP extraction yielded 1.5 % of  $^{13}\text{C}$ -LPS from dried biomass. As for Hot-Phenol/water LPS extract, subsequent enzymatic digestion and dialysis were performed and yielded 11% of extracted material from dried biomass.



**Figure 65. Phenol and aqueous phases from PCP followed by hot-phenol water extraction of LPS O157:H7 and their corresponding gel profiles.** LPS from *E. coli* O127:B8 was used as reference.  $^{13}\text{C}$ -LPS samples, extracted from water and phenol phases, were subjected to  $^1\text{H}$ -1D NMR spectroscopy to analyse their purity levels. The  $^1\text{H}$ -NMR spectra showed some differences in terms of purity and LPS molecules amounts. In fact, the spectroscopic profiles of the three samples that are  $^{13}\text{C}$ -LPS extracts from PCP, PCP+Hot Phenol/Water and either one or two enzymatic digestions and dialysis were compared (figure 66). Amide protons broad signals specific for proteins (around 6 and 8 ppm) were present in the Ph/Water fractions (green spectrum) with less amounts for the sample that was enzymatically treated twice (red spectrum). Those signals were completely absent in the case of the PCP  $^{13}\text{C}$ - LPS extract (blue spectrum) which showed, in addition, sharp signals corresponding to the anomeric protons mostly from the O-Antigen portion (four peaks around 4-5 ppm). Sugars rings protons signals (3-4.5 ppm) and

intense peaks of methyl groups from fatty acids of Lipid-A were also present. Thus, the purity of PCP  $^{13}\text{C}$ -LPS sample was confirmed and further NMR analyses were established on that sample.



**Figure 66. Examination of  $^{13}\text{C}$  LPS O157:H7 purity through (A) DOC-PAGE analyses and (B) 600 MHz  $^1\text{H}$ -NMR.** A) PCP and PCP/Hot-phenol water extracted LPS gel profiles, 1 mg/mL of the isolated  $^{13}\text{C}$ -LPS were solubilised in  $\text{H}_2\text{O}$  for electrophoresis (10  $\mu\text{g}/\text{lane}$ ). B) NMR spectra recorded at 298K of the isolated LPS samples prepared with different protocols and containing 5 mg of  $^{13}\text{C}$ -LPS extracts suspended in  $\text{D}_2\text{O}$ .

The isolated LOS and LPS molecules were further subjected to different methods to analyse and confirm their chemical structures, without going into detailed analyses since their chemical structures have been already determined. In the case of LPS, NMR spectroscopy experiments

could be performed on that native sample due to its high solubility. As for the LOSs, a delipidation step was required to allow their structural elucidation in solution through NMR spectroscopy.

#### **VI-1-1-4. LOSs delipidation**

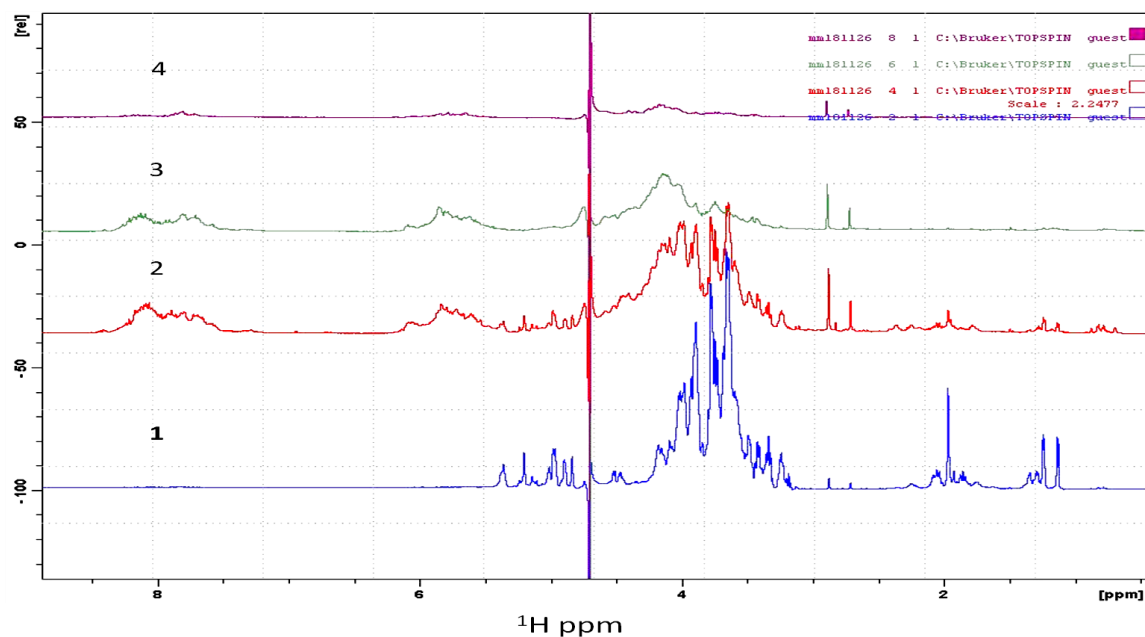
LOS molecules have the tendency to form vesicles in solution with low solubility in both aqueous and apolar organic solvents due to their amphiphilic nature. This self-aggregation phenomenon of *E. coli* LPS was investigated by Santos and coworkers by using light scattering spectroscopy and the apparent critical micelle concentration (CMC) was determined<sup>341</sup>. Above 14  $\mu\text{g mL}^{-1}$ , supramolecular structures of LPS molecules would form different shapes and sizes in solution. Consequently, structural analyses of LOS/LPS molecules becomes a difficult task, however, there are several solutions to overcome this obstacle. For example, chemical treatments are required for these molecules to improve their solubilisation in adequate buffers. Two chemical reactions commonly used in our laboratory for LPS/LOS analyses, consist in **LOS chemical degradation** to separate saccharide moiety (i.e., core OS) from fatty acids (c.f. Lipid-A) thus providing soluble fractions of LPS/LOS glycoconjugates and, to facilitate an eventual analysis of the lipid portion.

The first chemical treatment used for LOS R1 and R3 de-acylation consists of the removal of fatty acids from the Lipid-A linked to the OS, by means of alkaline treatment as introduced in chapter III-1. Briefly, de-*O*-acylation was established with mild hydrazinolysis in anhydrous conditions with anhydrous methyl-hydrazine (1 ml for 20 mg of sample), to remove ester-linked fatty acids. A strong alkaline hydrolysis namely de-*N*-acylation was subsequently performed to obtain the cleavage of amide-linked acyl chains. Parenthetically, amide-linked substituents like N-Acetyl groups linked to sugar residues, and perhaps base-labile substituents like pyrophosphate groups, could be lost during this procedure because of harsh reaction conditions (4 M KOH, at 120°C). Fatty acids from *E. coli* LOSs were immediately separated from the OSs by extracting them three times in acetone and in chloroform phases right after de-*O*-acylation and de-*N*-acylation, respectively.

The second chemical approach used herein consists of a mild hydrolysis for Lipid-A removal from LOS BC J2315. An acetic acid (at 1%) treatment of the LOS was sufficient to selectively cleave the glycosidic linkage, which is alkali-stable and acid-labile, between the Kdo residue of the core and the glucosamines of the lipid A. The oligosaccharide moiety was released in the water phase. As

for the released lipids, they were readily collected as brownish precipitates after delipidation, by centrifugation.

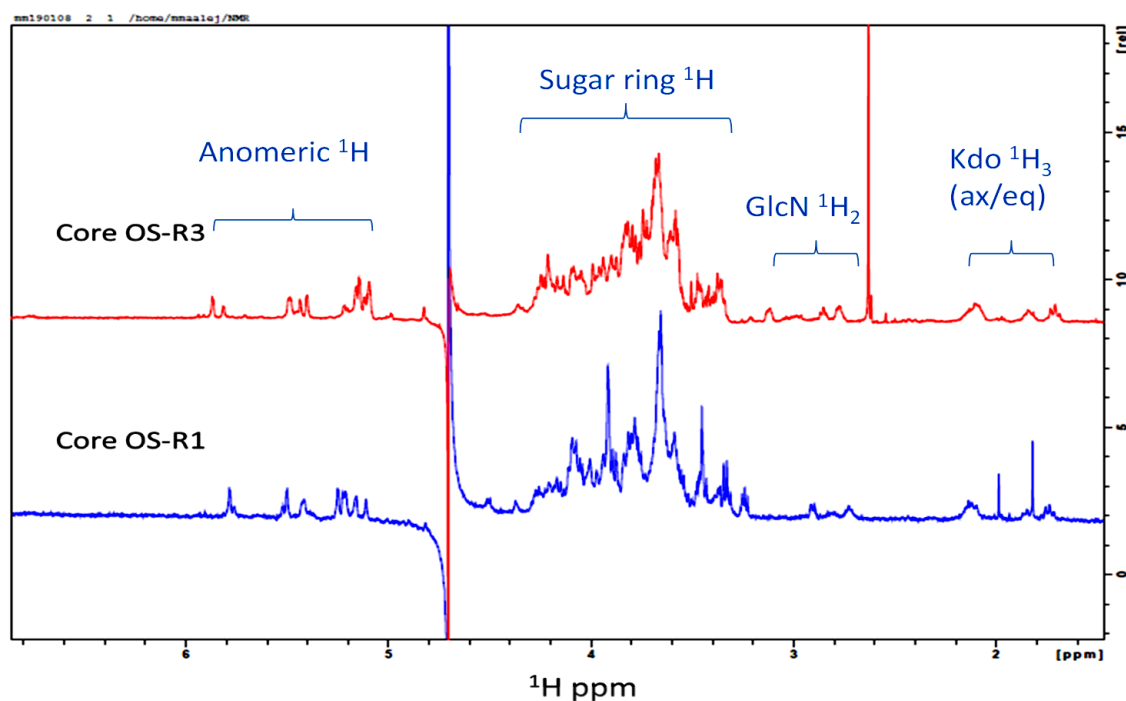
Gel filtration chromatography was then used to remove the salts formed during the de-acylation reactions. Briefly, samples were loaded in a Sephadex G10 column (for OS R1 and R3), followed by size exclusion chromatography on Biogel P2 and P4 (for OS J2315) and, eluted fractions were detected, collected, and lyophilized. P4 polyacrylamide bead was first used as chromatographic column for size-separation of the oligosaccharide OS BC J2315 released after chemical and enzymatic treatments and lyophilized after dialysis. Each of the collected fractions were analyzed by means of proton  $^1\text{H}$ -1D-NMR experiments to check the presence and the purity of the samples (see annexes figure2). Fractions showing the presence of carbohydrates were pooled and subjected to another gel filtration chromatography purification step using Bio-Gel P2 column (figure 67). The generated chromatographic profile corresponding to fraction **1**, showed a pure oligosaccharide with no signals from any protein contaminants, however very weak resonances from methyl groups (around 1.2 ppm), potentially from remaining lipids, were observed, which might indicate that the delipidation was incomplete. The purification of the deacetylated LOS extracted from BC J2315 strain was demanding and required the use of dialysis followed by the two subsequent gel filtration chromatography to end up with few milligrams (fraction **1**= 2.9 mg) of pure OS BC J2315.



**Figure 67. 600 MHz  $^1\text{H}$ -NMR spectra of 2.7 mM final concentration of OS J2315 fractions in  $\text{D}_2\text{O}$  at 298K after P4 then P2 columns chromatography.** The four fractions resulted from the second chromatography showed different NMR profiles highlighted by the presence of well-defined peaks in the anomeric region (4.6-5.5 ppm) of fraction 1 with respect to fractions 2-4.



**OS R1** and **OS R3** fractions were analyzed by means of  $^1\text{H}$ -1D NMR experiment and confirmed again the presence of sugar residues. In fact, by recording such NMR spectra, we obtained a global view on OSs sugar types and numbers. Hence, depending on three spectral regions, distinctive peaks can be assigned to the anomeric protons signals (between 5.8 and 4.6 ppm), the ring proton signals (overlapped signals between 4.6 and 3 ppm) and the deoxy positions signals (i.e. for kdo and acetylated residues) between 2.5 and 1.0 ppm (figure 68).



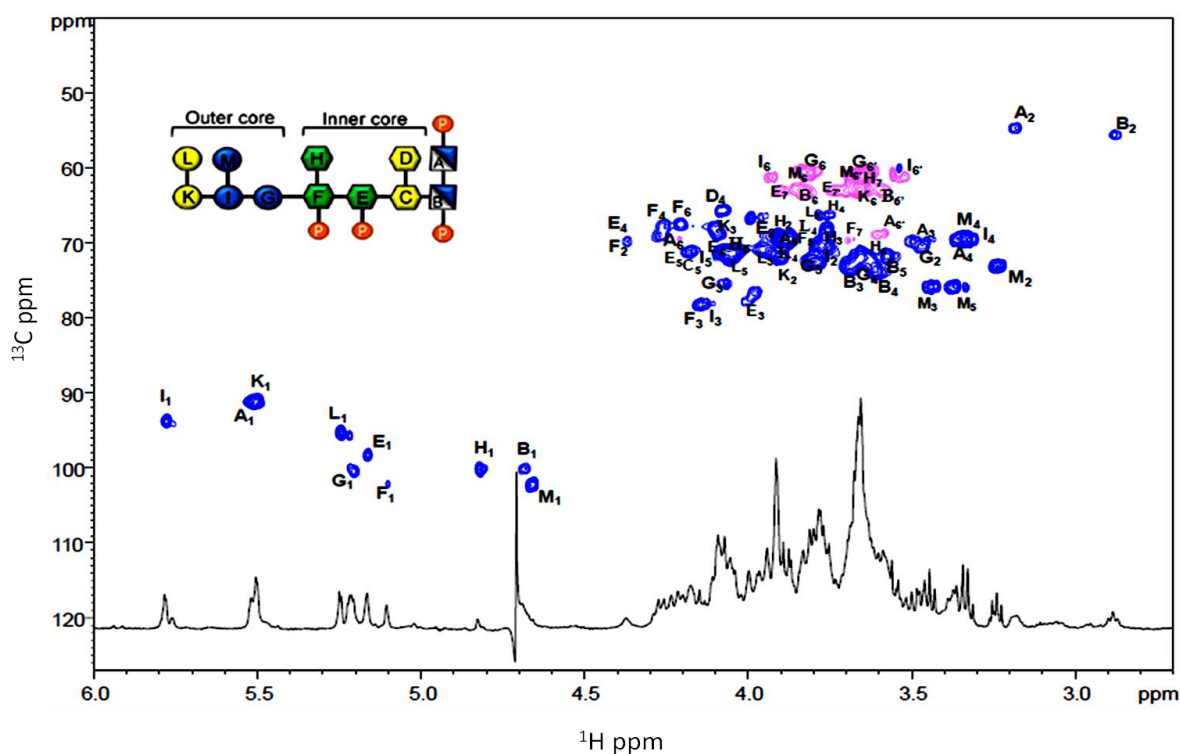
**Figure 68.** 600-MHz- $^1\text{H}$  NMR spectrum of 2.3 mM final concentration of deacylated LOS from *E. coli* R1 and R3. NMR samples were prepared in 25mM Tris- $\text{d}_{11}$ , 4mM  $\text{CaCl}_2$ , 150 mM NaCl pH 7 in  $\text{D}_2\text{O}$  (buffer A) and experiments were run at 298K. In the high field region of the spectrum 2 pairs of signals were observed originating from 3-deoxy protons with chemical shifts characteristic of Kdo residues. Signals of anomeric protons from the sugar sequence of the core oligosaccharide were also present at the region between 6.0-4.5ppm together with their ring protons between 3.0-4.3 ppm.

## V-1-2. Liquid and solid-state NMR for identification of LPS/LOS residues

### V-1-2-1. OS R1 characterization by liquid NMR spectroscopy

OSs isolated from *E. coli* R1 was analyzed by state-of-art NMR experiments that allow the determination of their structural features like the anomeric configuration of the linkages and the monosaccharides types and sequences. A combination of homo- and hetero-nuclear 2D-NMR experiments such as DQF-COSY and TOCSY (annex section figure4), ROESY, NOESY,  $^1\text{H}$ - $^{13}\text{C}$  HSQC and  $^1\text{H}$ - $^{13}\text{C}$  HMBC were all considered for the assignment of **OS R1** in order to attribute all the

spin systems by comparing to the already published *E. coli* OSs NMR structures including OS R1<sup>342,343</sup>, thus characterizing the monosaccharide sequence and, if applicable, determining the location and the nature of non-carbohydrate substituents.



**Figure 69. Zoom in the  $[^1\text{H}-^{13}\text{C}]$ -2D-HSQC spectrum of 5 mg of the isolated deacylated LOS R1.**

The NMR sample was prepared in buffer A and experiments were run at 298 K. The most relevant heteronuclear correlations are reported. Correlations corresponding to  $\text{CH}_2$  groups are colored in pink whereas  $\text{CH}$  groups are in blue. OS R1 structure is highlighted in the figure.

Briefly, in the anomeric region of the  $^1\text{H}$  NMR and  $^1\text{H}-^{13}\text{C}$ -2D-HSQC spectra (figure 69), ten anomeric signals were identified as composing the major glycoform of the oligosaccharide **OS**, excluding Kdo residues. The crowded region between 3.2 and 4.2 ppm includes sugar ring protons from the 12 sugars constituting OS R1 backbone. Up field cross-peaks (~50 ppm) were attributed to amino sugars A and B. Thus, NMR spectroscopic data allowed us to assign  $[^1\text{H}-^{13}\text{C}]$  resonances (Annex table1) belonging to the main glycoform of the isolated and deacylated LOS of the *E. coli* R1 strain. This finding was in good agreement with previously published structural data of OS R1<sup>342</sup>. The assignment of OS R1 is minutely described in our published article (Supplementary info. of <sup>344</sup>).

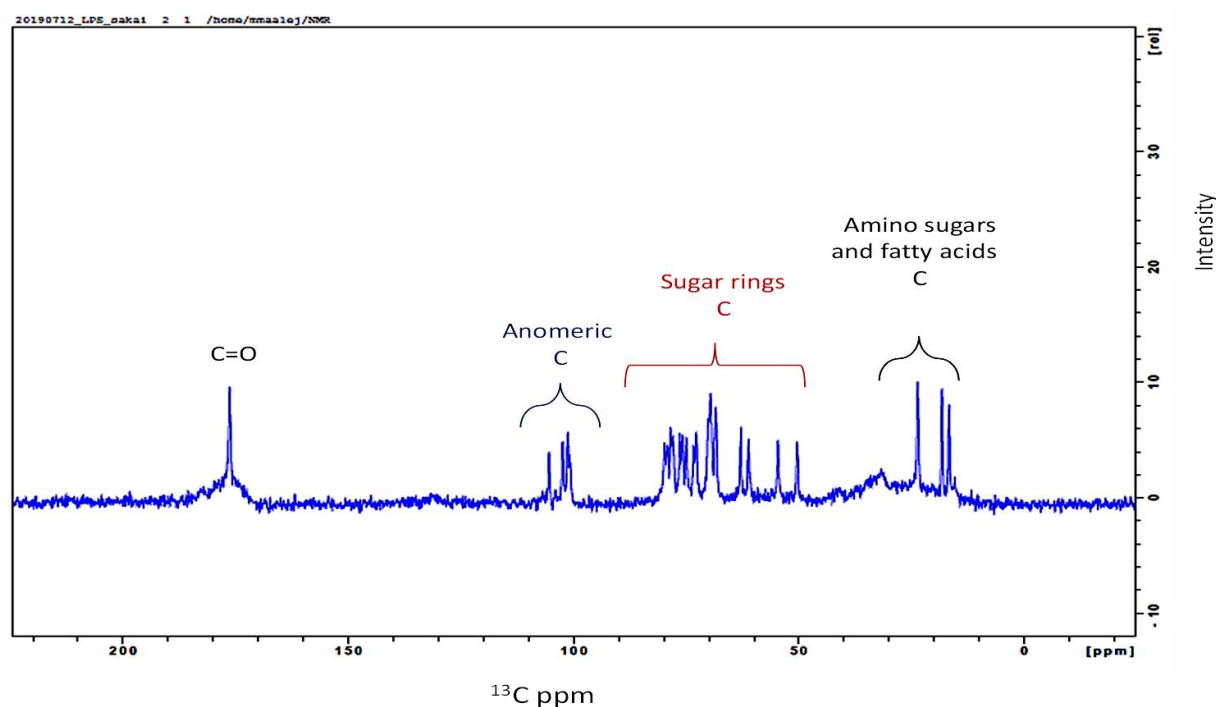
### VI-1-2-3. Structural analyses of $^{13}\text{C}$ -LPS O157:H7, the O-Antigen

In the previous part of this chapter, chemical analyses of OS R1 were presented. We mainly focused on quality control of the isolated *E. coli* glycoconjugates to facilitate further interaction

studies by NMR spectroscopy. Obtaining isotopically labelled sample is important for the establishment of NMR based analyses regarding glycoconjugates structures and functions.

As previously mentioned, we succeeded to produce  $^{15}\text{N}$ - $^{13}\text{C}$ -glycoconjugates (LOSs R1, R3 and LPS O157:H7). We first evaluated the quality of the isolated  $^{13}\text{C}$ -LPS O157:H7 by **solid state NMR spectroscopy (ss-NMR)** since we wished to use that approach for further studies considering C-type lectins interaction with native LOS/LPS glycoconjugates (in membrane-like structures). Sample preparation and packing into rotor with 1.3 mm outer diameter, is needed before starting ss-NMR analyses, contrarily to liquid state NMR that requires only liquid sample transfer into the NMR tube. Sample packing requires the use of ultra-centrifuge at high-speed rotation to allow its sedimentation inside the ss-NMR rotor (see chapter XI for more details).

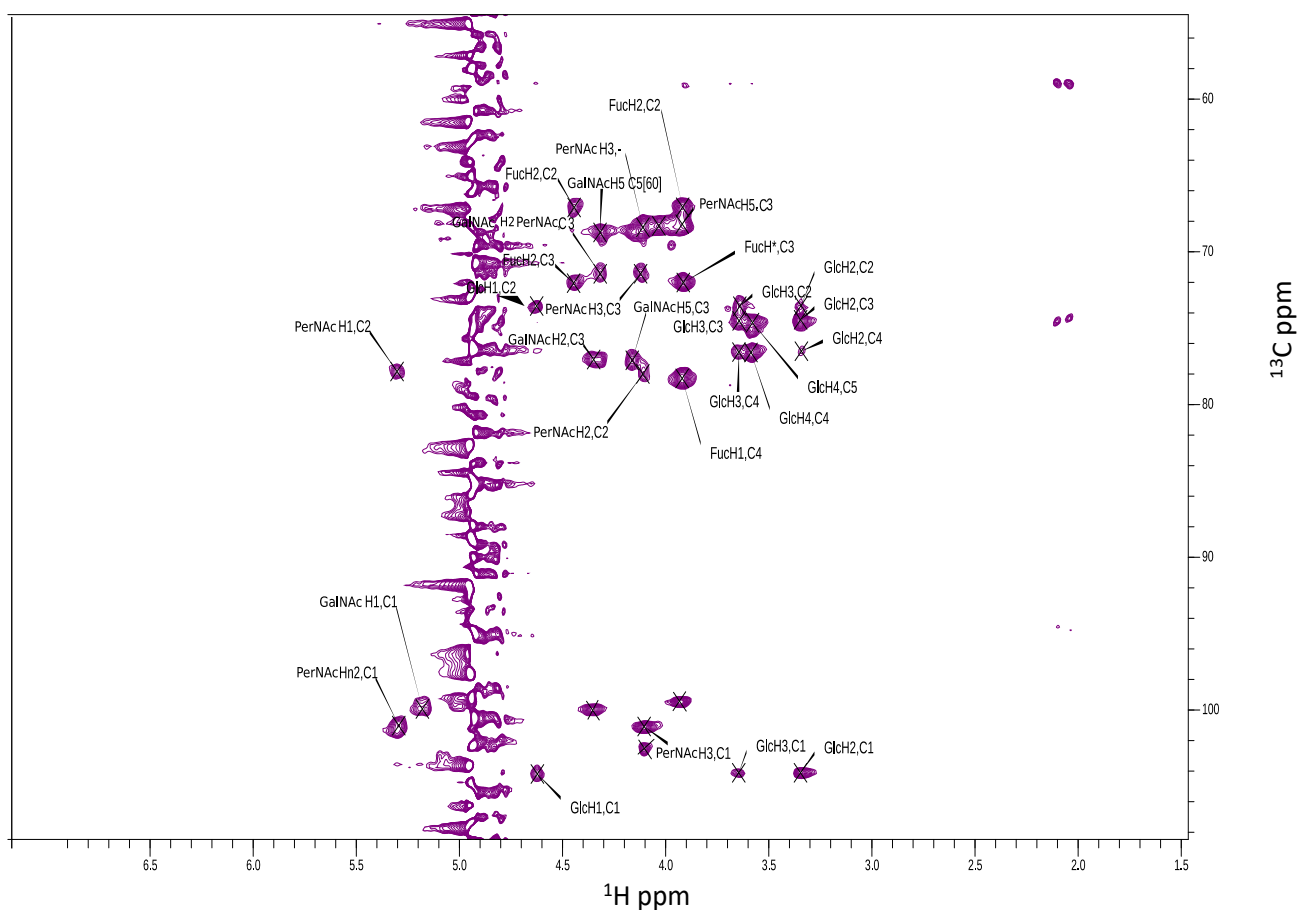
As said several times along this manuscript, insoluble LPS and especially LOS biomolecules form membrane-like assemblies in solution and give rise to broad signals when investigated by liquid state NMR. By using ss-NMR, we could obtain sharp signals thanks to sample spinning at magic angle spinning (MAS) thus mimicking its molecular motions in solution. Consequently, the results obtained from that strategy become more similar to liquid state NMR parameters with good quality.



**Figure 70.** 950 MHz MAS  $^{13}\text{C}$ -CP-1D-NMR spectrum of 2 mg of  $^{13}\text{C}$  LPS from *E. coli* O157:H7 strain. The sample was prepared in 25 mM Tris 150 mM NaCl 4 mM  $\text{CaCl}_2$   $\text{H}_2\text{O}$  (buffer B) and ss-NMR experiments were recorded at 270K and 25KHz.

As initial NMR data, a monodimensional  $^{13}\text{C}$ -CP-MAS NMR spectrum was recorded at a spinning frequency of 25KHz (figure 70), showing many sharp signals corresponding to the different chemical groups within the LPS O157:H7 biomolecule. Notably, the recorded  $^{13}\text{C}$ -NMR spectrum included **four** signals of **anomeric carbons** (99.4, 100, 101, 104 ppm) corresponding to the sugar residues constituting the O-Antigen. Besides that, two resonances corresponding to C-N signals from **PerNAc** and **GalNAc** residues were observed at  $\sim 49$  and 53 ppm and two other carbon signals of deoxy-C6 residues (**Fucose** and PerNAc sugars) appeared at 15.3 and 17 ppm, respectively. Fatty acids signals from Lipid-A portion were also visible in the recorded  $^{13}\text{C}$ -1D-NMR spectrum (especially from carbonyl resonances around 175 ppm, figure 70) however it was not possible to assign (Lipid-A carbons from C2 to C6, around 10-45 ppm) because of ambiguous data either from large cross peaks or missing spins correlations in the 2D-NMR spectra. As for the core structure, residues resonances were absent due to lack of sensitivity or to heterogenous structure owing to the huge/dominant signals from flexible O-Antigen portion at the spectral region of anomeric  $^{13}\text{C}$  resonances around 100 ppm.

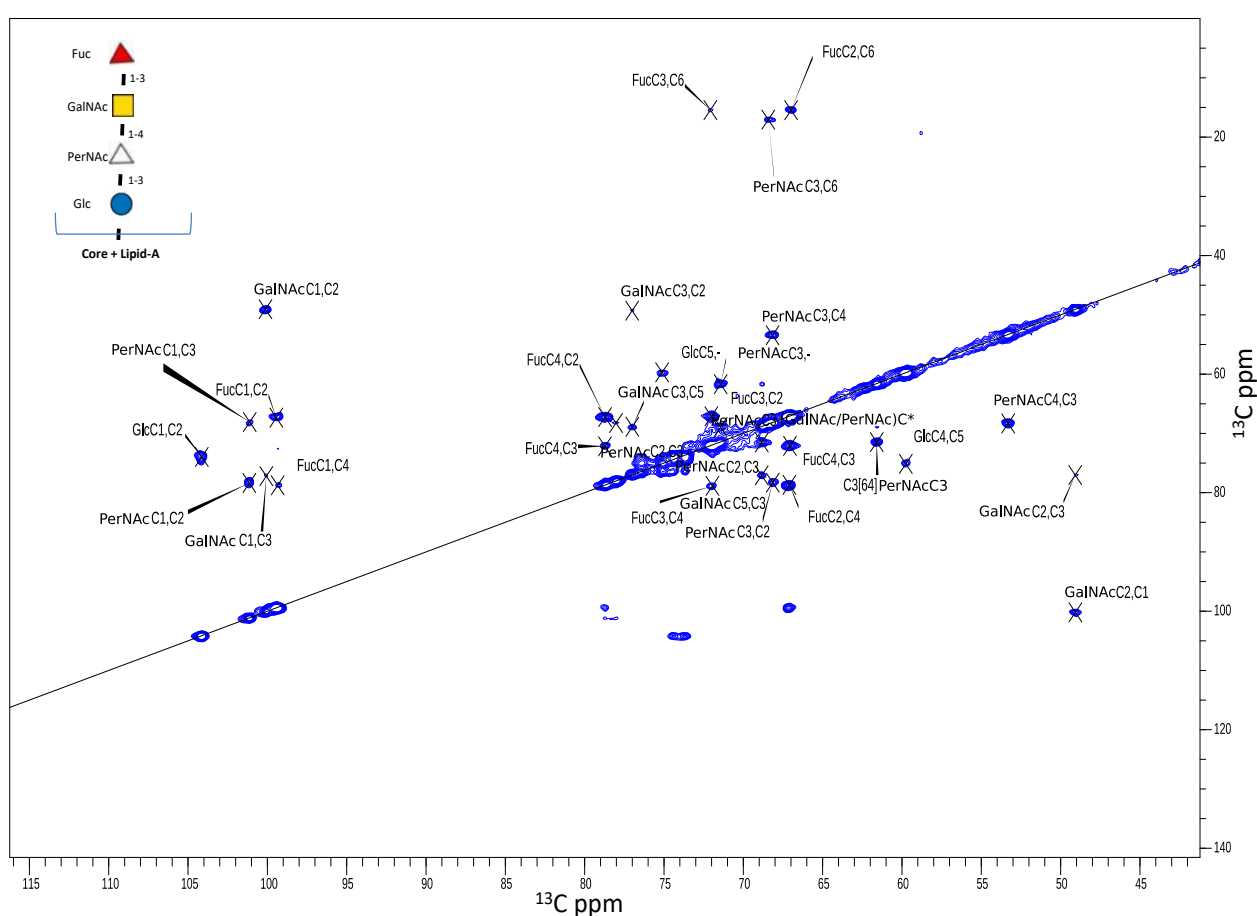
We managed to obtain sharp NMR signals with good quality, and we recorded multidimensional spectra afterwards to distinguish those signals and avoid their overlapping, thus facilitating O-Antigen assignments. The 2D experiment hCCH-2D-INEPT (figure 71) was recorded at high-speed rotation (MAS) of 50 KHz to improve  $^1\text{H}$  resonances resolution (providing sharp signals). This is a  $^1\text{H}$  (proton) detection experiment that permits to transfer magnetisation between adjacent carbons via  $^{13}\text{C}$ - $^{13}\text{C}$  scalar coupling, thus giving an information about the correlations within a sugar ring.



**Figure 71. 950MHz MAS [ $^1\text{H}$ - $^{13}\text{C}$ ]-hCCH-2D-INEPT spectrum (zoomed) of 2 mg of  $^{13}\text{C}$  LPS from *E. coli* O157:H7 strain.** The sample was prepared in buffer B and ss-NMR experiments were registered at 270K and 50KHz.

The signals provided from ss-NMR data mostly corresponded to the O-Antigen portion of LPS O157:H7, this is not surprising knowing that it is the most abundant sugar rich part of LPS molecules in general. O-Antigen residues were assigned mainly according to their chemical shift values and bidimensional correlations thanks to Ccpnmr.v2 software<sup>345</sup>. Ss-NMR data are reported in supplementary tables 2, 3 and 4 in Annexes part. Our findings were consistent with those obtained from liquid state NMR analyses by Nishiuchi and co-workers who studied the detailed structure of LPS *E. coli* O157:H7 O-Antigen through compositional analyses and in solution NMR investigations<sup>332</sup>. Briefly,  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts signals from both mono- and bi-dimensional spectra were in line with the published data despite the different experimental procedures that were employed. In fact, the authors have derived the O-Antigen from LPS O157:H7 (by removal of Lipid-A) before analyzing its structure by NMR whereas our O-Antigen sample was analyzed in the native state (contained in the intact LPS). Indeed, their data were considered as references for our LPS O157:H7 O-Antigen assignments.

In order to decipher  $^{13}\text{C}$ - $^{13}\text{C}$  correlations from rigid parts of the isolated LPS molecule,  $[^{13}\text{C}$ - $^{13}\text{C}]$ -hCC-DARR spectrum (figure 72) was recorded as well, at 15 kHz of MAS, and then analysed in parallel with the other spectra. It should be noted that O-Antigen residues were observed in both recorded ss-NMR spectra (hCCH-INEPT and hCC-DARR), indicating that LPS O157:H7 O-Antigen part contains both flexible and rigid oligosaccharides branches. Briefly, the four anomeric signals and ring sugar signals were detected in the  $[^1\text{H}$ - $^{13}\text{C}]$ -hCH-2D-INEPT spectrum (Annexes figure 5) and, analyses back and forth the recorded 2D spectra including  $[^1\text{H}$ - $^{13}\text{C}]$ -hCCH-INEPT and  $[^{13}\text{C}$ - $^{13}\text{C}]$ -hCC-DARR allowed us to assign most of the sugar  $[^1\text{H}$ - $^{13}\text{C}]$  and  $[^{13}\text{C}$ - $^{13}\text{C}]$  correlations, thus identifying O-antigen residues from the isolated LPS O157:H7.

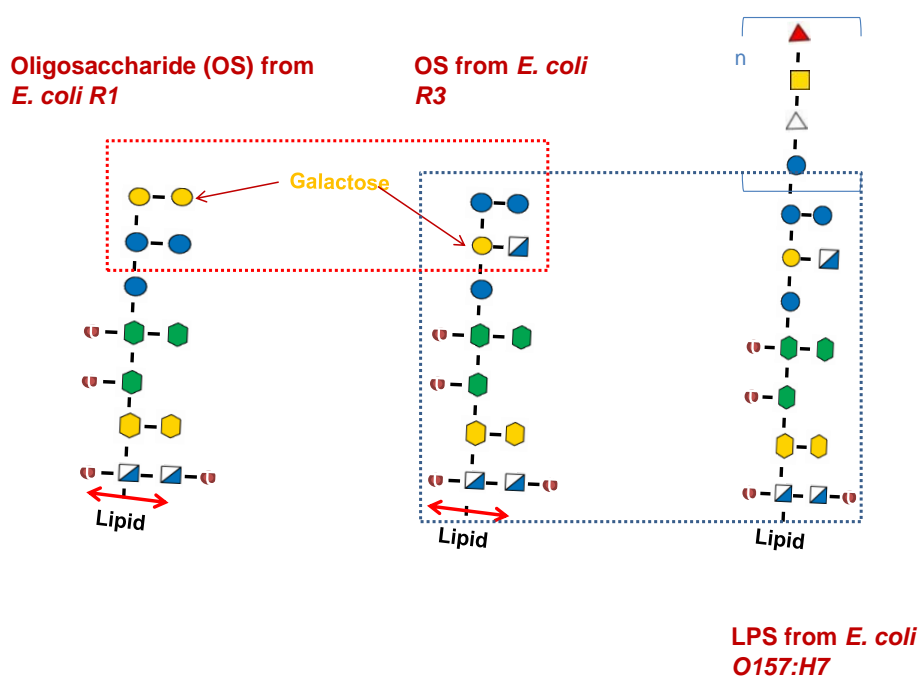


**Figure 72. 950MHz MAS  $[^{13}\text{C}$ - $^{13}\text{C}]$ -hCC-DARR-NMR spectrum focused on sugars resonances of 2 mg of  $^{13}\text{C}$  LPS from *E. coli* O157:H7 strain.** The sample was prepared in buffer B and ss-NMR experiments were recorded at 293-298K and 15KHz. For simplicity, the O-Antigen structure is shown using symbol nomenclatures from<sup>346</sup> and  $^{13}\text{C}$ - $^{13}\text{C}$  correlations from lipids are not shown in this spectrum.

## VI-2. Conclusion

A first point to be discussed in this part is that LOS R1 and LOS R3 were produced and purified in their native forms, which means that major and minor forms of each LOS were not separated. In other words, the mixture of two forms of LOS will not hinder in the study of their

recognition by human lectins, knowing that the recognition between CRDs and glycolipids or glycoproteins for instance occurs typically at terminal non-reducing residues<sup>347</sup> which are part of the outer core. The latter is identical in major and minor forms. Thus, by considering only major structures of the isolated LOSs (figure 73), the ensuing findings would be factual, in any case, because they may help in deciphering LOSs recognition by human CLRs in native conditions (as they are produced in the cell surface). Naturally, the present work may additionally help in considering further studies on individual LOS forms (major or minor) which may provide in-depth insights about the recognition of these pathogens by the immune system.



**Figure 73. Sugar sequences within the isolated *E. coli* LPS scaffolds for further interaction studies with human lectins.** Symbol nomenclatures of sugar residues has been used from<sup>346</sup>. Red and blue squares refer to structural differences and similarities within the isolated LPS species, respectively.

Structural diversity among the three *E. coli* glycoconjugates is a challenging feature. This includes not only intact (LPS) or truncated (LOS, OS) versions but also fine differences in the outer core structures. Those differences in the outermost carbohydrate moieties originate from gene mutations and they could be decisive for glycoconjugates function as PAMPs.

Here, we succeeded to extract and purify satisfactory amounts of *E. coli* glycoconjugates (in both isotopically labelled and unlabelled forms) which were chemically analysed via several methods to confirm their structures in terms of sugars composition and residues types. Parenthetically, as said all along this manuscript that these samples are experimentally difficult to explore because

of their low solubility and heterogenous compositions. Thus, the choice of each scientific approach that has been used in this work, considers the degree of solubility of the studied biomolecules. The following chapters are meant to present the results obtained using several methods that we applied on labelled (and unlabelled) species of the isolated glycoconjugate counting  $^{13}\text{C}$ -LPS, LOSs and OSs, so that we could investigate their interaction with human lectins at different levels.



## ***Chapter VII.***

## VII. Selective screening of an interactive LPS-Lectin system

This chapter is dedicated to present the preliminary data that we obtained at the beginning of the thesis as primary interaction tests. The idea behind these primary tests is about rapidly screening human C-type lectins (CLRs) to gain time in studying LPS-lectin interaction systems. In fact, we had a list of available lectins and LPS/LOSs to study within this PhD project and we had to select an active/functioning interaction system to start with our molecular investigations.

### VII-1. Results and discussions

#### VII-1-1. Preliminary results of the interaction between intact LPS and the human lectins

As presented in chapter VI, we successfully isolated three glycoconjugates from various *E. coli* strains with important clinical relevance (R1, R3 and O157:H7), in both intact (LOS/LPS) and deacetylated versions (oligosaccharide, OS) and, glycoconjugates from *B. cenocepacia* J2315 as well. Initially, we considered only *E. coli* glycoconjugates as first ligands to be tested and we scheduled an eventual study for *BC J2315*. In addition, *E. coli* ligands were expressed and purified as isotopically labelled biomolecules required for NMR studies. In parallel, our collaborators from Prof. F. Fieschi's laboratory at IBS, produced constructs (ECD and CRD) of human CLRs mainly Langerin, DC-SIGN and MGL, introduced in chapter II, ready for interaction studies. Thence, we started our research work through series of experiments to screen the couples of PAMP-CLR that could be potentially interactive.

##### VII-1-1-1. NMR spectroscopy and Electron microscopy results

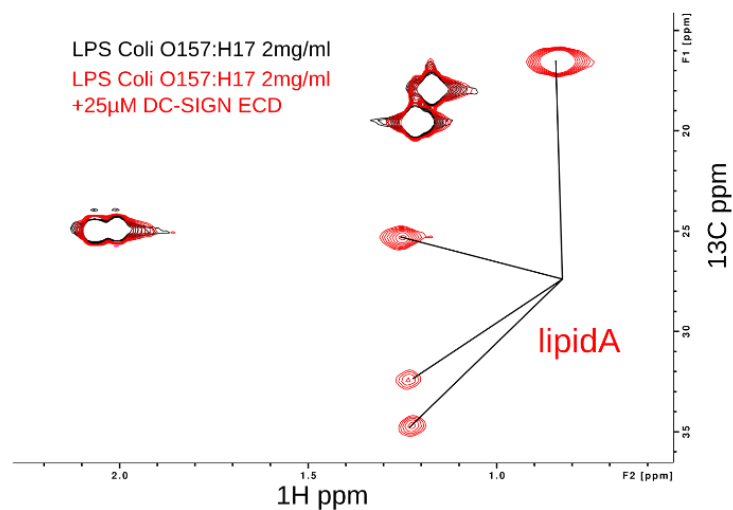
We used liquid state NMR spectroscopy associated with Transmission Electron Microscopy (TEM) considering intact LPS glycoconjugates and not the truncated versions because of LPS high solubility and large size at once. Investigating these properties (i.e., changes in LPS assemblies' sizes or shapes) would provide a preliminary vision about LPS-lectin interaction, if happening. The outcomes generated from those experiments are informative about changes in the environment surrounding certain residues either from Lipid-A or saccharide portions of <sup>13</sup>C-LPS in absence and in presence of human lectins. The appearance or disappearance (intensity changes) or chemical shifts variations of NMR signals will be used as local probes to monitor and characterise the interaction between the two biomolecules. In our case, we expected to see those changes by using liquid NMR spectroscopy. Subsequently, we used TEM to assess the morphology of LPS O157:H7 alone and in the presence of the lectins.

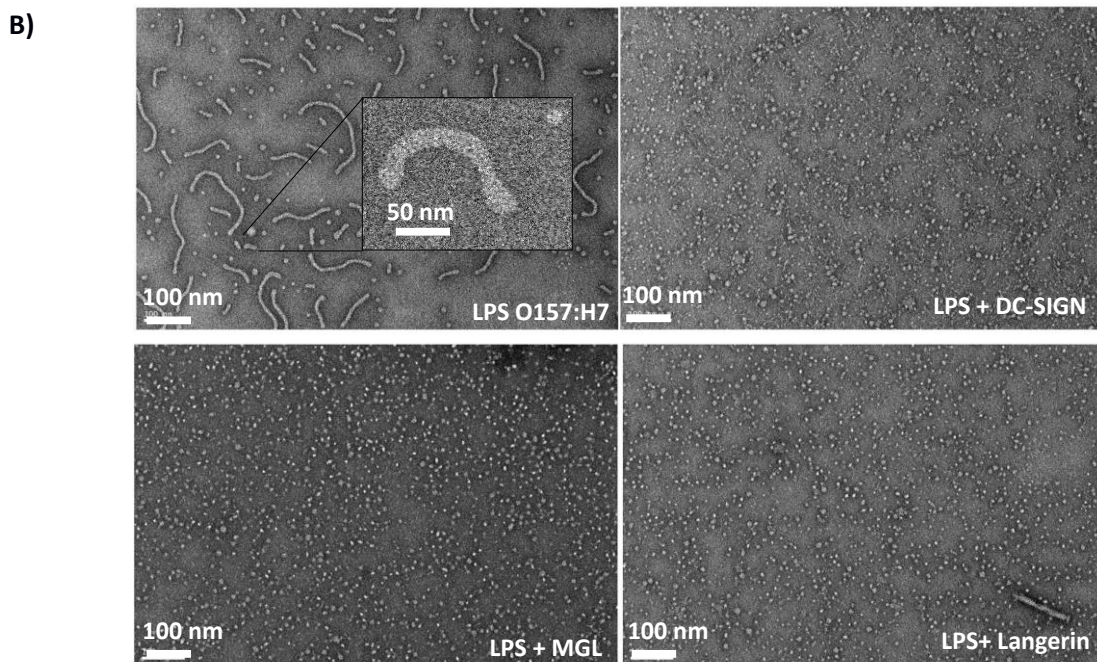
On that account, 1.2 mg/mL of  $^{13}\text{C}$ -LPS O157:H7 suspended in 50 mM  $\text{Na}_2\text{HPO}_4$  150 mM NaCl pH8 (Buffer C) (see chapter XI for details) was subjected to mono and bidimensional  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiment. Interestingly, the addition of the ECD of human lectins was accompanied by appearance of  $[^1\text{H}$ - $^{13}\text{C}]$  cross-peaks of terminal methyl and methylene groups ( $\text{CH}_3$  and  $\text{CH}_2$ , respectively) from Lipid-A fatty acids, especially for DC-SIGN and MGL (figure 74A). This is informative about a probable interaction between the two biomolecules. The appearance of fatty acids NMR signals could translate LPS particles size reduction following interaction with the lectin.

Furthermore, we sought to understand these spectral changes, by investigating these effects through TEM. Again, the later technique would allow us to visually observe the morphology of LPS particles (stabilized by lipids), alone and with the lectin, that might explain the effects detected in the NMR spectra.

In solution, LPS glycoconjugates form small vesicles and/or elongated micelles due to their amphiphilic nature and depending on the carbohydrate portion length. The external surface of elongated micelles is formed by the polysaccharidic portion of LPS and the internal layer is constituted by hydrophobic Lipid-A<sup>348</sup>.

A)





**Figure 74. Disaggregation of  $^{13}\text{C}$ -LPS O157:H7 in presence of human lectins including DC-SIGN observed in  $[^1\text{H}-^{13}\text{C}]$ -2D-HSQC NMR spectra (A) and TEM (B) graphs.** Electron microscopy micrographs are magnified to 23000x with 100 nm scale bar.

By EM, we managed to first confirm the nature of the isolated LPS O157:H7 assemblies, when they are free in solution, which exhibited elongated micelles with various lengths. Interestingly, we were able to visualize the effect of human lectins addition to LPS micelles at 1: 0.01 molar ratio of LPS: lectin through TEM images. We observed complete disaggregation or disruption of LPS elongated structures into small assemblies once the human lectins were added (figure 74B). The same phenomenon has been formerly observed for LPS interaction with LptE (subunit of LPS transporter complex) which disrupted micelles of LPS into smaller particles<sup>349</sup>. In another study, LPS disaggregation has been confirmed as resulting from *E. coli* O55:B5 LPS binding to LBP protein and the CRDs of caspases via TEM<sup>147</sup>.

In our case, this disaggregation effect possibly occurred via the recognition of LPS carbohydrates residues by the CRD of human lectins which led to these drastic changes in LPS assemblies. This is consistent with what we obtained from NMR data highlighted by the appearance of fatty acid signals from Lipid-A in the  $^1\text{H}-^{13}\text{C}$  HSQC spectra following DC-SIGN addition where terminal methyl groups became visible as sharp NMR signals in LPS due to the smaller size of micelles (Figure 74A). Anyhow, data from NMR spectroscopy and EM suggest that DC-SIGN and MGL bind to LPS O157:H7 even at relatively low concentrations of proteins. At this point, questions related to the molecular binding remain to be answered and, to do that successfully, additional approaches must be employed. As for Langerin, the NMR outcomes related to LPS binding

showed less effects following the addition of ECD langerin (data not shown). Besides, high amounts of langerin are required for the establishment of additional approaches like ss-NMR spectroscopy, and this limited our investigations for this human lectin.

Finally, we were able to make sure that human DC-SIGN and MGL interacted with LPS O157:H7 from *E. coli* pathogen. We then decided to test de-acetylated (oligosaccharides or OS) versions of *E. coli* glycoconjugates R1 and R3. The latter possesses identical core structure as LPS O157:H7, thus we should be able to look at these interactions more deeply with higher resolution by getting rid of lipids and keeping only the saccharide part of *E. coli* LPS. In other words, we aim to examine these interactions at the atomic level, and this is achievable only if we consider soluble ligands (OSs) and low molecular weight lectins (monomeric CRD domain), which would give well-resolved NMR spectra with reasonable sensitivity. Otherwise, studying the whole system by NMR would not provide factual results due to LPS insolubility and it would give unclear NMR data.

### **VII-1-2. Interaction between CRDs of human lectins and *E. coli* OSs**

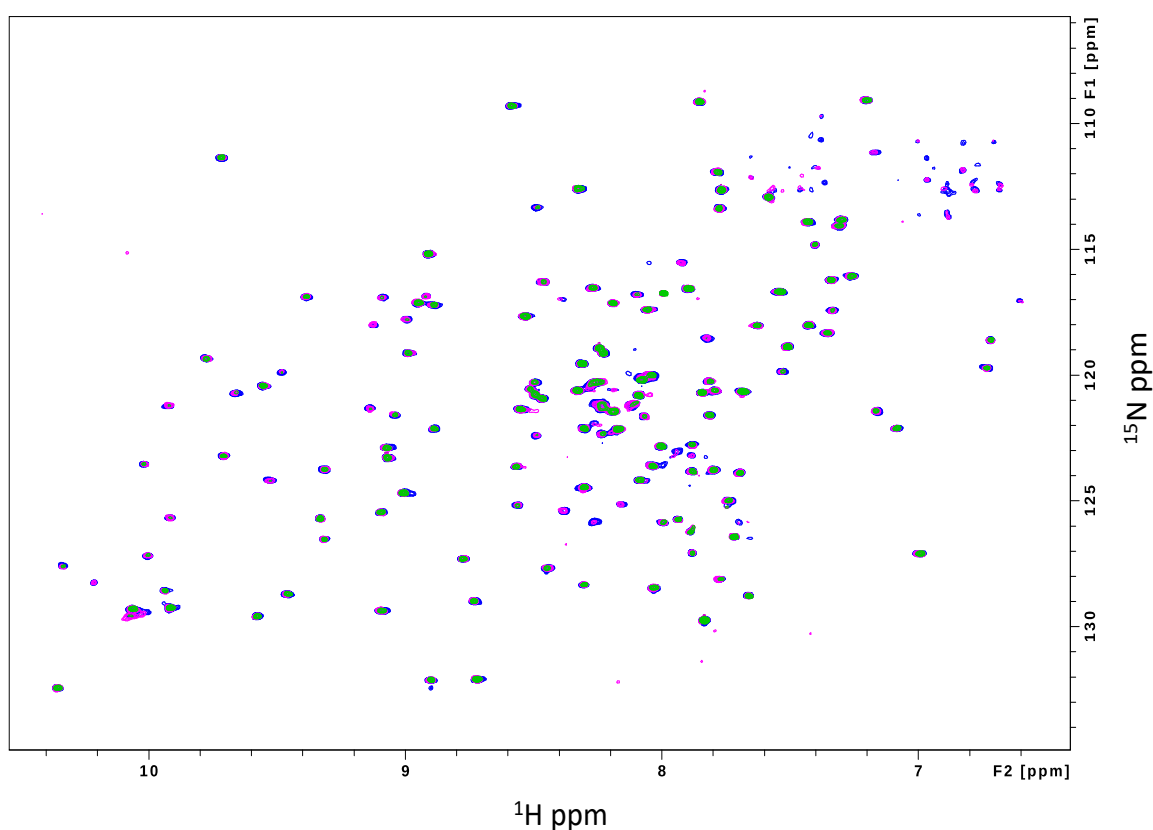
In chapter III, I have introduced a variety of technical approaches that are usually employed for carbohydrate-lectin interactions, among them NMR titrations experiments are useful for mapping and analyzing lectin Chemical Shifts Perturbation (**CSPs**) that may be observed after the stepwise addition of ligands.

In order to figure out the interaction between the human lectins (i.e., MGL and DC-SIGN) and the two *E. coli* LOSs R1 and R3, we tested the interaction between the OSs and short constructs of human lectins (CRDs). In fact, full lectins constructs (ECD) are high molecular weight proteins (28 and 39 KDa for MGL and DC-SIGN ECDs, respectively) presenting strong limitation for a sensitive detection by liquid state NMR spectroscopy. In addition, CRDs of MGL and DC-SIGN have been studied by other research groups by NMR spectroscopy<sup>188,350</sup> thus providing us a template to analyse our NMR titrations findings in terms of residues cross-peaks assignment.

#### **VII-1-2-1 *E. coli* OSs do not change human DC-SIGN spectroscopic profile**

NMR titrations are used to determine protein-ligand binding parameters. Bidimensional [<sup>1</sup>H-<sup>15</sup>N]-2D-BEST-TROSY NMR experiments were carried out first on human <sup>15</sup>N-CRD DC-SIGN in buffer B (25 mM Tris HCl 150 mM NaCl 4 mM CaCl<sub>2</sub> pH 8) on the 850 MHz spectrometer. From these data, we expected to observe chemical shifts of <sup>1</sup>H-<sup>15</sup>N-CRD-DC-SIGN specific residues in case of binding to the ligand, generated from changes in the chemical environment of the residues.

At the beginning of the experiment, a 2D spectrum was recorded showing many [ $^1\text{H}$ - $^{15}\text{N}$ ] correlations as “dots” or cross-peaks that correspond each to N-H group of each residue constituting the protein backbone in the Apo state (or free state). The stepwise addition of an increasing molar equivalents of OS R1 or OS R3 ligands (from 0.5 to 20 eq.) did not produce any perturbations (chemical shifts or peak intensity) in the NMR spectra of CRD DC-SIGN suggesting no interaction. Figure 75 shows a superimposition of the 2D spectra recorded for the protein in Apo state and following the addition of OS R1 and OS R3 ligands at 20 molar equivalents of protein.

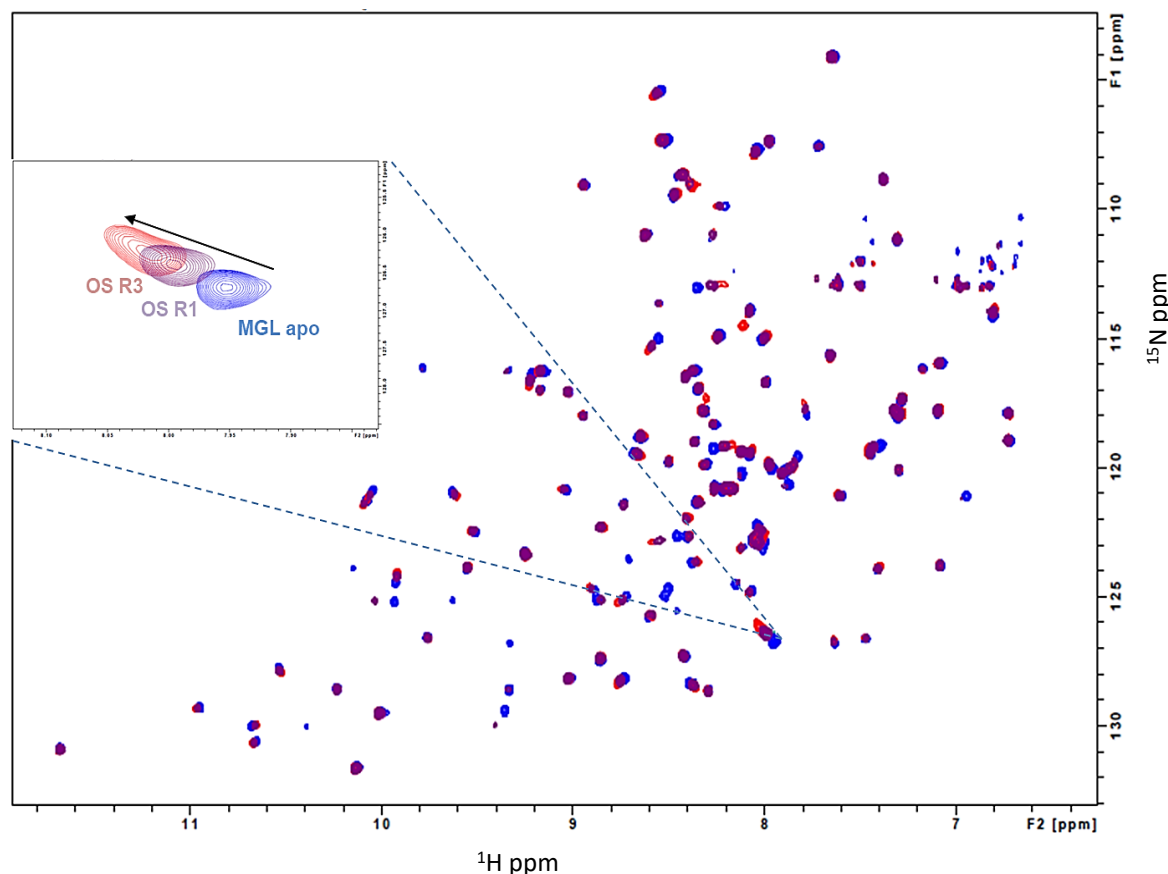


**Figure 75. Superimposition of  $^1\text{H}$ - $^{15}\text{N}$  backbone resonances of human CRD DC-SIGN (blue) in presence of 20 eq. of OS R1 (green) and 20 eq. of OS R3 (pink) in buffer B at 303 K on the 850 MHz spectrometer.**

Furthermore, NMR titrations were established on the isolated  $^{13}\text{C}$ -LPS O157:H7 in presence of CRD DC-SIGN (data not shown) and led to no chemical shift effects on the [ $^1\text{H}$ - $^{13}\text{C}$ ] correlations from LPS molecules (isotopically labelled, observable by NMR). Moreover, an attempt to study another aspect of the interaction by evaluating the effects of full DC-SIGN (ECD) addition to  $^{13}\text{C}$  LPS O157:H7 sample through solid state NMR analyses, resulted in no significant changes in the NMR spectra (Annexes figure7).

## VII-1-2-2 NMR titration of human MGL with *E. coli* OSs

When it comes to human CRD MGL, in-solution NMR studies provided distinct outcomes. Notably, CSPs in presence of both OSs were detected for the [ $^1\text{H}$ - $^{15}\text{N}$ ] backbone NMR resonances of residues (figure 76). First, we noticed that the same residues were affected by the addition of either OS R1 or OS R3 to the lectin in the same experimental conditions as for CRD DC-SIGN. We managed to speedily screen interacting lectin that is human MGL, and more detailed analyses are given in the next chapter.



**Figure 76.** Superimposition of [ $^1\text{H}$ - $^{15}\text{N}$ ] backbone resonances of human CRD MGL (blue) in presence of 20 eq of OS R1 (purple) and 20 eq of OS R3 (red) in buffer B at 303 K on the 850 MHz spectrometer.

## VII-2 Conclusions

By sequential use of coupled approaches like TEM and liquid state NMR analyses on different versions of human lectins and glycoconjugates, we sought to selectively determine relevant biomolecules and thus directing the choice of the interaction system to focus on. Taking into consideration the full set of experiments designed not only for large interaction system (ECD-LPS) but also truncated versions of the interactors (CRD-OSs), we succeeded to select the human lectin to start our study with, that is human MGL. However, *E. coli* glycoconjugates of interest did

not show any interaction with DC-SIGN residues from NMR profiles and this could be related to the absence of specific carbohydrate moieties on OS structures like mannoses. Studying human MGL is particularly interesting owing to its dual function somehow linked to paradoxical biological events such as pathogen clearance and T cells elimination<sup>205,351</sup> and since few papers have been published so far regarding its function as CLR especially when it comes to selective pathogen recognition and more importantly underlying *E. coli* pathogens uptake.



## ***Chapter VIII.***

## VIII. Does the human MGL CRD domain interact with *E. coli* ligands differently?

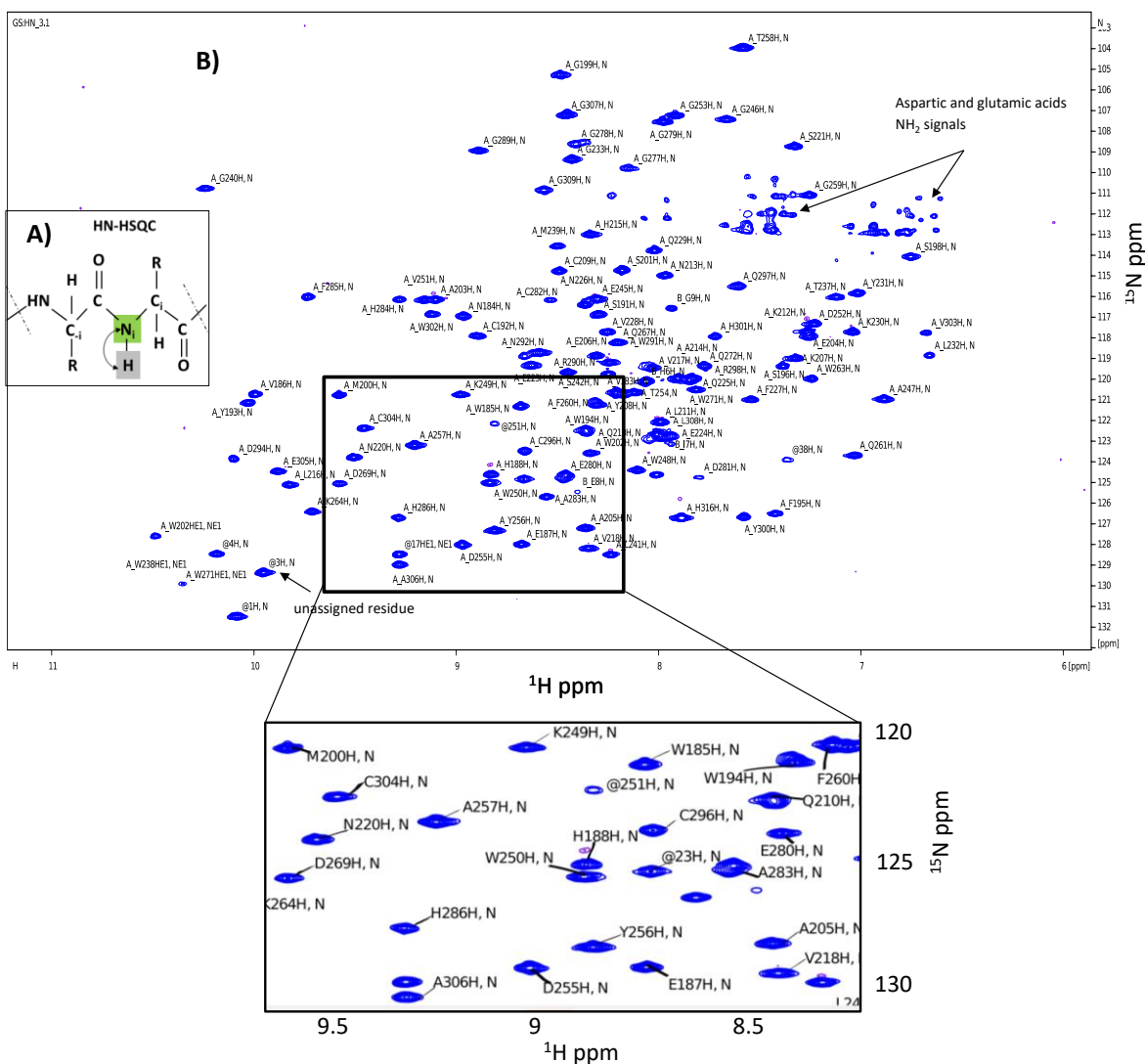
This chapter is meant to present the first results obtained from backbone assignment of human CRD MGL followed by NMR titration experiments using the isolated *E. coli* pathogenic ligands. The work discussed herein is focused on the lectin side of the interaction in terms of mapping of active residues and this would help in defining binding surfaces and, if applicable, estimating apparent affinity and more importantly comparing various ligands bindings on the same MGL CRD biomolecule. The scientific strategy illustrated in chapter V was employed and, the corresponding results are presented from now on. Briefly, the two domains of human MGL, initially CRD (the present chapter) then ECD (next chapter), were studied by using adequate methods. We aimed to figure out how interactions occur according to human MGL structure organization and, whether different *E. coli* LPS versions would induce distinctive binding modes.

### VIII-1. Results and discussions

#### VIII-1-1. Backbone assignment of human MGL CRD by NMR spectroscopy

Once the human CRD MGL is purified in its isotopically labelled and soluble form, the backbone assignment of this CLR was partially established (17% of total residues assignment missing or incomplete, due to ambiguous or absent peaks) by recording several NMR experiments that allow looking at atoms correlations (cross-peaks) from amino acids composing CRD MGL backbone, and thanks to specific NMR software used, that is ccpnmr.v3<sup>®</sup> 352. Backbone assignments of the lectin is, as explained in chapter III, required for further NMR titrations with the ligand to permit the identification of binding surfaces. Binding contributions from unassigned residues, if happening, cannot be used to carefully define the interaction surface.

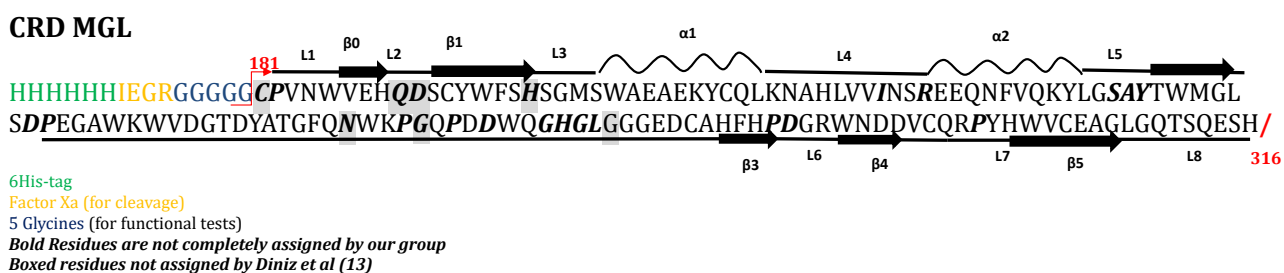
Mono- and bi-dimensional NMR spectra are routinely recorded as initial experiments providing primary observable signals from residues, indicating well-structured (folded) protein, and affirming the choice of the right range of temperature for its structural stability. Figure 77 shows the fingerprint of <sup>15</sup>N-<sup>13</sup>C-CRD backbone constituting human MGL CRD where each cross-peak corresponds to amide backbone correlations between <sup>1</sup>H and <sup>15</sup>N atoms.



**Figure 77. (A) Graphical description for transfer of magnetization during a  $[^1\text{H}-^{15}\text{N}]$ -2D-Best-TROSY-HSQC experiment applied to proteins (B) 950 MHz  $[^1\text{H}-^{15}\text{N}]$ - BEST-TROSY-HSQC spectrum of human  $^{15}\text{N}$ - $^{13}\text{C}$  -CRD MGL in buffer B at 303K. For clarity, A- and B- labels in residues nomenclatures (on the entire spectrum) refer to CRD MGL NMR chains.**

For  $^{15}\text{N}$ - $^{13}\text{C}$ -CRD MGL backbone assignment, we recorded different heteronuclear 3D-experiments (HNC(O), HN(CA)CO, HN(CO)CACB and HNCACB) highlighted in table 4 of chapter III using optimized pulse sequences<sup>310</sup>, and with a nice spectral quality.

The assignment strategy using Ccpnmr® software permitted us to characterize the backbone of CRD MGL thus assigning cross-peaks in the  $[^1\text{H}-^{15}\text{N}]$ -2D-Best-TROSY-HSQC spectrum to their corresponding amino acids, which is crucial for further NMR titration experiments of CRD MGL with LPS ligands. 82% of human MGL CRD backbone resonances were assigned and compared to NMR assignment established by Diniz et al<sup>350</sup>. Unassigned residues were basically located at Loop 2 and Loop 5 that harbors the first canonical binding site. Figure 78 shows human MGL sequence highlighting assigned and non assigned residues, by our group and by<sup>350</sup>.



**Figure 78. Illustration of amino acid sequence backbone assignments of human MGL CRD.** The CRD MGL backbone assignment is overlaid with the one from Diniz et al<sup>350</sup> who have used different experimental conditions (lower temperature and higher concentrations of lectin and of Ca<sup>2+</sup> ions in the buffer). In our CRD MGL sequence, the 15 initial residues were added by our collaborators for the purification of the protein and for functional tests belonging to them.

## VIII-1-2 Human MGL interaction profiles with *E. coli* LPSs by NMR titrations

### VIII-1-2-1 Interaction studies between OSs and CRD of human MGL

OS R1 and OS R3 ligands (deacylated LOSs versions) were produced and purified as described in chapter VI and subsequently subjected to interaction studies through solution NMR spectroscopy analyses. Three ligands (GalNAc, OS R1 and OS R3) were added stepwise to the protein at different number of equivalents up to 20 eq.

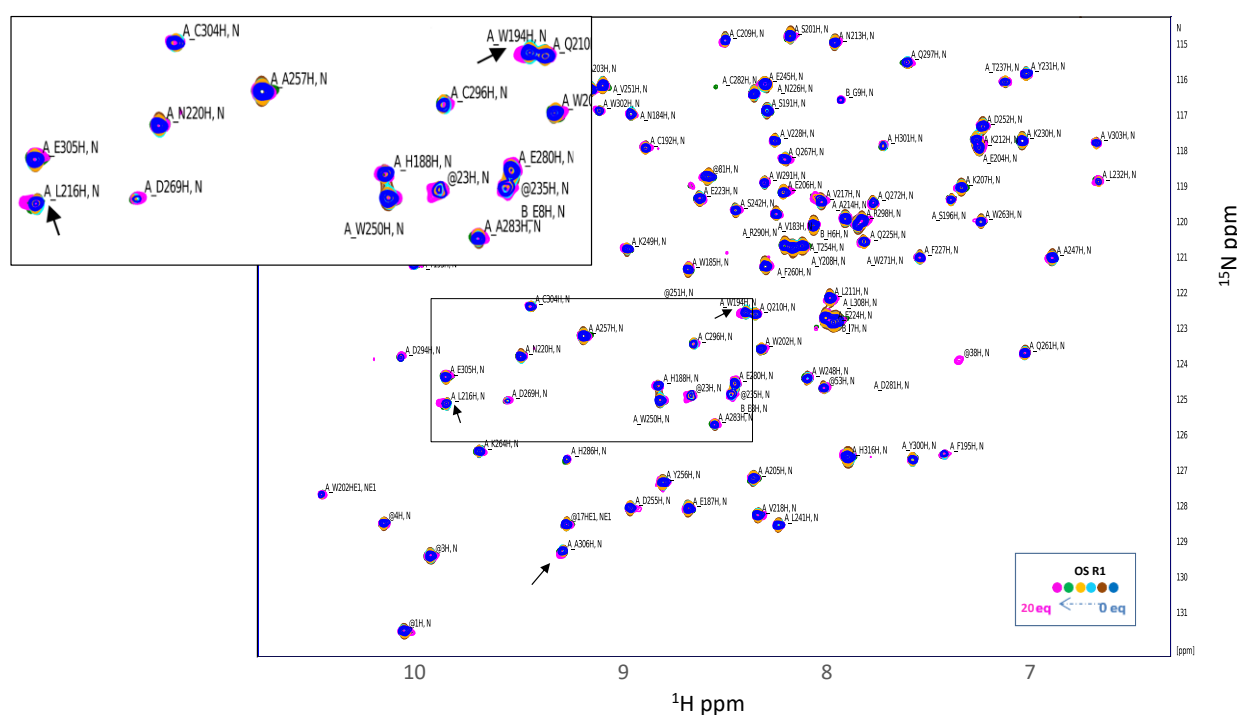
#### VIII-1-2-1-1. Human MGL CRD interacts strongly with GalNAc sugar

The capability of MGL to bind to GalNAc sugars has been confirmed<sup>353</sup>. We used single GalNAc sugar, which is known to be a calcium dependent binder to MGL, as a positive control to compare with the ligands tested including OS R1, and this has helped us to find out interesting insights concerning the binding surface. Globally, the CSP profile produced after the addition of GalNAc showed similar binding regions as for GalNAc-bound human MGL CRD that has been investigated by Diniz et al<sup>350</sup> by NMR titrations (spectra in annexes part). Important residues (e.g., D269) from the QPD motif that are specific for GalNAc binding in CLRs exhibited significant CSP in the presence of GalNAc residues. This residue, together with Q267, are crucial for complexation with Ca<sup>2+</sup> which, in turn, mediates contact with OH-3 and OH-4 groups of GalNAc, as demonstrated by<sup>350</sup>. The authors have also detected a slow exchange rate from that sugar during the titration that is associated with a low dissociation constant ( $K_{off}$ ) between free and GalNAc-bound CRD MGL states. Consequently, they did not determine the affinity for individual GalNAc residues. The latter is known to exhibit high to medium affinity ranges for MGL binding depending essentially on its anomeric configuration ( $\alpha$  or  $\beta$ , with high affinity for  $\alpha$ -GalNAc)<sup>354</sup>. Finally, by comparing our results with another study, we validated the specificity related to the

functionality of CRD MGL and the similarity of our findings with the published results and, this would facilitate further analyses.

### VIII-1-2-1-2. *E. coli* OSs R1 and R3 perturbed identical residues on MGL CRD

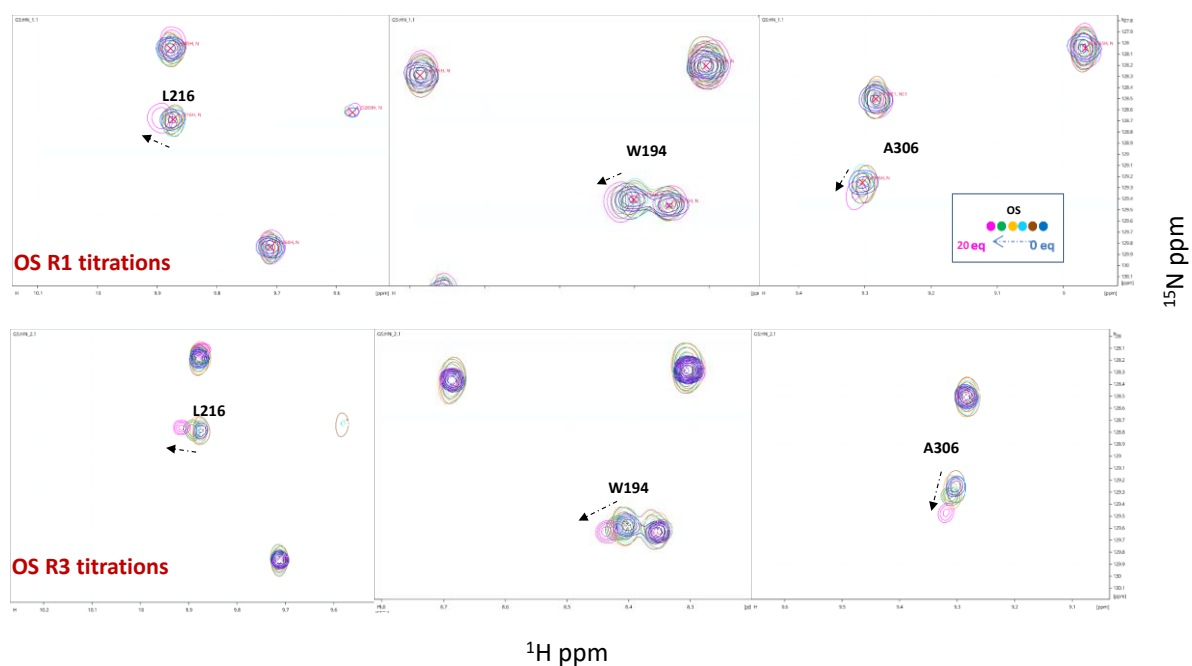
To get deeper knowledge about human MGL CRD binding to *E. coli* OSs, we continued NMR titrations in the same conditions. As expected, OS R1 produced weak chemical shift perturbations (CSPs) on some of the protein <sup>1</sup>H-<sup>15</sup>N backbone resonances. Although the CSPs are small in magnitude, specific residues were affected following OS R1 addition at the different molar ratios i.e., 1, 2, 5, 10 and 20 molar equivalents (figure 79). NMR titration results from OS R1-bound state of MGL showed proximal region to Ca<sup>2+</sup> binding site being perturbed with GalNAc, this is foreseeable knowing that the structure of this particular ligand contains galactoses Gal in the core OS. In addition to those protein residues, another region was involved in OS R1 binding, close to the N-Terminus of the protein.



**Figure 79. 700 MHz (top) [<sup>1</sup>H, <sup>15</sup>N]-2D-Best-TROSY-HSQC superimposed spectra of 50 μM <sup>15</sup>N-CRD MGL at the different OS R1 molar ratios.** Backbone resonances are showed during stepwise addition of the ligand in buffer B at 303K.

These results were similar to NMR titrations with OS R3 however chemical shifts perturbations were slightly more pronounced for OS R3 compared to OS R1 (figure 80). At this stage, it is not possible to estimate K<sub>d</sub> because the protein was not fully saturated by OSs ligands, even at 20 eq., and so titration curves did not reach saturation and does not allow to calculate a K<sub>d</sub>

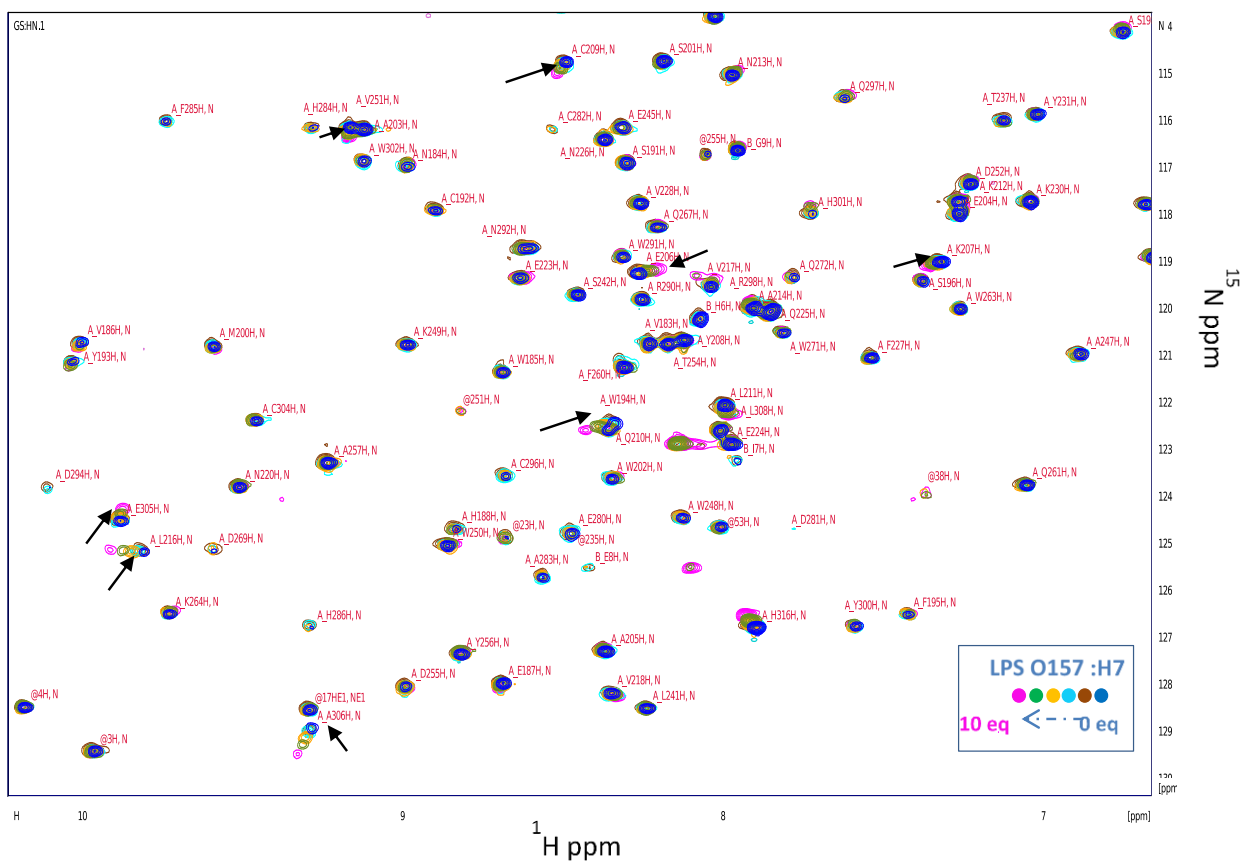
accurately. Nevertheless, higher CSPs that were observed in the case of OS R3 suggested a slightly favoured affinity for that ligand.



**Figure 80.** 700MHz [ $^1\text{H}$ ,  $^{15}\text{N}$ ]-2D-Best-TROSY-HSQC superimposed spectra (zoomed at specific regions) of 50  $\mu\text{M}$   $^{15}\text{N}$ -CRD MGL at the different OS R3 and OS R1 molar ratios. Backbone resonances are showed during stepwise addition of the ligand in buffer B at 303K.

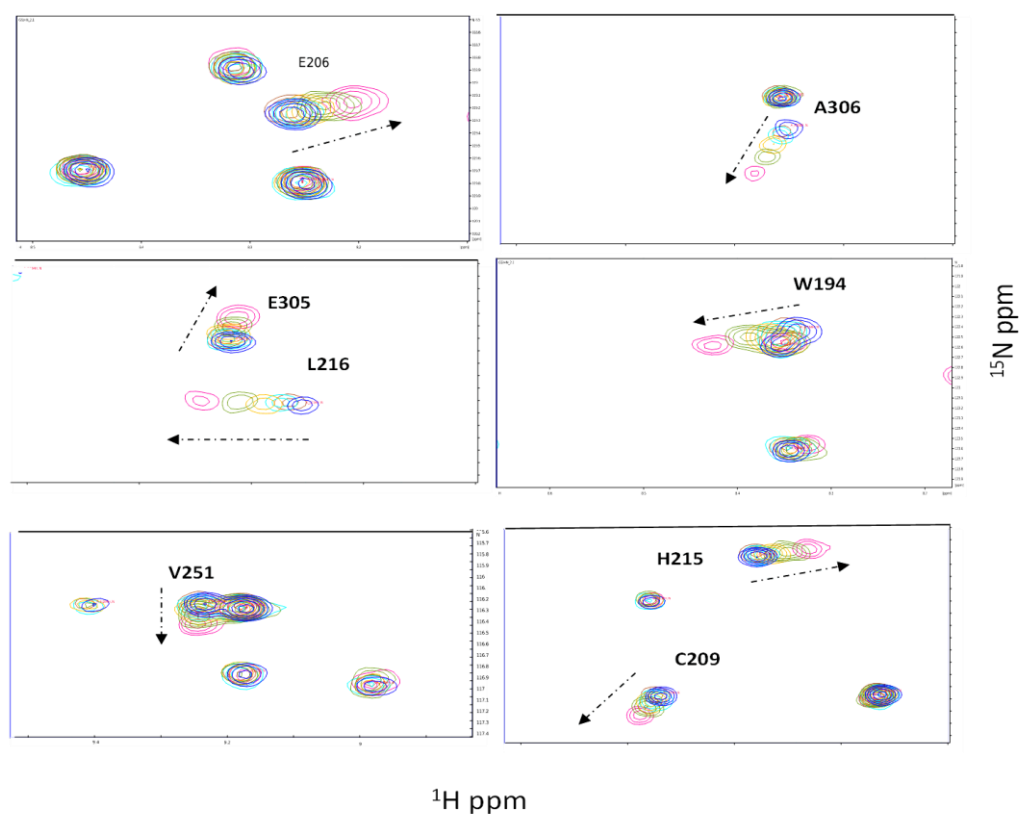
### VIII-1-2-2 CRD of MGL recognizes intact LPS O157:H7 scaffold

The previous observations opened the question if more extended glycoconjugates scaffolds from *E. coli* pathogens would behave the same way or differently considering their native state (without any prior chemical treatment). Having in hands glycoconjugates like LPS O157:H7 with high solubility and structural similarity to the OSs (e.g., identical core structure to OS R3 ligand), we performed NMR titrations on that ligand and the [ $^1\text{H}$ - $^{15}\text{N}$ ]- BEST-TROSY-HSQC spectra were recorded during stepwise addition of LPS at the different molar ratios (figure 81). Indeed, we managed to observe distinct contributions from the isolated  $^{13}\text{C}$ -LPS O157:H7. Yet again, LPS O157:H7 provoked CSPs were higher with respect to OS R3.



**Figure 81. 850MHz [ $^1\text{H}$ ,  $^{15}\text{N}$ ]-2D-Best-TROSY-HSQC superimposed spectra of  $50\ \mu\text{M}$   $^{15}\text{N}$ -CRD MGL at the different LPS O157:H7 molar ratios. Backbone resonances are showed during stepwise addition of the ligand in buffer B at 300K.**

We succeeded to spot residues that exhibited valuable CSPs (figure 82), these include W194, E206, C209, H215, L216, V251, E305, A306, H316. Many of those residues are identical to residues experiencing CSPs from OSs R1 and R3, according to the NMR titrations spectra and to the backbone assignments.



**Figure 82. Chemical shift changes of human  $^{15}\text{N}$ -CRD MGL active residues following LPS O157:H7 binding.**

For a given molar ratio, at 10 eq. for example, some of those signals (e.g., A306, L216 and W194) were shifted also in the case of the other OSs especially OS R3, however CSP values for LPS O157:H7 presented higher magnitudes and well-defined shifts trajectories compared to CSPs provoked by the OSs. It is note worthy to mention that the molarity of LPS O157:H7 is just an estimate due to its polydispersity. Despite that drawback, we observed higher CSPs for lower molar ratios.

Titration curves permitted us to estimate binding affinity for LPS O157:H7. Using ccpnmr.v3 software, NMR titration data were fitted and showed typical curves for weak binding with an apparent  $K_d$  of 5.8 (+/- 5) mM (Annexes figure9). These weak affinity bindings are usually observed for carbohydrates lectins interactions in fast exchange regimes<sup>355</sup>. According to titration data, it seems to us that binding strength is reduced for the case of the OS interaction with CRD MGL in comparison with LPS O157:H7. This could be correlated with structural features of both ligands, for instance, LPS and OS R1, which expose GalNAc and Gal in their outer carbohydrate core, respectively. Indeed, GalNAc is much stronger ligand than Gal, thanks to the establishment of additional H-bonds with CRD MGL<sup>203</sup>. This may explain our interpretation of



NMR titration for LPS O157:H7 that could possibly interact with MGL either through the O-Antigen and/or the core. Presentation of extended carbohydrates in O-Antigen LPS O157:H7 would increase indeed the probability of binding multiplicity<sup>356,357</sup>. At this stage, binding contributions from the O-Antigen portion is a hypothesis that should not be discarded.

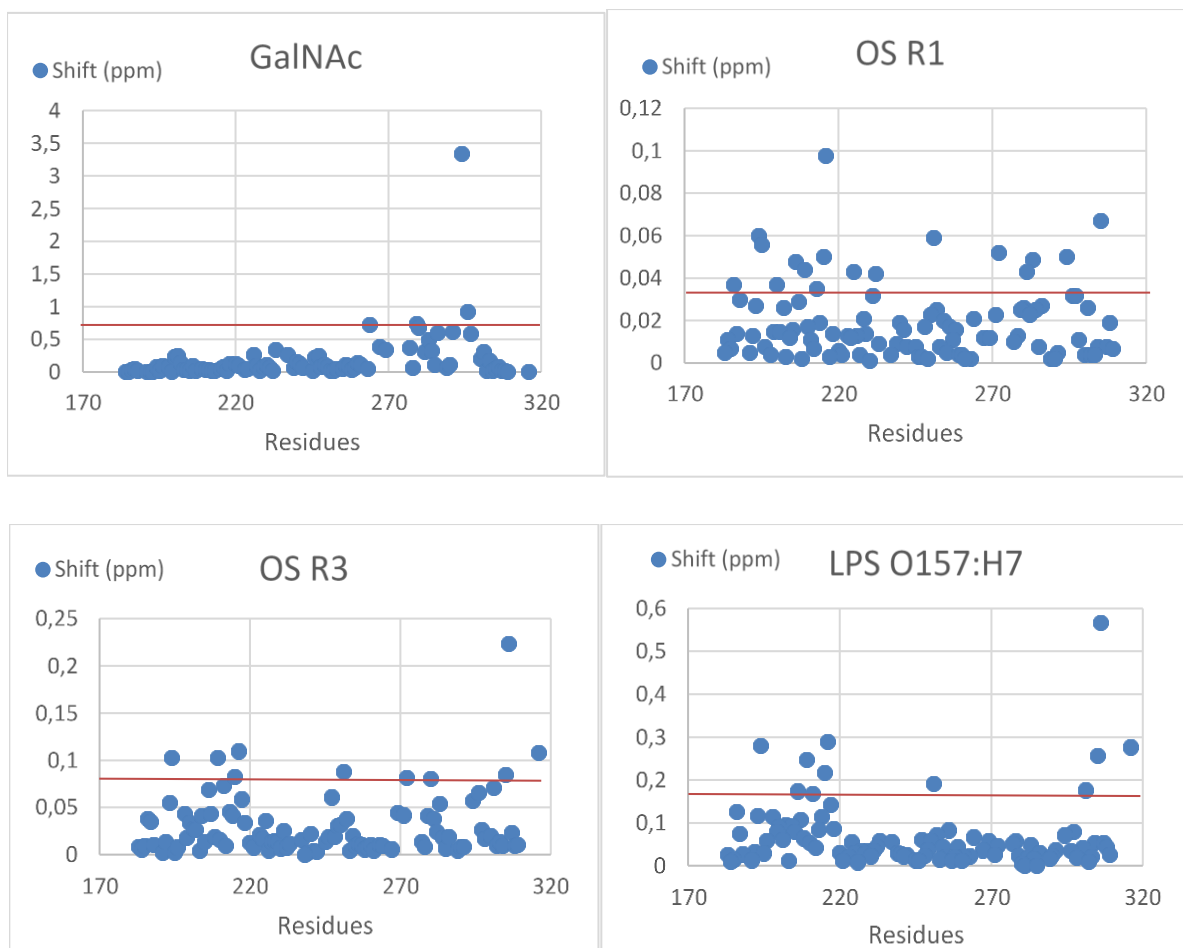
#### **VIII-1-2-2-1. LPS O157:H7 is not believed to interact with MGL through lipids**

Our previous findings may suggest that CRD MGL has stronger binding preferences for intact LPS or LOS structures, over chemically treated or modified glycoconjugates specimens. One may think that the Lipid-A could also be recruited by CRD MGL for interaction, however we observed, via titrating N-dodecyl- $\beta$ -D-maltoside (DDM, a di-glucose linked to a fatty acid chain) (Annexes figure10) to human  $^1\text{H}$ - $^{15}\text{N}$ -CRD MGL, that residues cross-peaks did not move, so we presume that there are no interactions between LPS O157:H7 Lipid-A and the human CRD MGL.

#### **VIII-1-2-3 A second potential binding site on CRD MGL**

As shown using titrations experiments, NMR profiles of some residues from human CRD MGL were affected following *E. coli* ligands binding and this was assessed by CSPs analyses. These analyses include protein mapping for binding site identification from NMR data. We sought to have a closer vision of *E. coli* ligands binding sites on CRD MGL and compare interaction results. This is an important step for our study because it allows a relevant prediction of binding surfaces that the CRD domain may arrange depending on the ligand and on its oligomerization state.

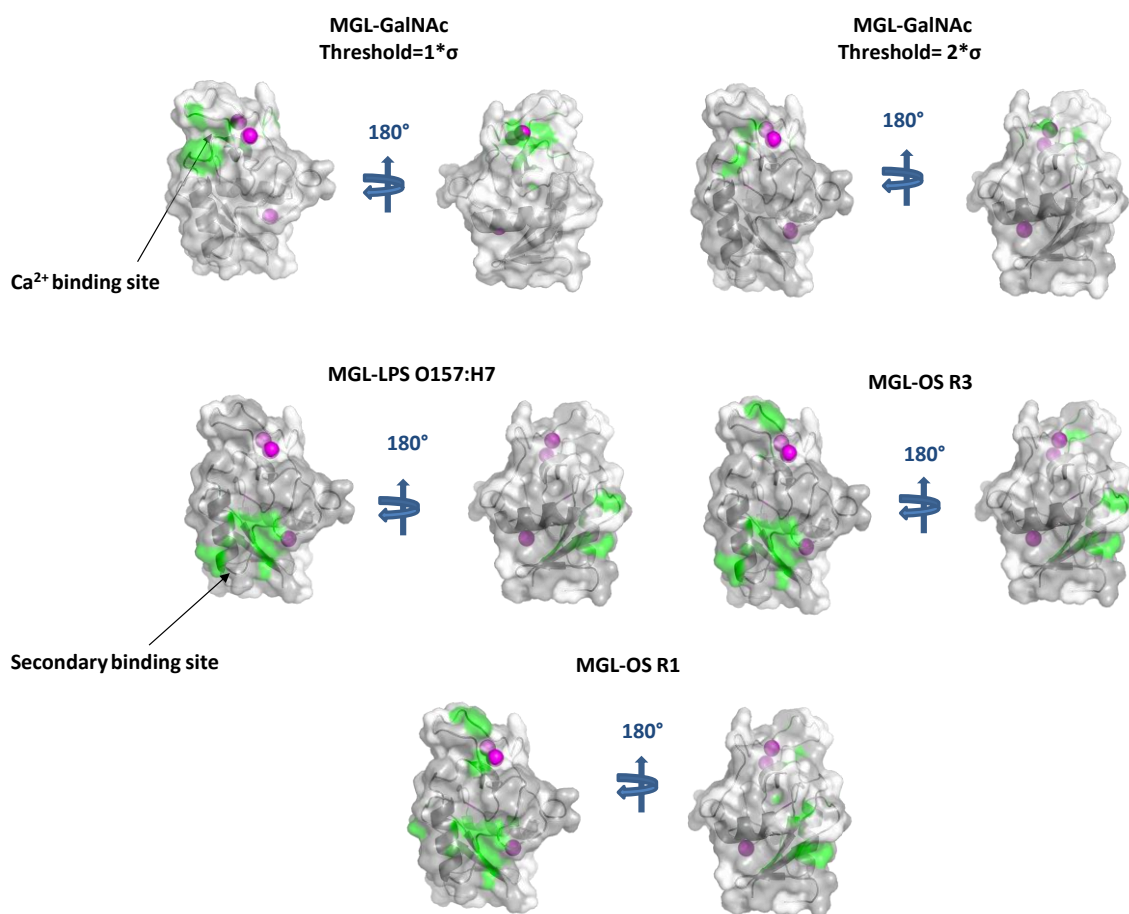
Figure 83 presents CSPs distributions based on NMR chemical shift changes (CSPs) and according to the assigned amino acids backbone. We immediately visualized a secondary binding region being affected by the presence of the three glycoconjugates, that we estimated to be likely close to the N-terminal side of the lectin. We suggest that there is a potential secondary binding site implicated in all tested *E. coli* glycoconjugates ligands interactions, knowing that ability of human CRD MGL to dynamically exploit binding surfaces when interacting with GalNAc-containing ligands as it has been demonstrated through NMR spectroscopy and MD simulations<sup>350</sup>.



**Figure 83. Representation of chemical shifts scatters of the four studied ligands over the backbone of human MGL CRD.** Data derived from NMR titrations considering single  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift values for each residue, considering the corresponding threshold values ( $2\sigma$ ) delimited by the red line.

#### VIII-1-2-3-2. CRD MGL mapping highlighted an extended binding surface

CSPs were measured following *E. coli* ligands additions to CRD MGL and then represented into a CRD MGL model that we predicted considering the strong CRD structure similarity between human MGL and Asialo-glycoprotein (pdb 1PV8) lectin and according to the backbone assignments of the studied CRD MGL, knowing that the 3D structure of this lectin was not available at that time.



**Figure 84. Potential binding surfaces following the interaction of *E. coli* glycoconjugates ligands deduced from NMR titrations with human CRD MGL.** Threshold values were calculated from two times standard deviation ( $\sigma$ ) from  $^1\text{H}$  and  $^{15}\text{N}$ , for all four ligands, giving these values 0.72, 0.16, 0.035 and 0.066 for GalNAc, LPS O157:H7, OS R1 and OS R3, respectively. Residues that are involved in ligand binding are depicted in green and those who were not assigned in white for clarity sake and, calcium ions are colored in pink.

By comparing the different complexes obtained from the projection of NMR titrations data (i.e. chemical shifts) on CRD MGL predicted model (figure 84), MGL-OS R3 and MGL-LPS O157:H7 displayed very similar results regarding the predicted secondary binding site. This may result from the strong similarities between LPS O157:H7 and OS R3 in terms of core oligosaccharide composition and, especially that both ligands derive from the same mother strain (i.e., *E. coli* O157:H7).

Commonly, CSPs from the three tested *E. coli* ligands were mainly located at a secondary binding surface and this does not completely exclude the involvement of canonical binding sites. Similarly, Aretz et al have studied the DC-SIGN (CRD) lectin's capability of screening various druggable molecules, as ligands, by using NMR-based approaches and SPR essays<sup>358</sup>. The authors

supported the existence of secondary sites (up to five) on DC-SIGN CRD domain for the establishment of key interactions.

Additionally, Hanske et al have suggested that, for human langerin which belongs to the same lectin family of MGL, the existence of a secondary binding surface, adjacent to the canonical one, is strongly correlated with the size of the glycan ligand<sup>359</sup>. It should be noted that the well-defined disposition of sugars on elongated ligands scaffolds usually comes up with a complex multiple binding system that defines several types and positions of binding.

## VIII-2. Conclusion

Overall, we observed binding events at an extended surface on CRD MGL, independently of which glycoconjugate ligand is tested, this leads us to think about the existence of a large interaction surface. NMR titrations of *E. coli* glycoconjugates (i.e., OSs R1 and R3, LPS O157:H7) on <sup>15</sup>N-CRD MGL showed similar chemical shifts perturbation profiles with apparent K<sub>d</sub> value in the millimolar range resulting from fast exchange binding regimes observed from LPS O157:H7 binding to CRD MGL. Furthermore, this is the first work that proposes a secondary binding site used by human MGL for *E. coli* pathogens recognition. We expect that this finding might contribute to the fine-tuning of carbohydrates recognition by CLRs like human MGL through a secondary binding surface, frequently observed on CRDs of CLRs, thereby empowering MGL specificity and function as a receptor for *E. coli* ligands. Yet, this interpretation has to be confirmed by further investigations.

Although the second binding site is basically shared between the three ligands, various binding modes may occur. This is possibly due to the different ligand structures (especially the outer core), thus directing CRD MGL Ca<sup>2+</sup>-dependent binding in favor of OS R1 and R3 (primary binding site) at the first binding site which could be perhaps not specific for LPS O157:H7. Still, adding EDTA to chelate calcium ions, for instance, may confirm or not this hypothesis but this chelator agent may strongly affect CRD MGL structure possibly rich in Ca<sup>2+</sup> sites, thus hampering NMR analyses. There is a way to ascertain the involvement of Ca<sup>2+</sup> binding site in the interaction with our ligands, which is about mutating residues close to that binding area and then evaluating mutagenesis effects on the recognition of the ligands.

We questioned about the new binding site contribution; does it really stabilize the interaction with a specific *E. coli* ligand? Is it mainly based on contacts from the inner core of *E. coli* ligands considering its identical shared structure? To answer the first question, at least in part, ligands

from other strains should be considered in the future as controls for further validation of CRD MGL specificity to *E. coli* ligands. The second interrogation could be resolved for example by titrating unlabelled CRD MGL on isotopically labelled versions of the same *E. coli* ligands to try to localize the binding area on the ligand.

Finally, a good strategy to understand if and how the binding may happen in nature is to extend our investigations on larger systems counting both the lectin and the glycoconjugates sides. The next chapter will describe our results obtained from ECD MGL interactions with our *E. coli* glycoconjugates using a combination of additional NMR methods (STD-NMR) and biomolecular (BLI) and biophysical (EM) essays together with microscopy and cytometry analyses.

## ***Chapter IX.***

## **IX. Investigations of the large interaction system offer new findings**

In the previous chapter, the CRD construct of human MGL was studied despite the lack of structural information. We succeeded to predict the ability of this lectin to recognize the isolated *E. coli* ligands (via the saccharides portions, OSs) through more than one binding site, even with partial outcomes regarding the affinity of the CRD binding to OSs R1 and R3. To gain more insights about these interactions, we sought figure out whether if the whole **ECD construct** of the same lectin would accommodate *E. coli* ligands and if it is the case, what are the sugars involved? In this chapter, results will be focused on the ECD construct of MGL. Several methods will be presented, based on NMR spectroscopy and additional methods that we initially carried out on ECD MGL with different versions of *E. coli* ligands until reaching the large recognition system including whole *E. coli* bacteria.

### **IX-1. Results and discussions**

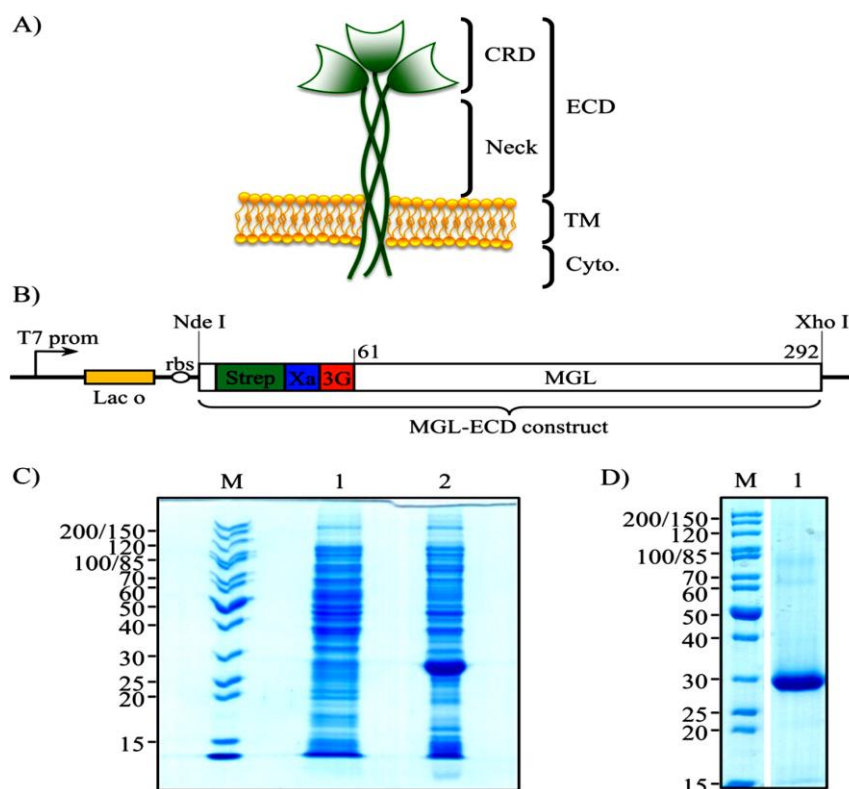
#### **IX-1-1. Results from ECD-MGL interaction with *E. coli* OSs through mono-dimensional STD-NMR**

As presented in chapter III-2, STD-NMR spectroscopy is dedicated for large proteins interaction studies by fast screening of relatively small ligands. By using this technique, it is possible to achieve ligand mapping without the need of processing and analyzing NMR parameters from the lectin side, knowing besides that this approach tolerates high molecular weight and unlabelled receptors like the intact ECD MGL. We employed STD NMR technique to investigate the interaction between the ECD MGL and the oligosaccharides (OSs R1 and R3) previously isolated from *E. coli* strains.

##### **IX-1-1-1. The extracellular domain ECD of human MGL as a receptor in STD-NMR**

Once [<sup>1</sup>H-<sup>13</sup>C] NMR resonances were assigned for OS R1 as presented in chapter VI, the capability of ECD MGL to recognize and interact with OSs R1 and R3 was assessed by means of monodimensional STD NMR spectroscopy.

Trimeric ECD of human MGL was produced, purified, and refolded by our collaborators at IBS (F. Fieschi's research group) (figure 85). Detailed procedure for human MGL ECD expression and purification could be found in supplementary information of <sup>344</sup>.



**Figure 85. Cloning and purification of recombinantly expressed MGL-ECD.** A) Domain organisation of MGL. TM: transmembrane domain. Cyto.: cytoplasmic domain. B) Construct for the overproduction of MGL-ECD. Strep: StrepTag II. Xa: factor Xa protease cleavage site. 3G: tri-glycine linker. C) SDS-PAGE analysis of MGL-ECD overexpression. Lane M: Page Ruler unstained protein ladder; lane 1: total proteins before induction; lane 2: total proteins after induction. A band at about 28 kDa corresponding to MGL-ECD MW is overexpressed. D) SDS-PAGE analysis of purified MGL-ECD. Lane 1: A unique band at about 28 kDa corresponding to pure MGL-ECD is observed.

In order to have an idea about the oligomeric organisation of ECD MGL, we employed diffusion ordered 2D NMR spectroscopy (DOSY) and this would enable the determining of MGL oligomeric state in solution. A DOSY experiment generates a 2D spectrum with NMR chemical shifts in one dimension (F1, horizontal axis) and diffusion coefficients in the other one (F2, vertical axis). In fact, molecules are analyzed according to their diffusion coefficients resulting from molecular association of oligomers, thus providing information about the averaged molecular weight of those molecules. The Stokes-Einstein equations (eq x, y)<sup>360</sup> hold the relation between the diffusion coefficient of a certain molecule in solution and its molecular size:

$$D = k_B T / 6\pi\eta R_h \text{ (m}^2 \text{ s}^{-1}\text{)} \quad \text{eq. x}$$

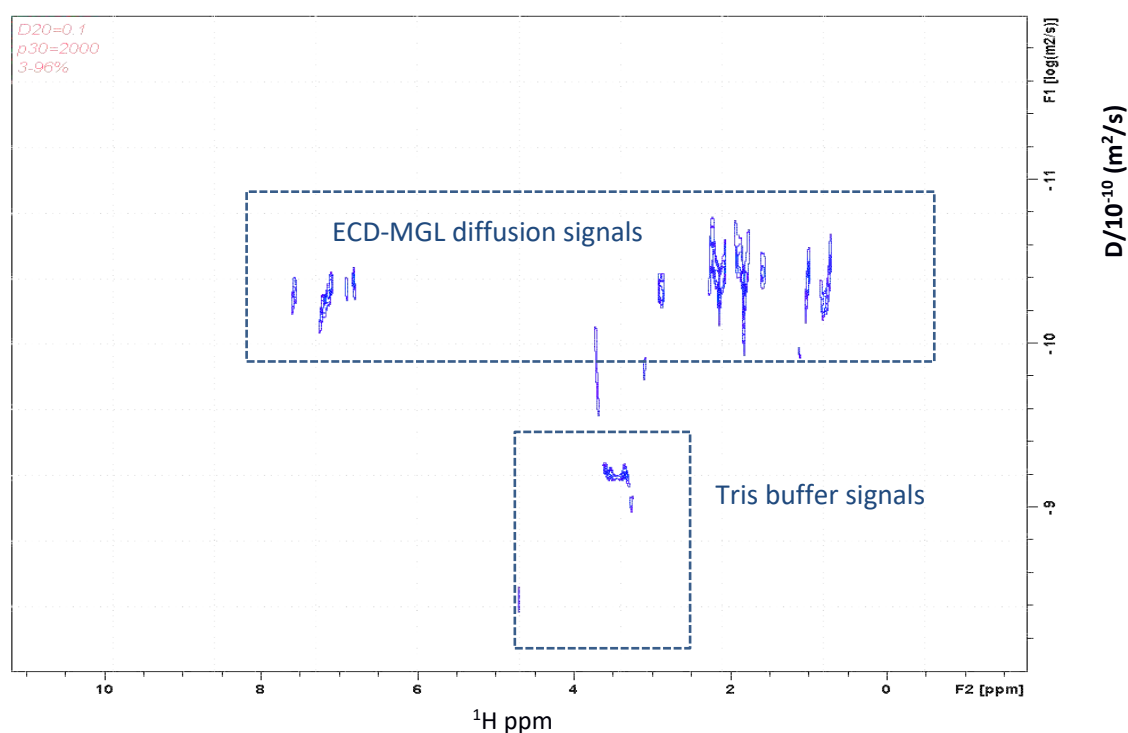
$$D = K \times MW^\alpha \text{ (m}^2 \text{ s}^{-1}\text{)} \quad \text{eq. y}$$

Where D is the diffusion of spherical molecules, such as a globular protein in a continuum medium (where the solvent molecules are much smaller than the solute), which is indeed related



to the molecule's hydrodynamic radius,  $R_h$ , and to the solvent viscosity,  $\eta$ .  $k_B$  is Boltzmann constant and  $T$  the absolute temperature of 25 °C.  $K$  and  $\alpha$  are appropriate scaling parameters.

A quantitative analysis was established, and diffusion coefficients data were fitted starting from  $^1\text{H}$ -1D NMR signals integration. The interpolation of DOSY NMR diffusion data (figure 86) by linearization of Diffusion equation (eq y) provided us the averaged molecular masse for ECD MGL. The molecular weight of ECD MGL (86000 Da) was roughly estimated confirming the formation of a **trimer** in solution, since ECD MGL monomer has a molecular weight of 28 KDa as reported in figure 85.

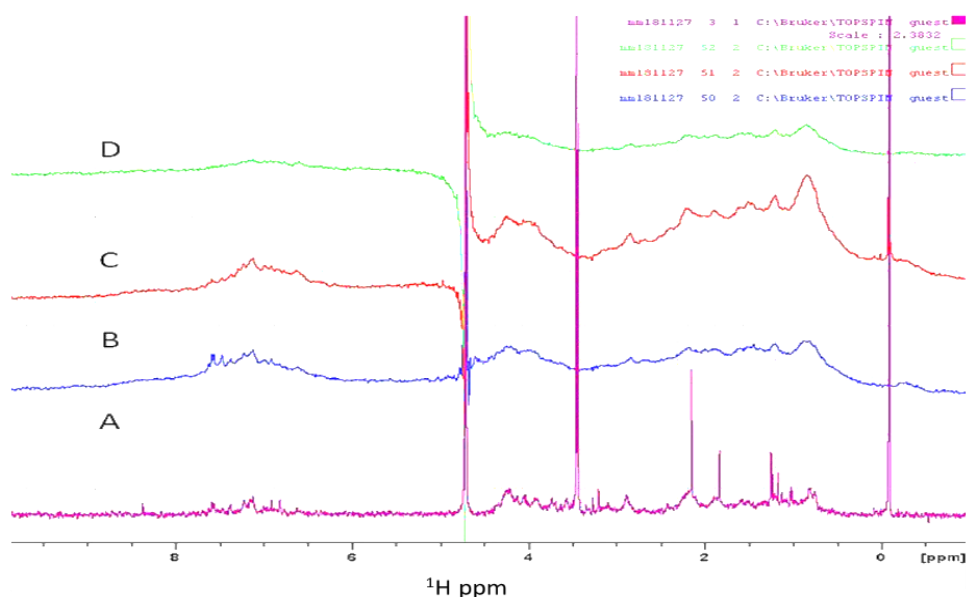


**Figure 86. 600 MHz-2D DOSY spectrum of 20  $\mu\text{M}$  final concentration of ECD MGL.** Lectin sample was prepared in 25mM Tris- $\text{d}_{11}$ , 4mM  $\text{CaCl}_2$ , 150 mM NaCl pH 7 in  $\text{D}_2\text{O}$  (buffer A). Diffusion signals from ECD MGL are present in the upper part of the spectrum between -10 and -11  $\log(\text{m}^2/\text{s})$ . In the lower part of the spectrum, diffusion signals from molecules which move fast (i.e. small molecules like Tris at 3.45 ppm and water molecules HOD at 4.77 ppm) are also detected.

### VIII-1-1-2. The deacetylated LOS R1 is recognized by human ECD MGL

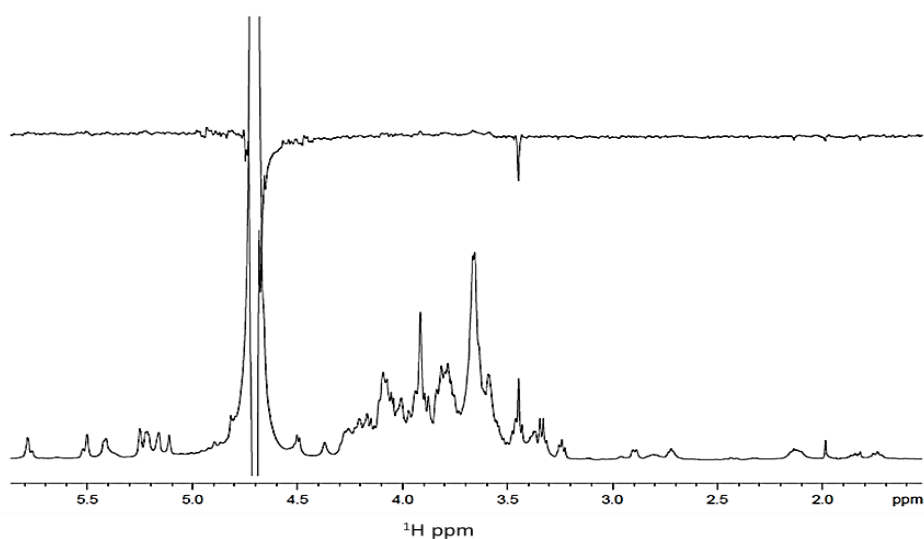
STD-NMR analyses were performed on both OSs R1 and R3 in the presence of trimeric MGL ECD. Lectin sample was first subjected to STD-NMR measurements to select the optimal irradiation frequency far from sugar resonances (figure 87). The on-resonance frequency was set at 7.5 ppm (aromatic residues) since no STD ligand signals were observed in the corresponding

reference spectrum, while 100 ppm was set as the off-resonance pulse frequency. Low quantity of human MGL is required since a high protein: ligand ratio is often used for STD NMR analyses.



**Figure 87. 600 MHz  $^1\text{H}$ -1D-NMR spectra of ECD MGL (27.8  $\mu\text{M}$ ) in buffer A.** The efficiency of on-resonances 7.5 ppm, 0 ppm and -1 ppm in STD-diff spectra (B, C and D, respectively) were compared. Spectrum A corresponds to the  $^1\text{H}$ -NMR profile of the lectin.

Control STD NMR experiments were performed on the ligands as well, OS R1 and OS R3, in the free state at different irradiation frequencies, and by using the same experimental conditions set for the next experiments on ECD MGL-OS mixtures to avoid artefacts due to false-positive STD signals. The STD spectrum of the sugar OS R1 alone does not show any signals in the STD-spectrum (figure 88).



**Figure 88. 600 MHz  $^1\text{H}$ -1D-STD-NMR spectra of OS-R1 alone in buffer A (top) and reference (bottom).**

### IX-1-1-2-1. STD-NMR enhancements provided OS R1 epitope mapping

Once STD NMR conditions are optimized, STD NMR experiments were performed on ECD MGL:OS R1 and in parallel on ECD MGL:OS R3 mixtures, at 1:100 molar ratios.

OS R3 ligand did not show any significant signals from carbohydrates in the STD difference spectrum despite its high similarity with OS R1. A peak from a potential contaminant at 2.63 ppm that was already observed when performing NMR analyses on our ligands (see chapter VI), emerged instead, in the  $^1\text{H}$  STD-NMR spectrum (figure 89).

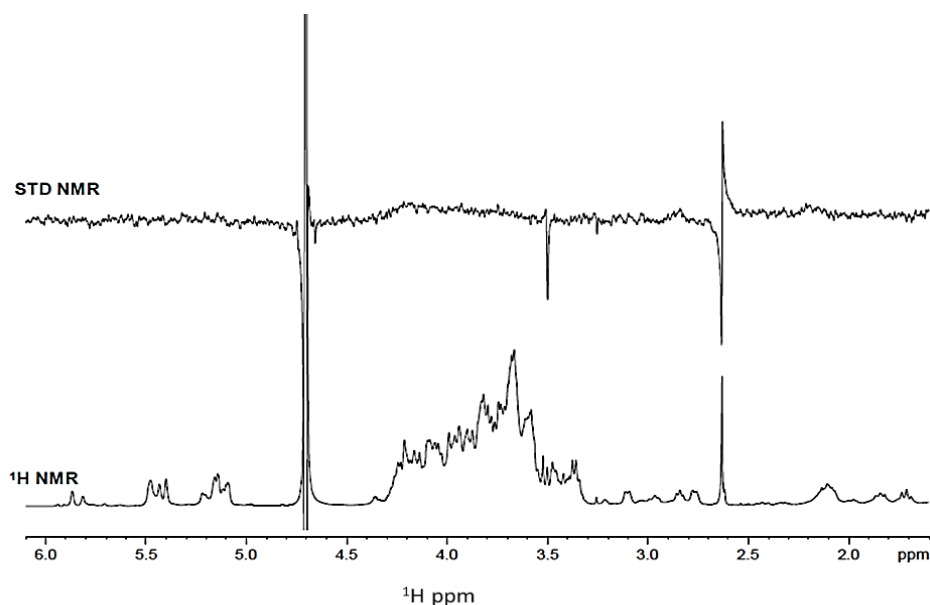
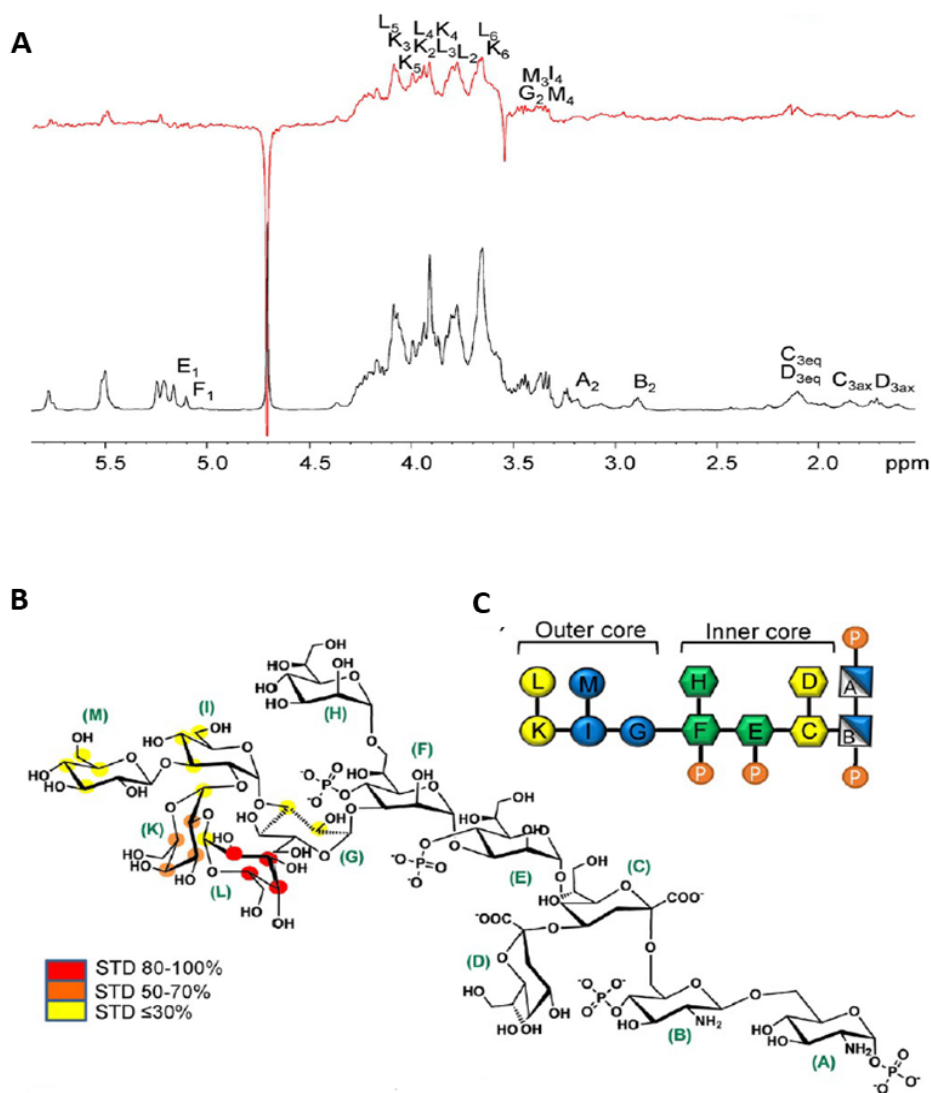


Figure 89.  $^1\text{H}$ -1D-STD-NMR spectroscopy analysis of ECD MGL: OS R3 mixture. On top STD NMR spectrum (up) of the 1:100 mixture of MGL–OS R3 is displayed and the related  $^1\text{H}$  reference spectrum is shown at bottom.

The observation of STD enhancements in the STD spectrum of the mixture of MGL: OS R1 (figure 90A) clearly revealed the protein–ligand interaction, in deep contrast to OS R3 STD results.

The great overlap between the NMR resonances of sugar ring protons of OS R1 hampered an extensive and quantitative study of the STD effects. However, analysis of signals resonating in isolated regions of the spectrum, together with the strong differences in multiplicity and relative intensities of STD signals with respect to the corresponding reference, OS R1 epitope mapping was carried out (figure 90B).



**Figure 90. STD NMR spectroscopy and epitope mapping analyses of ECD MGL:OS R1 mixture.**

A)  $^1\text{H}$ -NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:100 mixture of MGL:OSR1. Some key proton resonances are labelled. B) STD-derived epitope mapping of the MGL:OSR1 interaction, with colour coding from the highest (red) to lowest (yellow) observed STD effects. C) Schematic structure of OS R1. Phosphate groups are depicted as white letter P in orange circles.

The outer core was the moiety the most engaged in binding with the human MGL. In detail, the strongest STD effects derived from the terminal galactose units (L and K) and protons H-2–H-5 of residue L (see the discussion on the computational analysis below). This observation was consistent with the previously reported ability of the human MGL to accommodate not only GalNAc sugar, but also the galactose moiety, in its  $\text{Ca}^{2+}$  binding site<sup>354</sup>.

Saturation transfer to the glucose residues, I, G and M, adjacent to the terminal galactose disaccharide, was also detected, although to a lesser extent, thus suggesting lesser involvement in the interaction with human MGL. On the contrary, the residues of the inner core region did not seem to be directly involved in the binding interface.

We showed that the involvement of OS R1 outer core in MGL binding occurs mainly through **terminal Di-Galactose** residues that are not present in OS R3 outer core thus abolishing its interaction with MGL. This is an additional proof that OS R1 is recognized by human MGL mainly via interaction with those residues and more precisely through defined sugar OH groups which underwent STD enhancements.

### IX-1-1-2-2. Computational studies by using homology modelling and molecular docking

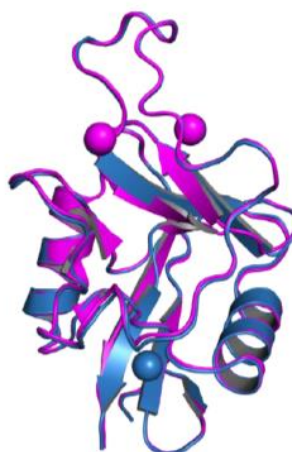
Having showed that OS R1 is recognized by human ECD MGL from 1D-STD-NMR findings, we were interested about building an MGL/OS R1 complex model. To do that in the simplest and effectiveness way, we pointed back our direction to the CRD given its shorter amino acids sequence. We investigated the possible binding modes between OS R1 and the CRD of human MGL via combining Homology Modeling and Molecular Docking studies that were established by R.E. Forgione from our group. It was possible to have a look at possible lectin residues that are involved in binding process.

#### IX-1-1-2-2-1. Human CRD MGL homology modelling

Figure 91 shows structural details regarding the homology modelling study that were established based on human Asialoglycoprotein amino acids sequence (1DV8) to make the CRD structure of human CRD MGL, the same template used to derive our predicted MGL represented in chapter VIII. The 3D CRD MGL model was successfully predicted which enabled us to continue with further analyses.



b

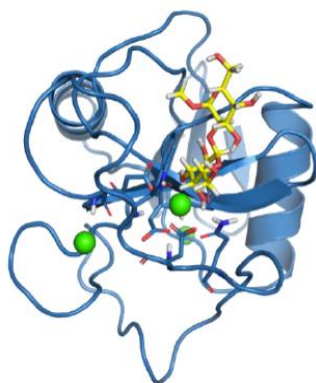


**Figure 91. Homology modelling of human MGL. a)** Multiple sequence alignment of human MGL with the homologous protein ASGR. The sequence intervals from Cys157 to Leu284 was considered as CRD MGL and as a template used for the homology modeling, namely ASGR (Cys154- Leu181) are highlighted in grey. The residues important for galactoside binding are highlighted in cyan. **b)** Superimposition of MGL-CRD homology-based model (in blue) and the X-ray crystal structure from Asialo-glycoprotein (in magenta).

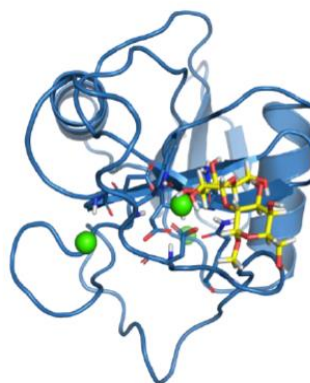
#### IX-1-1-2-2-2. Molecular Docking of OS R1 into CRD MGL

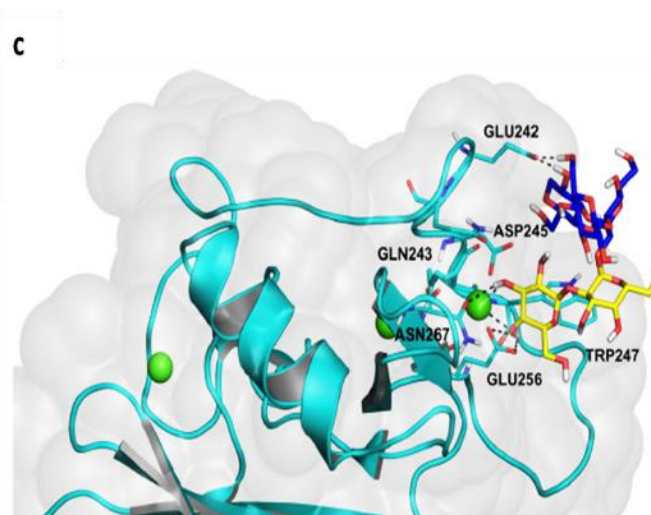
Employing computational studies has become crucial nowadays for understanding and complementing experimental data. In fact, those studies were carried out exclusively to complement the STD-NMR experiments and we had no idea about the existence of a secondary binding site on CRD MGL for *E. coli* glycoconjugates at that time. We used Docking calculations in a gradual way, starting from the Di-Galactose unit and the outer core (figure 92), so that the complete OS R1 structure could be easily docked onto the calcium binding site of CRD MGL (figure 93).

a



b



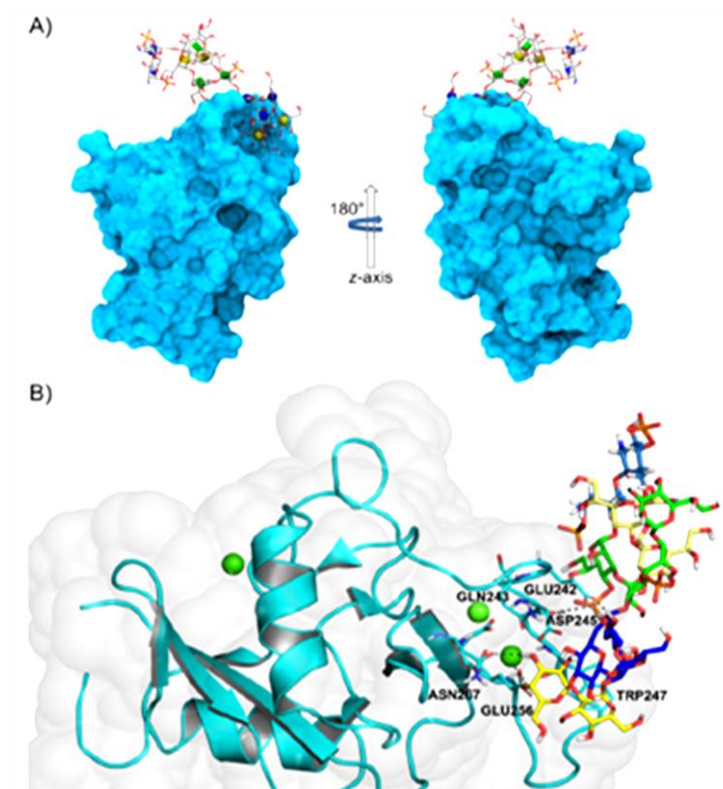


**Figure 92. The 3D model building of the MGL- outer core R1 complex.** Gradual molecular docking was done by using Autodock program, starting from the Di-Galactose unit (a and b) with two different sugar orientations, to the pentasaccharide of OS R1.

Two binding modes were observed including  $\text{Ca}^{2+}$  binding through OH-3 and OH-4 of the galactoside ring, polar interactions, and hydrogen bonding, as suggested from Di-Gal docking calculations. From the outer core docking results, we found out that the interaction mainly occurs via polar interactions, involving terminal Gal OH-3 and OH-4 and the side-chains of Asp245, Asn268, Asp269 and Glu256 residues and OH-2 of the glucose for hydrogen bonding with the carboxylate group of Glu242.

These bindings showed a high stability thanks to (i) additional stacking interactions between the aromatic system of the conserved Trp247 and (ii) steric hindrance of the penta-saccharide scaffold, thus guiding OS R1 preferences for  $\text{Ca}^{2+}$  binding and polar interactions as sole binding mode. More details are available in our publication<sup>344</sup>.





**Figure 93. A 3D model of the MGL–OSR1 complex.** A) Binding poses of the OS R1 docked into the binding pocket of the MGL starting from the 3D model of the MGL–pentasaccharide complex and based on NMR spectroscopy data. B) Specific interactions between binding-site residues and the OS are depicted by dotted lines for polar contacts.

The molecular insights obtained from STD-NMR coupled to computational studies considering  $\text{Ca}^{2+}$  coordination together with polar interactions are in complete accordance with usually observed lectin-carbohydrates binding modes. The availability of many OH groups and, in some cases N-Acetyl groups besides, in carbohydrate structures contributes to lectin affinity and selectivity towards a given sugar type.

### IX-1-2 Bio-Layer Interferometry confirmed MGL binding to the isolated LOS R1

Along the previous chapter, we initially sought to test the interactions between human MGL and *E. coli* glycoconjugates and we found out that the CRD domain potentially recognizes our ligands with a weak-binding mode. As for ECD MGL, we showed its potential in distinguishing between the two similar OSs R1 and R3 ligands by recognizing selectively terminal di-Gal moieties. Having no idea about larger ligand versions including the LOSs and LPS O157:H7 interaction with the ECD MGL construct, we tried to answer the following question: does ECD MGL conformation increases binding affinity thereby inducing the interaction with the three *E. coli* ligands?



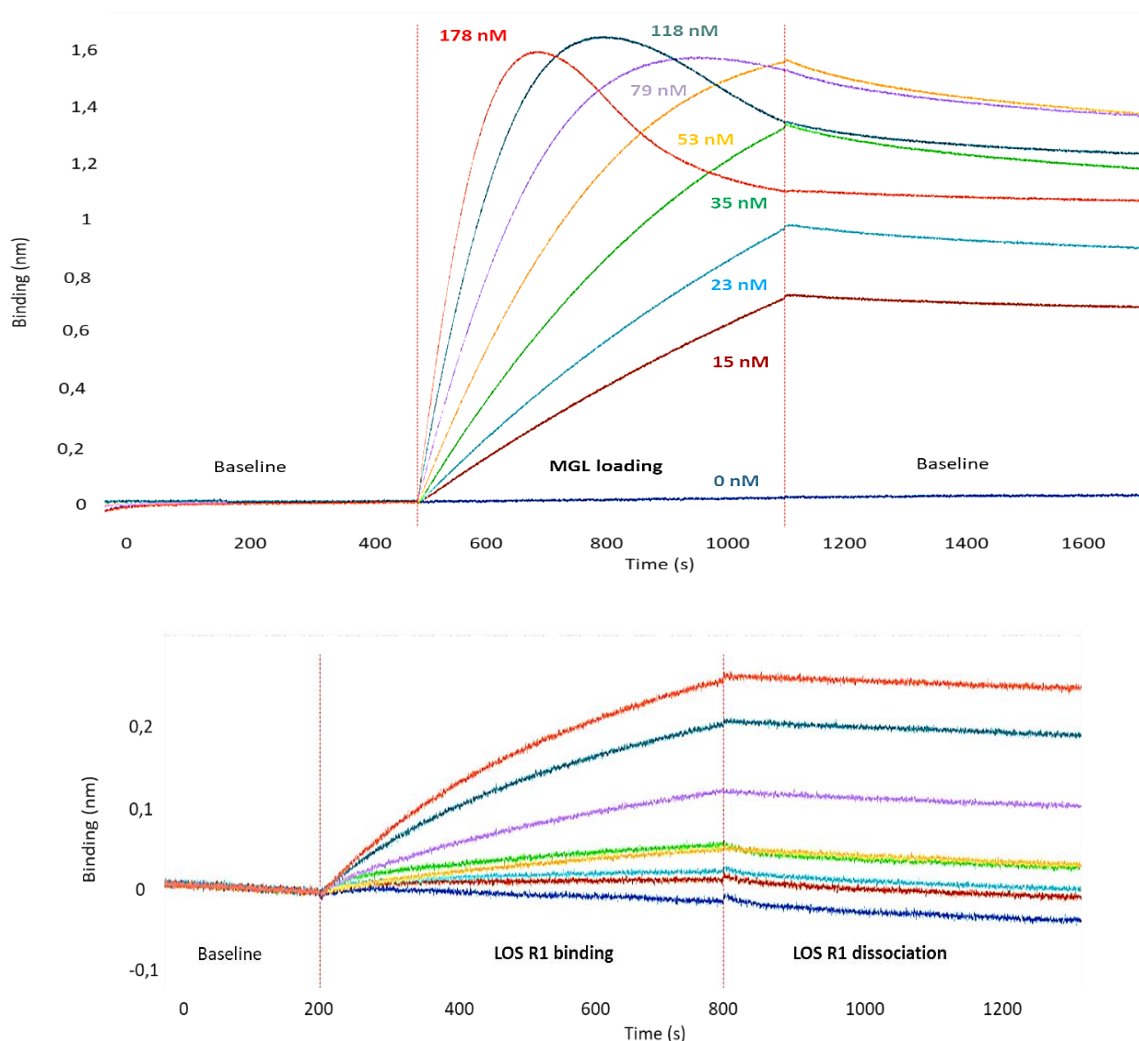
To gain deeper comprehension into *E. coli* glycoconjugates binding to human ECD MGL, we conducted direct measurements of biomolecular interactions by using Biolayer Interferometry (BLI) to evaluate the recruitment of such biomolecules by trimeric ECD MGL. Essentially, this technique measures binding kinetics from the detection of interference patterns of light changes. In fact, by exposing the interactors to white light where the lectin is fixed on a biocompatible surface (fiber-optic biosensor tips) and the ligand is flowed in the appropriate buffer. If an interaction happens during BLI measurements, this will form a layer that is thicker than the initial one coated only by lectins, thus inducing interference patterns changes or shifts<sup>361</sup>. In this respect, real-time binding data are recorded in real time.

### **IX-1-2-1 Optimisation of BLI conditions**

Conducted with the idea that the oligomeric structure of lectins is important to induce avidity-based binding to their ligands such as OS R1, as it could be suggested from our STD-NMR analyses, we used the intact ECD MGL for BLI experiments. Lectin coupling to biotin is required for its stable immobilisation on the surface of BLI biosensor tips thanks to the high affinity between biotin and streptavidin contained in those tips. Indeed, ECD MGL biotinylation was successfully performed giving satisfactory amounts of biotinylated MGL, sufficient for efficient covering of BLI tip surface and for the interactions with the ligands.

We did many optimisations on the system before starting principal BLI experiments to guarantee the appropriate binding conditions. Optimized experimental conditions would automatically provide accurate kinetics data. These include the choice of an efficient concentration of MGL for which the tips surface is not totally covered by the protein, thus avoiding steric hindrance from fully saturated surfaces and, the adequate buffer (i.e. 50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl pH8 in H<sub>2</sub>O or buffer C). Besides that, different concentrations of ligands were tested showing an optimal concentration at 1.6 μM that is the lowest median concentration (considering ligand density effects) for which binding is analyzable.

Figure 94 shows real time MGL loading at different concentrations followed by LOS R1 binding in each condition. We selected the most effective MGL final concentration to be 53 nM for which the biosensor surface is not oversaturated and for which LOS R1 binding is measurable from the sensorgram traces.



**Figure 94. BLI sensorgram traces from (top) human ECD MGL binding to streptavidin tips and (bottom) from binding to 1.6  $\mu$ M LOS R1 at 25°C in buffer C.** Colored lines represent data traces at different ECD MGL concentrations. Before LOS R1 loading, MGL charged tips were moved to a fresh buffer for the second baseline permitting the standardization of the initial binding value to zero. Binding was followed from  $t=0$  to 1200 seconds and BLI steps are mentioned for each measurement.

Moreover, we assessed the functionality of ECD MGL towards GalNAc-BSA (Bovine Serum Albumine) ligand by carrying out BLI measurements in the same conditions (data not shown). MGL binding to GalNAc-BSA was proved thus confirming the reliability of this method for ECD MGL specificity and affinity predictions.

### IX-1-2-3. Biomolecular findings estimate the affinity between MGL and LOS R1

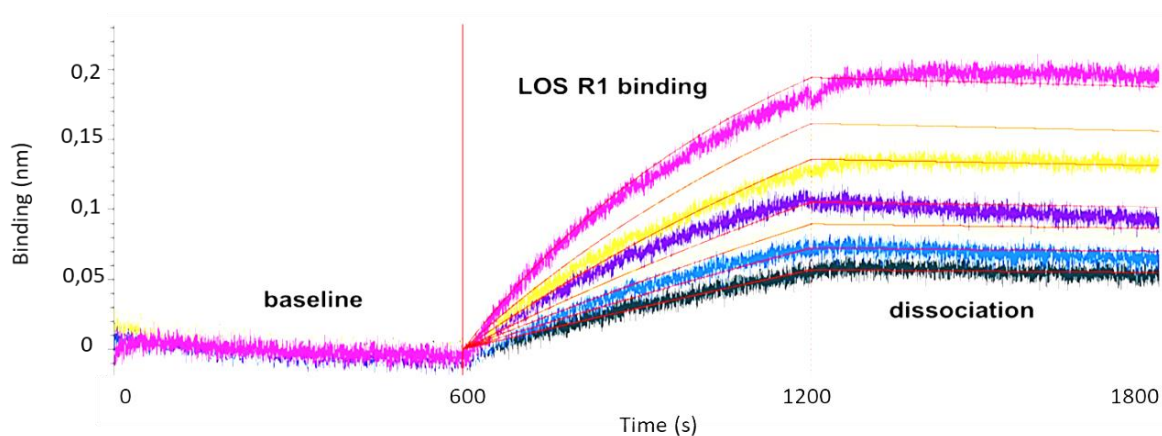
#### IX-1-2-3-1. LPS O157:H7 and LOS R3 exhibited unspecific bindings

All isolated glycoconjugates were evaluated by BLI tests in the same experimental conditions (e.g., Temperature, ECD MGL optimal concentration and buffer C). BLI sensorgrams traces that we recorded for LOS R3 and LPS O157:H7 ligands showed strong non-specific

interactions (see Annexes figure13) where both MGL charged and uncharged biosensors tips manifested unspecific binding profiles.

#### IX-1-2-3-2. The affinity of LOS R1 binding to ECD MGL is estimated

As for LOS R1, specific interaction was uniquely observed for MGL charged biosensors tips, and not for uncharged tips. In fact, 0.7 to 8  $\mu\text{M}$  final concentrations of LOS R1 were flowed in the appropriate buffer during real time BLI measurements. The choice of this range of ligand concentrations was confirmed by many essays considering a large row of concentrations (0.7-40  $\mu\text{M}$ ). Our selection was directed to lower range because this permit the interaction to be the closest to ideal curve fit by ensuring lower ligands densities on the binding surface (figure 95).



**Figure 95. BLI analyses of human ECD MGL (53 nM) binding to different concentrations of LOS R1 ligand at 25°C in buffer C and B.** Colored lines represent sensorgrams traces at different LOS R1 concentrations. Binding was followed from  $t=0$  to 1800 seconds and BLI steps are mentioned for each measurement.

We processed the apparent  $K_d$  prediction of **LOS R1** binding to human MGL ECD by applying global fitting respecting 1:1 Langmuir model through Data analysis HT 11.1.0.25<sup>®</sup> software<sup>361</sup>. Apparent dissociation constant  $K_{d_{app}}$  was estimated to be in the nanomolar range, that is **288 nM +/-6**. This strong affinity has been formerly observed for LOS *E. coli* B binding to a bacteriophage adhesin (gp37) and was suggested to be a strong binding by using the same approach together with AFM analyses (providing piconewton range of binding forces)<sup>291</sup>.

The biomolecular interactions between ECD MGL and LOS R1 was confirmed by BLI experiments. These observations are consistent with the STD-NMR analyses that we obtained from ECD MGL-OS R1. We think that the strong affinity for *E. coli* R1 glycoconjugates, that this human lectin has, is induced the avidity phenomenon driven by ECD construct organization of human MGL that provides multivalent interactions hence reinforcing and stabilizing LOS R1 (c.f. OS R1) binding.

### **IX-1-3. Disaggregation of *E. coli* LOSs vesicles in presence of ECD MGL**

Along with observing real-time MGL binding to *E. coli* R1 LOS and not to *E. coli* R3 LOS by means of binding essays, we wanted to go further with interaction analyses by evaluating the global dispersions and morphologies of both LOS specimens in presence of human MGL. This experiment would help in determining another aspect of the interaction. LPS disaggregation by human Mannose Binding Lectin (MBL) and caspase-4 protein has been experimentally shown and it is considered as LPS recognition prior to TLR4 activation in a biological context<sup>133,147</sup>. We used TEM technique to observe LOSs R1 and R3 morphological behaviours in different conditions, as we did for LPS O157:H7 (chapter VII) and we managed to get interesting observations.

#### **IX-1-3-1. *E. coli* LOS R1 and R3 morphologies in solution**

We chose to study those aspects considering the capability of this lectin to form invaginations at the bacteria membrane (chapter III, figure 51) when the whole interaction system is considered. Does ECD MGL disrupt LOS vesicles in vitro which would lead to (or correlate with) membrane curvatures in vivo? In order to visualize these biophysical effects, we assessed first free LOS morphology in solution.

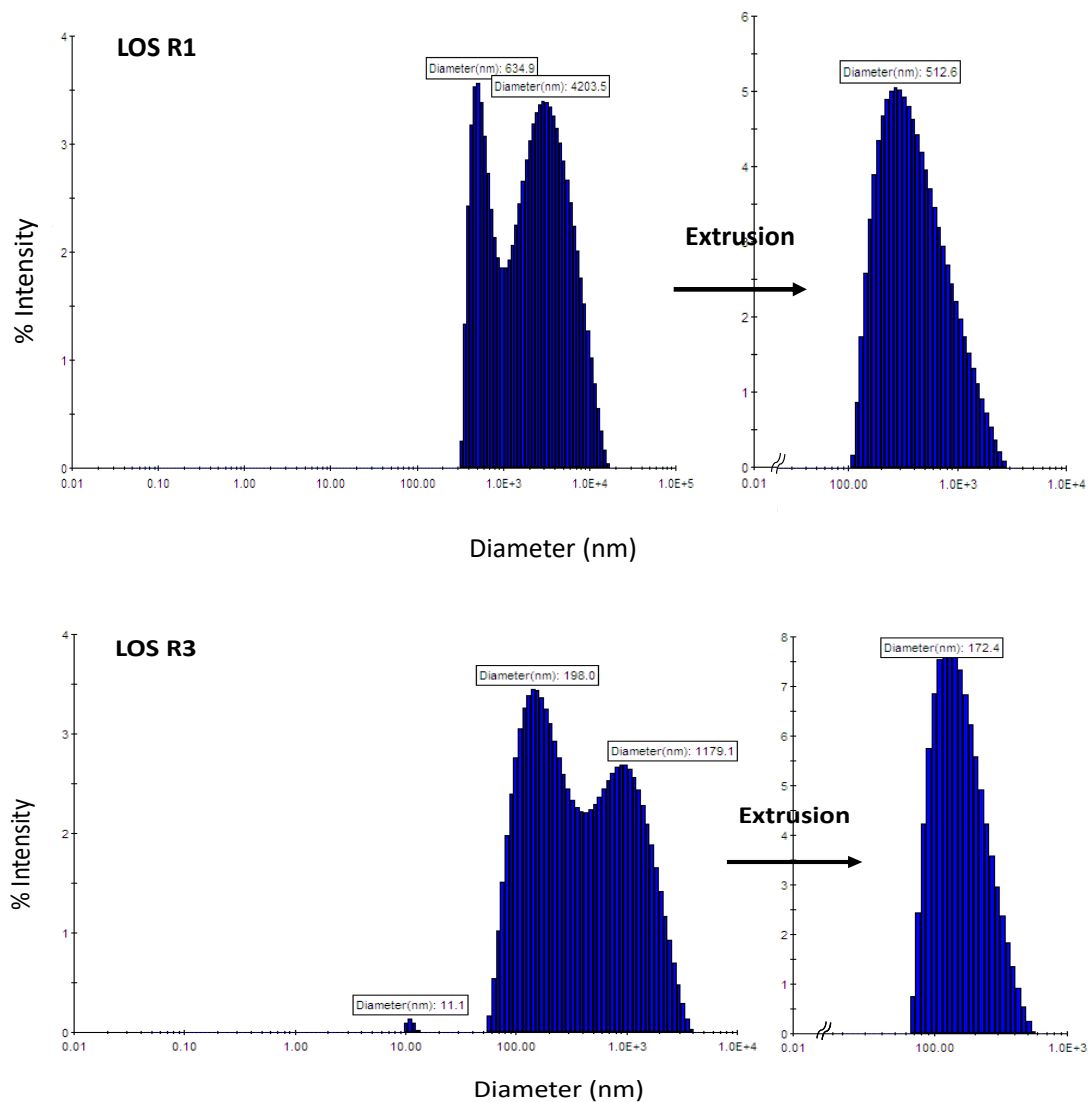
Isolated LOSs specimens tend to adopt various sized shapes in solution highlighted by big and/or small vesicles due to the presence of fatty acids (Lipid-A) and relatively few sugar residues from the core (with respect to the LPS). The employment of chloroform on DPPC liposomes induced size distribution variations from 135 to 583 nm of vesicle diameter due to partial solubilisation of fatty acids<sup>362</sup>. Hypothetically, LOS R1 and R3 would experience the same changes during their PCP extraction providing vesicles with different sizes in solution. Thus, LOS size homogenisation is required before interaction tests.

##### **IX-1-3-1-1. Preparation of homogenous LOS vesicles**

To obtain relevant MGL binding observations, we first accomplished size homogenisation of LOSs R1 and R3 specimens by an extrusion protocol adapted from Avanti® polar lipids system<sup>363</sup>. Briefly, LOSs were subjected to extrusion by passing each sample through membranes with decreasing sizes, successively exchanged, to end up with unified size, in theory. Extruded LOSs particles were assessed afterwards by Dynamic Light Scattering (DLS), a technique that we used to evaluate size distribution of extruded LOS biomolecules<sup>341</sup>. This method is based on monochromatic light (laser) scattering by the sample that yields Brownian motions in solution, and so this scattering will take on different intensities. These intensity fluctuations will be in turn

analyzed and converted to an autocorrelation function yielding particle size using the Stokes-Einstein relationship (previously mentioned in this chapter as equation x).

After being passed through a sequence of extruder membranes (0.1-0.8  $\mu\text{m}$  pore diameter), LOSs particles were subjected to DLS and showed apparent sizes changes from bimodal (200 – 4000 nm of diameter) to unimodal (~500 nm for LOS R1 and ~200 nm for LOS R3) size distributions indicative of a successful homogenisation (figure 96).



**Figure 96. DLS analyses of LOSs R1 and R3 size distributions before and after extrusion.** The size distribution is shown as bimodal peaked with the appropriate diameter sizes in nm, while after extrusion the size became monomodal for both LOSs.

#### **IX-1-4-2. LOS R1 vesicles are disaggregated by human MGL *in vitro***

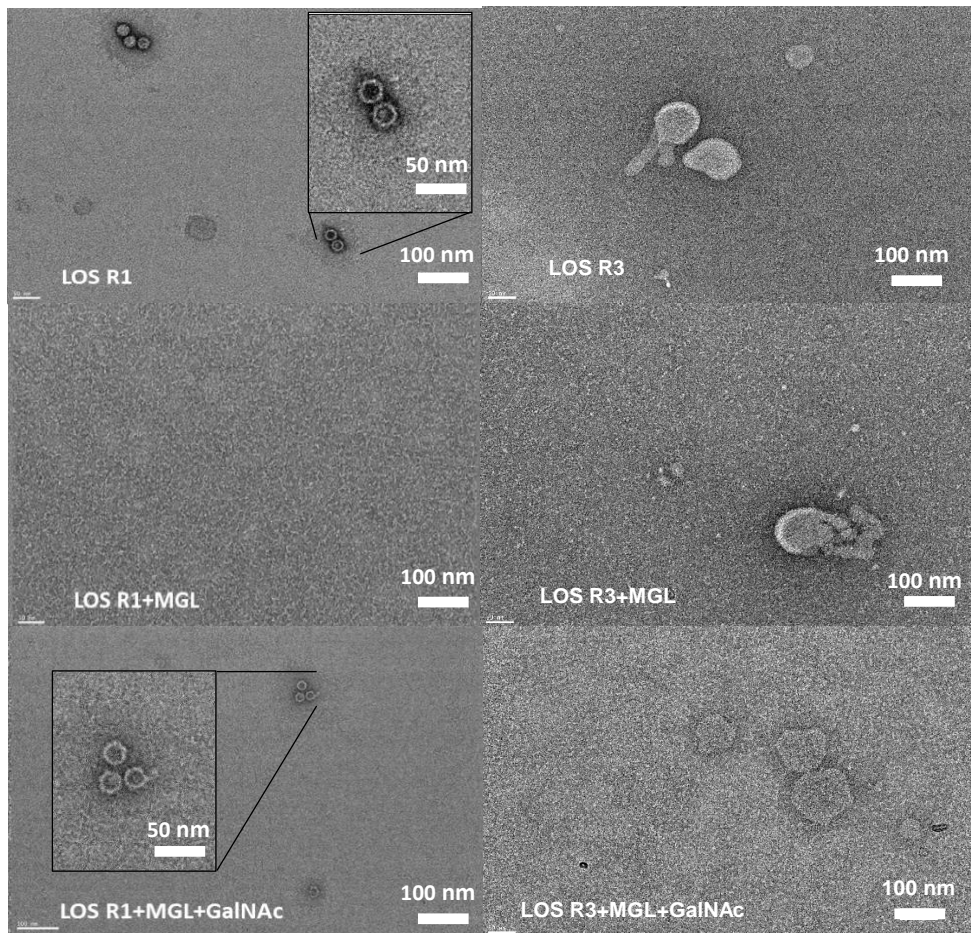
Once size distributions of LOSs R1 and R3 were standardized, we analysed the effect of ECD MGL on these LOS structures. We aimed to double verify ECD MGL effect on *E. coli* ligands

during the interaction and, to confirm (or not) our previous findings regarding its specificity to *E. coli* R1 LOS.

To do that, we incubated 1.2 mg/mL of LOS in the appropriate buffer (50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl pH8 in H<sub>2</sub>O, buffer C) and divided each preparation into two different samples of which one was dedicated to interaction test by addition of 3.5 μM final concentration of human ECD MGL. We found out that, in the absence of MGL, both LOSs formed uniform vesicles with different assemblies highlighted by distinct sizes. In contrary to what was observed through DLS measurements, the particles formed by the LOSs showed different relatively smaller sizes and controversial distributions. This could be due to the different sizes of LOS membrane-like assemblies that the LOSs may arbitrarily adopt in solution, with their tendency either to agglomerate (DLS) or to assemble in dried form (TEM). For these reasons, particles sizes are more properly construed by considering TEM results, since DLS data provide the hydrodynamic radius or diameters with focus on large agglomerates whereas TEM shows better emphasis on smaller assemblies (~50-200 nm for LOS R1 and R3 vesicles). The latter approach would be more informative about the size of LOS.

Interestingly, MGL addition strongly disaggregated LOS R1 assemblies into very small particles, even at a low lectin concentration. Conversely, it seems that MGL did not significantly affect LOS R3 vesicles organization.





**Figure 97. TEM micrographs showing human ECD MGL disaggregation effects on LOS R1.**

Complex formation of LOS R1-MGL, at 1:0.01 molar ratio, is manifested in the corresponding image by small aggregates whereas no morphological changes were observed in the case of LOS R3 in presence of the protein. TEM profiles following GalNAc addition showed significant insights for LOS R1-MGL binding. When adding 2 mM final concentration of GalNAc sugar to the LOS-MGL mixtures, disrupted LOS R1 vesicles resurfaced (figure 97). This observation results from competitive interactions of LOS R1 and GalNAc with MGL. Accordingly, we suggest that the exposition of MGL to high concentration of its favourite ligand (single GalNAc residues) disfavoured its binding to LOS R1 in order to accommodate stronger binder.

We managed to follow ECD MGL binding in many conditions for both *E. coli* LOSs using TEM approach. LOS disaggregation was observed uniquely for LOSR1-MGL complexes. Furthermore, we accomplished competition essays by evaluating the strong GalNAc interaction which allowed us to ascertain ECD MGL binding specificity and possibly the strength of the binding to LOS R1. Interestingly, our results demonstrated competing interactions between LOS R1 and GalNAc for MGL binding. This would incite us to consider the involvement of canonical binding site in LOS R1-MGL interaction, knowing that Tn-Antigens bearing GalNAc residue binds specifically to MGL

through calcium coordination<sup>203</sup>. Otherwise, ECD MGL binding to LOS R1 might be not fully dependent on direct contacts with calcium, knowing that C-type lectins classification is paradoxical for some lectins that have CTLD (CRD domain with Ca<sup>2+</sup> independent binding)<sup>161</sup> with no calcium coordination.

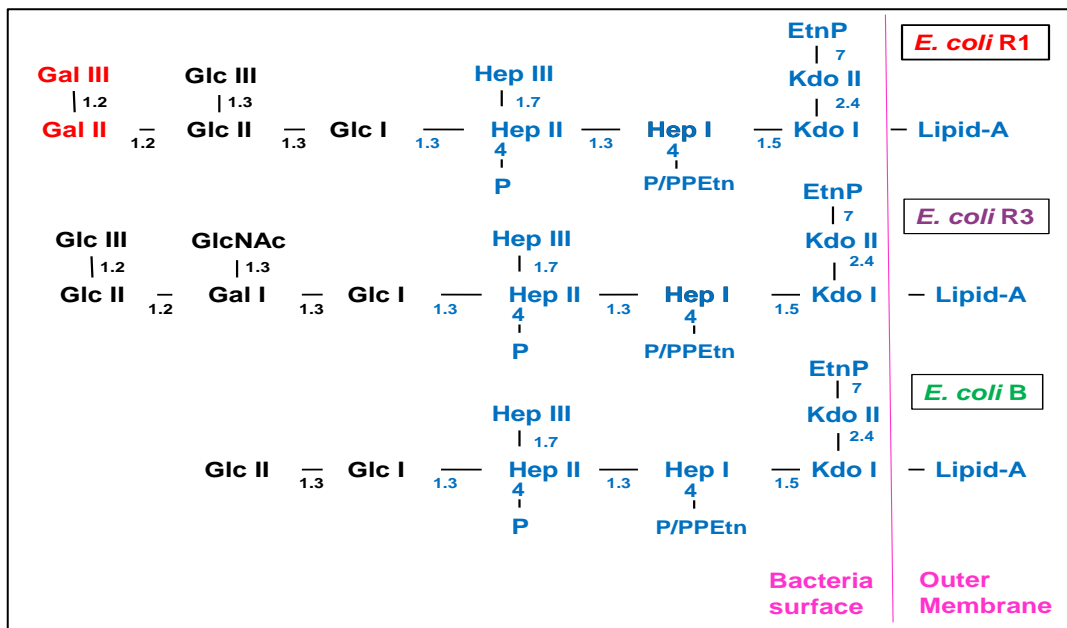
Considering the crosstalk between our STD-NMR data and TEM competition results and, knowing that MGL has higher affinity to GalNAc (micromolar range) over Gal residues (low millimolar range), we assume that ECD MGL binding potentially occurs on the terminal di-Gal of LOS R1. Fittingly, one may believe that the binding affinity and or strength between LOS R1 and ECD MGL is increased thanks to, not only the trimeric organisation of ECD MGL, but also to the multiplicity of binding prompted by the two terminal galactoses branched on the OS. Eventually, it is possible to have additional information by evaluating ECD MGL disaggregation effects on LOS R1 using competition with high concentrations of single Gal residues.

#### **IX-1-4. Human MGL tracking in *E. coli* bacterium, LOS R1 producer**

Human MGL binding specificity was predicted to be in favor of LOS R1. To further assess this specificity and validate (or not) this interaction system, we investigated ECD MGL binding to the whole bacteria. That way, we expect to obtain insights about the recognition of *E. coli* pathogens, which would be plentifully more informative and realistic than with reconstituted vesicles.

We compared human MGL interactions with *E. coli* strains namely R1, R3 and *E. coli* B strains, in different experimental conditions. The latter bacterium produces LOS with no terminal galactoses in its saccharides portion neither an outer core<sup>364</sup> (a glucose is present instead), thus presenting a negative control (figure 98). We managed (in collaboration with J.P Kleman at the cell-imaging platform) to assess human MGL strain dependency and to evaluate its LOS specificity as well, by fluorescence microscopy coupled to flow cytometry measurements.





**Figure 98. Structural differences between major LOSs exposed extracellularly by *E. coli* R1, R3 and B bacteria tested by fluorescence microscopy.** Identical LOS portions are colored in blue and terminal di-galactoses of LOS R1 are depicted in red. The scaffold of *E. coli* B LOS with a di-Glc instead of an outer core<sup>364,365</sup>. Ethanolamine phosphate (EtnP) aminoethyl diphosphate (PPEtn) and phosphate (P) are phosphate groups branched to the inner core.

#### IX-1-4-1. *E. coli* R1, R3 and B bacteria are investigated by Epifluorescence microscopy

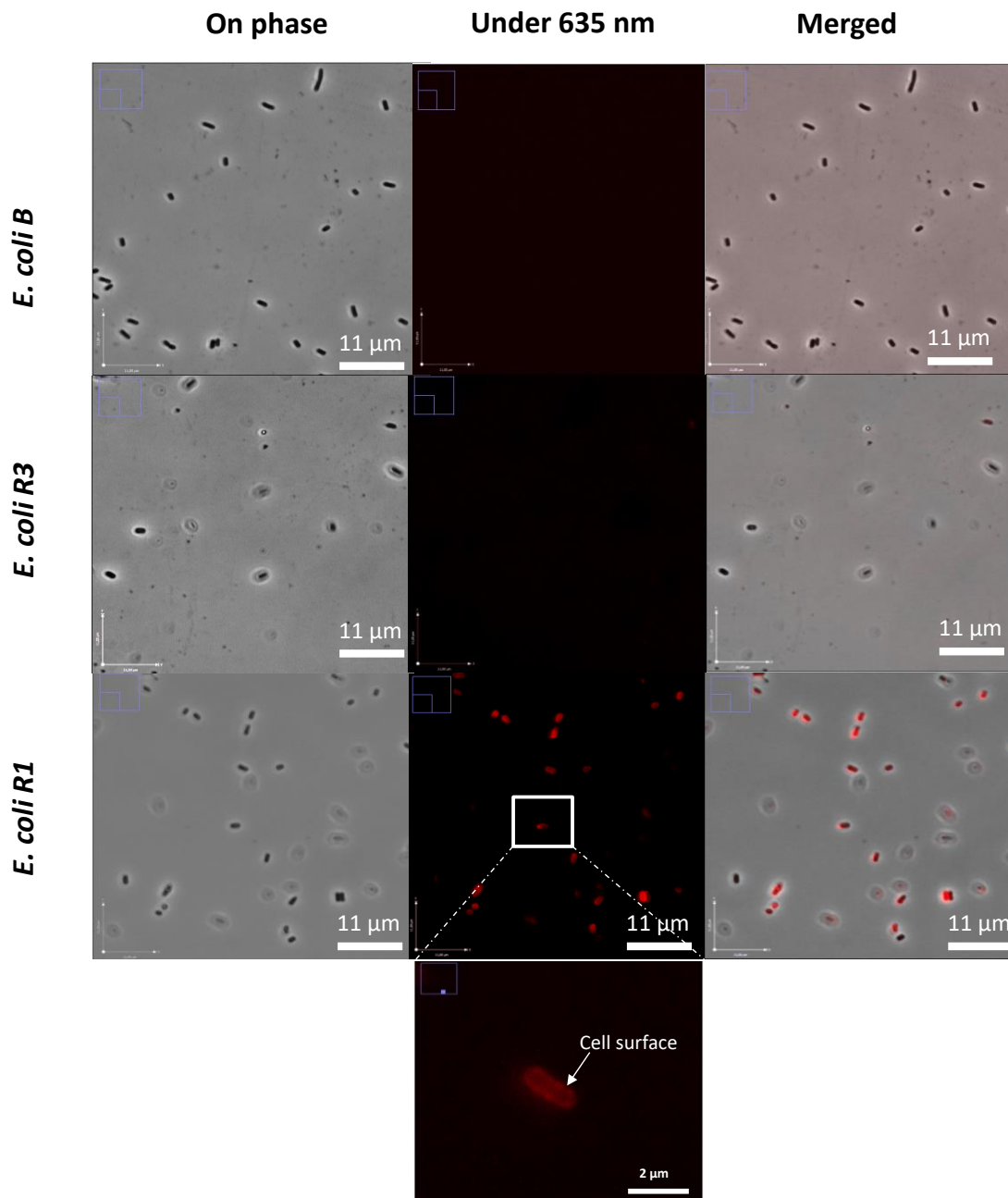
First, human ECD MGL must be labelled with a fluorophore that emits sufficiently detectable fluorescence at a specific wavelength. For that, we chose Alexa 647 to conjugate it with MGL (see chapter XI for more information). We managed to obtain reasonable quantities of fluorescently labelled ECD MGL, enough for performing several experiments.

Once the lectin was prepared for single-cell interaction tests, we monitored cells fluorescence at different stages of bacterial growth. We first grew the three strains of *E. coli* (R1, R3 and B) then incubated them with fluorescently labelled ECD MGL in the adequate conditions (chapter XI for more details). *E. coli* bacteria were afterwards visualized under the microscope at exponential and stationary growth phases. Strikingly, we observed fluorescent bacteria from some of the bacterial samples incubated with human MGL.

##### IX-1-4-1-1. Human MGL recognizes specifically *E. coli* R1 bacteria

The previous findings incited us to first optimize experimental parameters such as determining optimal lectin concentration, and sample washes steps, which enabled getting rid of insignificant brightness and decreasing fluorescence noise as well. Three ECD MGL concentrations were tested (at 1, 2 and 5  $\mu\text{M}$  as final concentrations) showing the best resolved

images at 2  $\mu\text{M}$  of MGL, otherwise, additional washing steps would be needed for higher concentrations. We selected that concentration being the lowest that provides sufficiently detectable brightness.



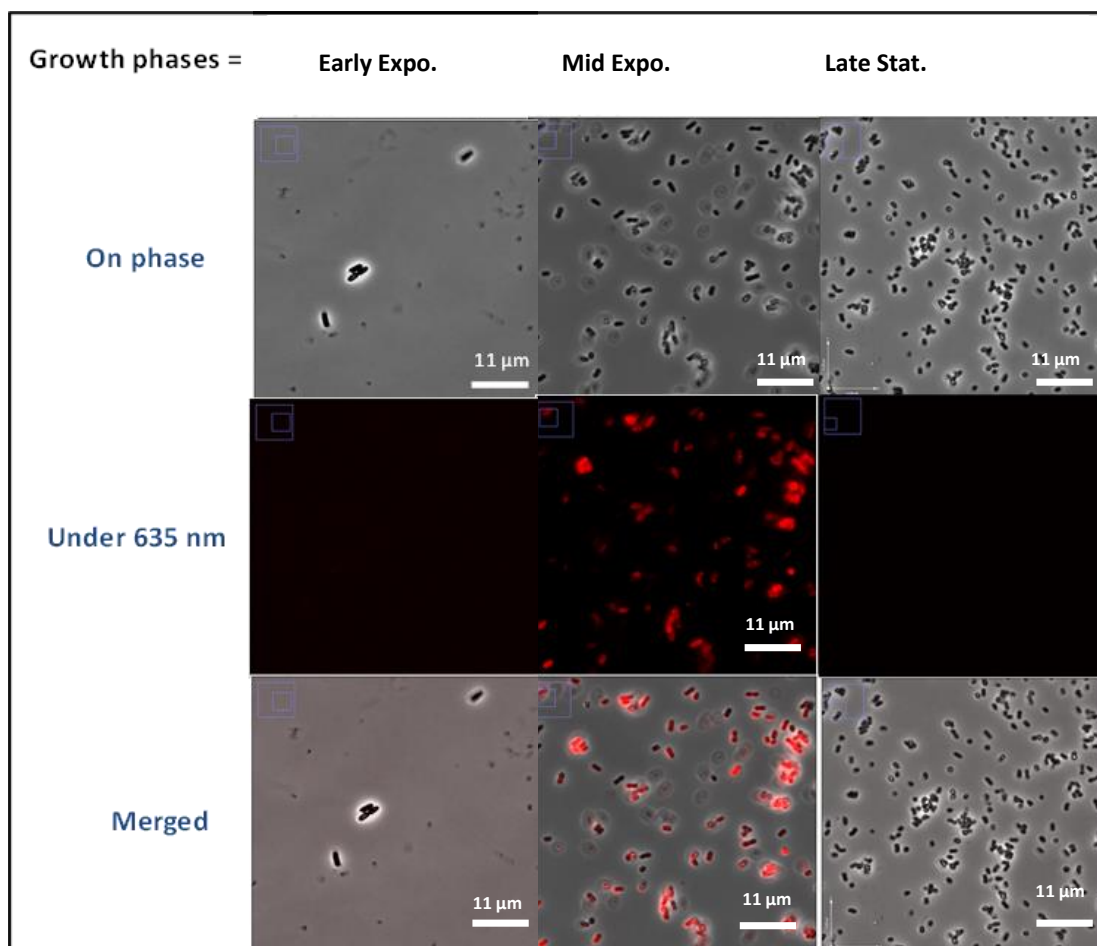
**Figure 99. Phase contrast and fluorescence microscopy images of *E. coli* R1, R3 and K12 bacteria mixed with 2  $\mu\text{M}$  ECD MGL<sup>A647</sup>. Cells are concentrated to  $19 \times 10^8$  cells/mL and incubated with fluorescently labelled human MGL in buffer C.**

The growth of *E. coli* strains during mid exponential phase (at 1.2 OD) led to the observation of fluorescent bacteria resulting from MGL binding. In deep contrast, MGL bound *E. coli* R1 bacteria were visually fluorescent while *E. coli* R3 and B cells were not (figure 99). This observation is in a

good agreement with our previous results, regarding ECD MGL specificity towards LOS R1 terminal sugars (di-Gal). Besides, this may help in understanding sugar preferences for human MGL now that this lectin binds to *E. coli* R1 and not *E. coli* R3 while both strains are homologous. Our results are in the line with a study established on parasite glycoconjugates recognition by lectins showing a strong binding from concanavalin A (ConA) and wheat germ agglutinin (WGA) lectins<sup>366</sup>.

#### IX-1-4-1-2. The recognition between MGL and *E. coli* R1 is growth phase dependent

Our results showed that ECD MGL interacts with *E. coli* R1 bacteria at mid-exponential phase of bacterial growth. To evaluate MGL bound *E. coli* R1 fluorescence at the different stages of growth including lag phase, mid-exponential phase and late stationary phase, we conducted interaction experiments in the optimal conditions described above.



**Figure 100. Microphotographs taken for *E. coli* R1 in presence of human MGL at different growth stages.** Exponential growth phases are labelled Expo., for early and mid-exponential phases; stationary phase is labelled Stat. corresponding to optical density (OD) values of 0.3, 1.2 and 4, respectively.

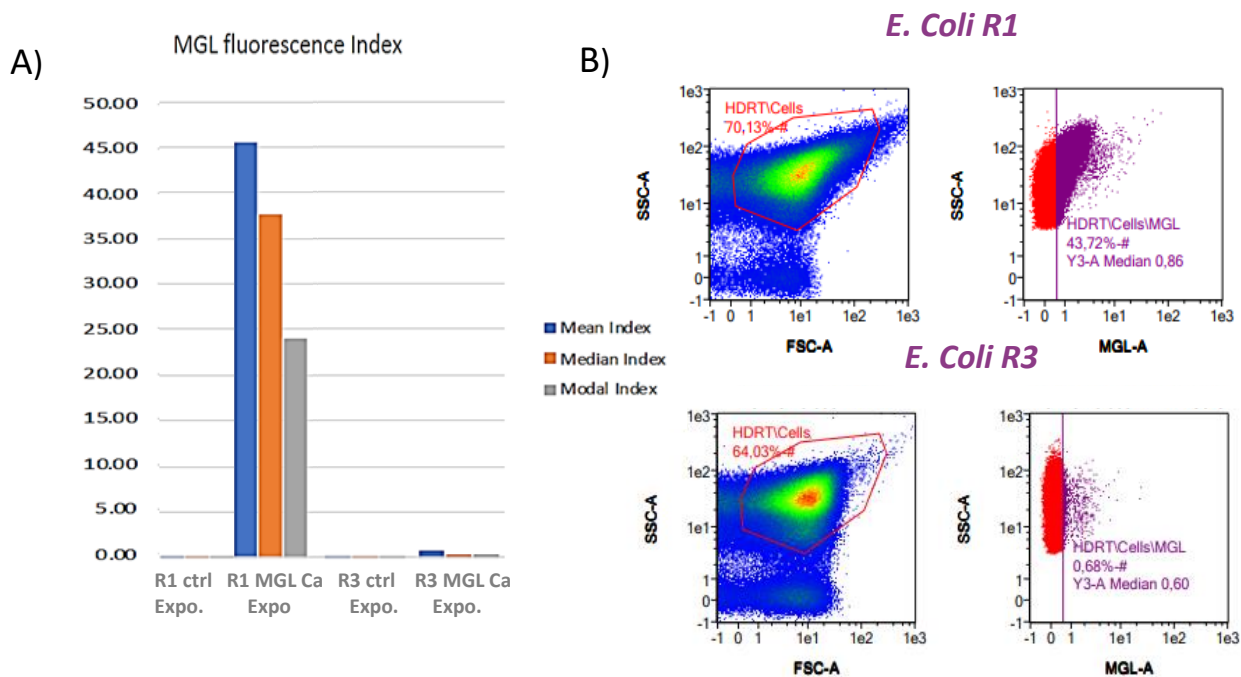
We found that *E. coli* R1 binding to ECD MGL is strictly growth phase dependent which is testified by the absence of fluorescently labelled MGL bound cells neither at early exponential phase nor at late stationary stage. Notably, MGL bound *E. coli* R1 fluorescent bacteria were observed exclusively at mid exponential phase however their fluorescence tends to be extinguished at mid to late stationary phase. This may come from cell death after being at stationary phase for long time (figure 100). When we visually analyzed microphotographs, we noticed that the morphology of cells was affected during that growth phase highlighted by porous and shrunk cells which is indicative for cell death. Still, we were not aware about if MGL binding to cell wall depends or not on cell viability. To answer that question, we carried out a first attempt to co-stain *E. coli* R1 bacteria with specific fluorophores (i.e., Syto-9 and Propidium iodide PI) that allow distinguishing live bacteria with intact cell membrane and the dead ones with degraded cell wall. We managed to register nice images with different fluorophores colors however our results were not conclusive (data not shown) because it was not possible to sensitively detect the different fluorescence emissions by using a flow cytometer equipped with only one laser for both fluorophores (i.e. A<sup>647</sup> for MGL and PI for cells). Further optimizations can be made by testing suitable combinations of fluorophores.

#### **IX-1-4-2. Quantification of MGL-bound *E. coli* R1 by using flow Cytometry**

##### **IX-1-4-2-1. *E. coli* R3 fluorescence is of zero order compared to *E. coli* R1**

To attain a high accuracy when comparing MGL bound *E. coli* R1 and R3 samples, we quantified fluorescent bacteria for samples for which fluorescence was observable (R1) or doubted (R3). By using flow cytometry, it becomes possible to obtain factual and precise quantitative data from fluorescent cells since, for instance, cells debris could be eliminated from cells sub-selection when treating data (HDRT/cells, figure 101B). As introduced in chapter III-2, this technique allows the detection of fluorescent populations and their quantification.

Interestingly, flow cytometry data confirmed fluorescence microscopy analyses. MGL fluorescent populations counting yielded up to 40 % of MGL bound *E. coli* R1 bacteria whereas it exhibited less than 1 % of MGL bound *E. coli* R3. Fluorescent populations dispersion was observed only when bacteria interacted with ECD MGL and not for control samples (figure 101A).

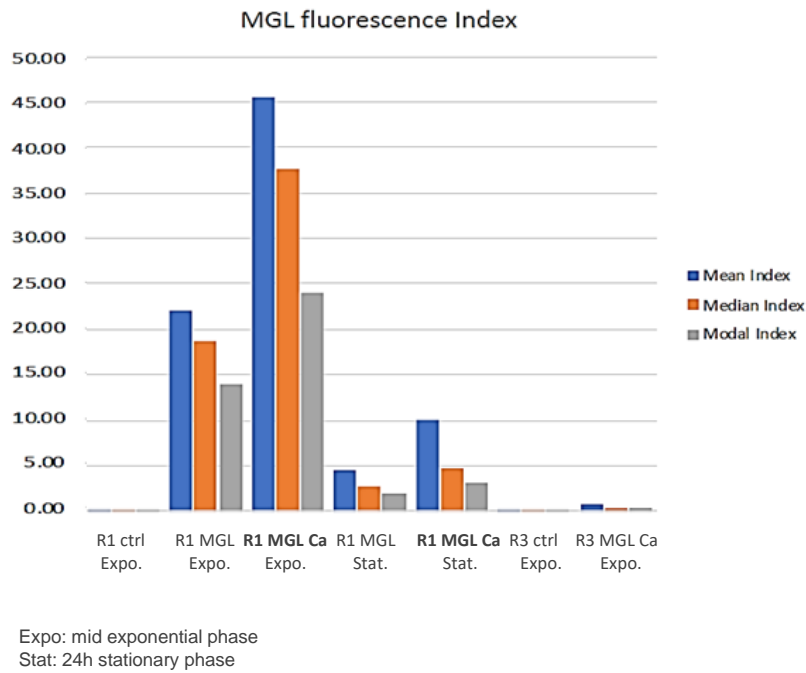


**Figure 101. Flow cytometry analyses for fluorescent *E. coli* R1 and *E. coli* R3 quantification.** A) MGL-bound cells Fluorescence index showing different estimations by comparing both *E. coli* strains. B) dot plot distribution of *E. coli* R1 and R3 populations at a concentration of  $9 \times 10^7$  cells/mL. FSC: Forward scattering; SSC: Side scattering.

A similar in vitro study about plant lectin WGA binding to cell-surface carbohydrates has been conducted by Hendrickson and co-workers via flow cytometry complemented by fluorescence microscopy<sup>367</sup>. Our findings are consistent with that work where several bacterial strains were investigated for their recognition by lectins, showing distinct fluorescence intensities resulting from singular carbohydrates interactions respecting strain recognition dependency. Another lectin called peanut agglutinin (PNA), specific for terminal galactoses, was evaluated by flow cytometry measurements on parasites *Perkinsus marinus*, and its specificity was confirmed for cell surface glycoconjugates.

#### IX-1-4-2-2. The addition of calcium ions increases MGL-bound *E. coli* R1 fluorescence

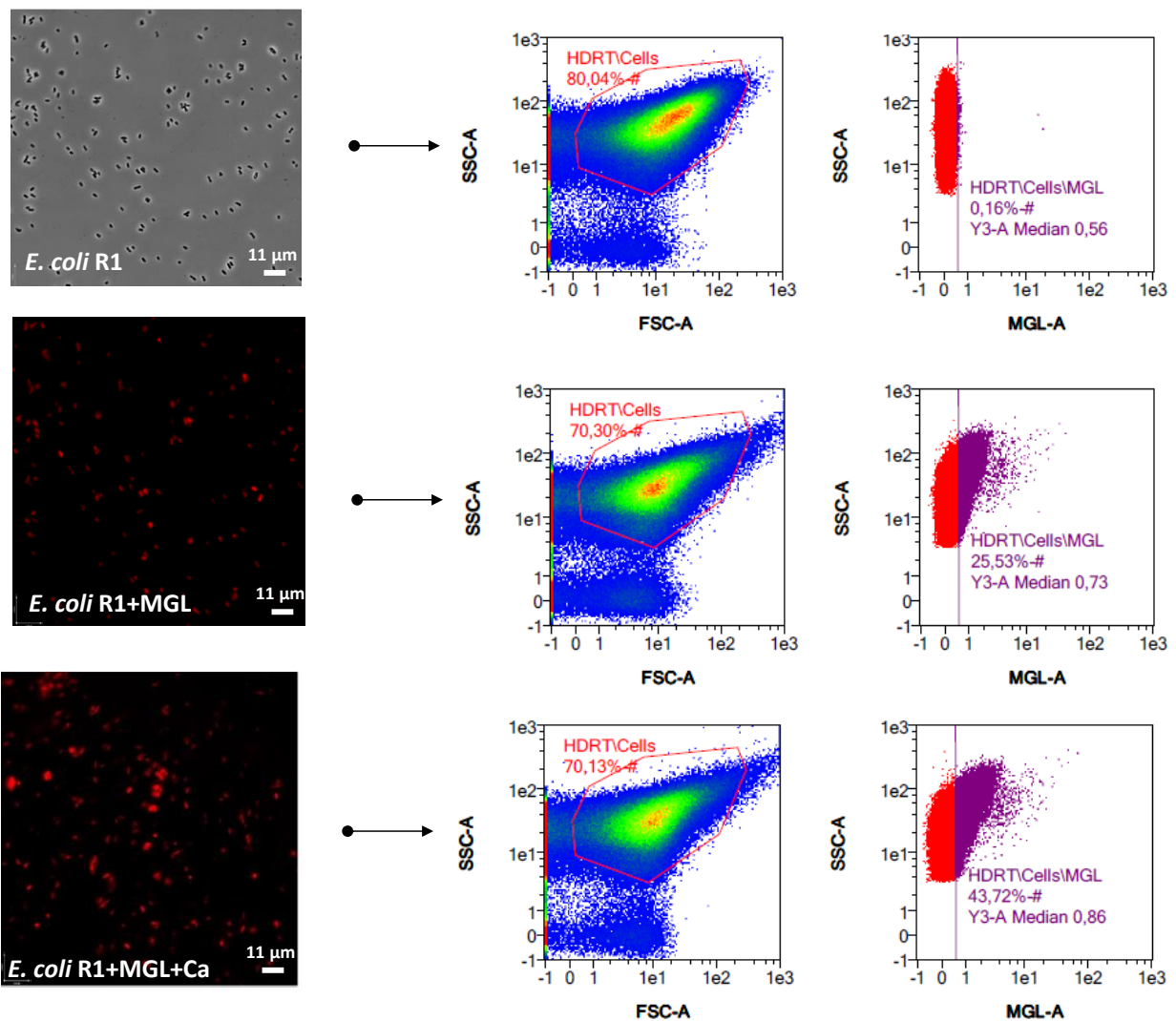
Supplementing the interaction buffer with additional calcium ions, besides the ones that are naturally present in MGL structure and *E. coli* cell wall, would aid in understanding its implication in MGL binding to *E. coli* R1 bacteria. To our knowledge, this is the first preliminary single-cell study that evaluates effects of calcium addition on the fluorescence of MGL-bound *E. coli* R1 bacteria. Strikingly, their incubation with fluorescent human MGL in buffer supplemented with 2 mM final concentration of calcium produced to two times higher fluorescence than without calcium (figure 102).



**Figure 102. Growth phase dependency and Ca<sup>2+</sup> addition effect on human MGL bound *E. coli* R1.**

This metal binds to the cell wall of *E. coli* bacteria, and more precisely, Ca<sup>2+</sup> binding sites are located on LPS phosphate and carboxylate groups<sup>93</sup>. Calcium (c.f. Mg<sup>2+</sup>) has a crucial role in rigidifying the cell wall of *E. coli* against AMPs either by binding native LPS or model LPS layer<sup>368,369</sup>. By considering that calcium presence contributes to cell rigidity, one may think that calcium addition strengthens human MGL binding to *E. coli* R1 by tightening LOS biomolecules on the outer membrane thus favouring more multivalent interactions.

It seems that calcium addition increases MGL-bound *E. coli* R1 fluorescence. Figure 103 shows real time microscopic analyses of *E. coli* R1 bacteria that became fluorescent only following the interaction, and, when calcium was added to the cell suspension, human MGL bound *E. coli* R1 exhibited the highest fluorescence, as confirmed by dot plot analyses of those populations.



**Figure 103. Calcium addition intensifies MGL-bound *E. coli* R1 fluorescence.** Left panel: Micrographs of *E. coli* R1 on phase-contrast (a) and under 635 nm (b and c). Right panel: flow cytometry measurements corresponding to each condition.

## IX-2. Conclusions

Studying the large interaction system between human MGL and *E. coli* pathogens allowed us to better understand this lectin's fine selectivity for *E. coli* R1 glycoconjugates. The binding was suggested to take place mainly at the outer core of OS R1 as concluded from STD NMR data together with molecular docking. These data supported molecular interactions between terminal di-galactoses of the OS and MGL according to epitope mapping results. Additionally, we succeeded to estimate the binding affinity to be high with a  $K_d$  of ~300 nM from our real-time BLI measurements. Fittingly, an avidity-based mechanism is suggested for MGL ECD binding to *E. coli* R1 glycoconjugates. Those outcomes are consistent with typical multivalent lectin-carbohydrates binding.

Moreover, we proved, by means of TEM, that human MGL strongly affects LOS R1 vesicles assemblies and we furthermore evaluated binding specificity by testing LOS R1 competition with single GalNAc residues for assessing the specificity of the binding and estimating canonical binding site involvement. Calcium sites could be numerous on CRD structures and they usually possess double roles including structural stability and ligand coordination. Consequently, the contribution of  $\text{Ca}^{2+}$  ions is hard to construe accurately at this stage because of MGL structure susceptibility towards interventions on calcium ions.

Human MGL specificity and selectivity to *E. coli* R1 glycoconjugates was again confirmed by fluorescence microscopy data and the interaction was quantified using flow cytometry. By using single-cell experiments on fluorescently labelled human ECD MGL and the three *E. coli* strains, we managed to evaluate MGL binding at the bacterial cell wall. Again, despite that we observed an increased fluorescence after adding  $\text{Ca}^{2+}$  ions to the interaction media, its contribution to the binding is still not conclusive and needs to be further investigated.

Taken together, our results led to a clear vision about ECD MGL specificity towards native glycoconjugate LOS R1 relative to its homologous LOS R3, at different levels including molecular, morphological, and cellular aspects. Those promising results could open new perspectives regarding pathogens recognition by human MGL. The next chapter is meant to summarize and conclude about the totality of our scientific outcomes, considering both MGL constructs (CRD and ECD). Various suggestions and perspectives are also presented in the next part of the manuscript.



## Published article



# Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of *Escherichia coli* Lipooligosaccharide

Meriem Maalej,<sup>[a, b]</sup> Rosa Ester Forgione,<sup>[a]</sup> Roberta Marchetti,<sup>\*[a]</sup> François Bulteau,<sup>[b]</sup> Michel Thépaut,<sup>[b]</sup> Rosa Lanzetta,<sup>[a]</sup> Cedric Laguri,<sup>[b]</sup> Jean-Pierre Simorre,<sup>[b]</sup> Franck Fieschi,<sup>[b]</sup> Antonio Molinaro,<sup>[a]</sup> and Alba Silipo<sup>[a]</sup>

Carbohydrate–lectin interactions intervene in and mediate most biological processes, including a crucial modulation of immune responses to pathogens. Despite growing interest in investigating the association between host receptor lectins and exogenous glycan ligands, the molecular mechanisms underlying bacterial recognition by human lectins are still not fully understood. Herein, a novel molecular interaction between the human macrophage galactose-type lectin (MGL) and the lipooligosaccharide (LOS) of *Escherichia coli* strain R1 is described. Saturation transfer difference NMR spectroscopy analysis, supported by computational studies, demonstrated that MGL bound to the purified deacylated LOS<sub>R1</sub> mainly through recognition of its outer core and established crucial interactions with the terminal Gal $\alpha$ (1,2)Gal epitope. These results assess the ability of MGL to recognise glycan moieties exposed on Gram-negative bacterial surfaces.

Bacterial cell surfaces are decorated with highly diverse glycoconjugates, in the form of capsular polysaccharides, peptidoglycans, lipopolysaccharides (LPSs) and other glycolipids,<sup>[1]</sup> which perform several functions, ranging from structural to protective roles.<sup>[2]</sup> Bacterial glycans take part in many key biological events, including pathogen recognition, receptor activation, cell adhesion and signal transduction. Additionally, these structures often serve as molecular patterns that are recognised by specific glycan-binding receptors of the host immune system, thus triggering a pathogen-specific immune response.

It is well known that LPSs, the major constituents of the outer membrane of Gram-negative bacteria,<sup>[3]</sup> are one of the main virulence factors of several feared bacterial strains, including enteropathogenic *Escherichia coli*, which is implicated in severe foodborne and urinary tract infections.<sup>[4]</sup> From a structural point of view, LPS is composed of three structural motifs that can be distinguished because they are encoded by different gene clusters. Lipid A, which represents the glycolipid portion, is an acylated bis-phosphorylated glucosamine disaccharide that anchors the LPS to the outer membrane. Lipid A is covalently linked to a core oligosaccharide (OS) that can be further divided into two different portions: the more conserved inner region, which is characterised by the presence of peculiar sugar residues, such as 3-deoxy-D-manno-oct-2-ulopyranosonic acid (KDO), and the more variable outer core. Finally, the O-antigen, which is a polysaccharide composed of several OS repeating units, extends to the extracellular medium and acts as a hydrophilic coating surface.<sup>[3,5]</sup> However, Gram-negative bacteria can also produce rough-type LPS, namely, lipooligosaccharide (LOS); a truncated version of LPS that lacks the O-antigen.

To date, the receptor complex formed by the toll-like receptor 4 and the small secreted MD2 protein is among the main species involved in bacterial LPS recognition by host immune cells.<sup>[6]</sup> More recently, it has been shown that LPSs are also intracellularly detected by caspases.<sup>[7]</sup> In both cases, the glycolipid part seems to be heavily involved in binding and in the follow-up of signalling.

Given their ability to selectively recognise carbohydrate structures, human lectins have emerged as potential LPS receptors specifically for their carbohydrate moiety.<sup>[8]</sup> Lectins are ubiquitous oligomeric proteins that have multiple roles in cell–cell communication, cellular trafficking and regulation of immune cell functions; therefore, representing validated therapeutic targets.<sup>[9,10]</sup> Within the many different families of lectins described in humans, to date, the C-type lectin (CTL) class is the biggest and most diverse, including both transmembrane and soluble receptors. Host immune cells, such as dendritic cells and macrophages, are known to express a wide range of CTLs that contain at least one carbohydrate recognition domain (CRD), which is responsible, in most cases, for the recognition of glycan structures in a Ca<sup>2+</sup>-dependent manner. In the last decade, the use of glycan arrays has provided valuable details on the ligand specificity of different human CTLs; thus highlighting their ability to recognise carbohydrate epitopes

[a] M. Maalej,<sup>+</sup> R. E. Forgione,<sup>+</sup> Dr. R. Marchetti, Prof. R. Lanzetta, Prof. A. Molinaro, Prof. A. Silipo  
Department of Chemical Sciences, University of Naples Federico II  
via Cintia 4, 80126 Napoli (Italy)  
E-mail: roberta.marchetti@unina.it

[b] M. Maalej,<sup>+</sup> F. Bulteau, Dr. M. Thépaut, Dr. C. Laguri, Prof. J.-P. Simorre, Prof. F. Fieschi  
CNRS, CEA, Institut de Biologie Structurale  
Université Grenoble Alpes  
71 avenue des Martyrs, 38000 Grenoble (France)

[<sup>+</sup>] These authors contributed equally to this work.

Supporting information and the ORCID identification numbers for the authors of this article can be found under <https://doi.org/10.1002/cbic.201900087>: experimental procedures and additional data, including detailed NMR spectroscopy analysis of the core oligosaccharide structure.

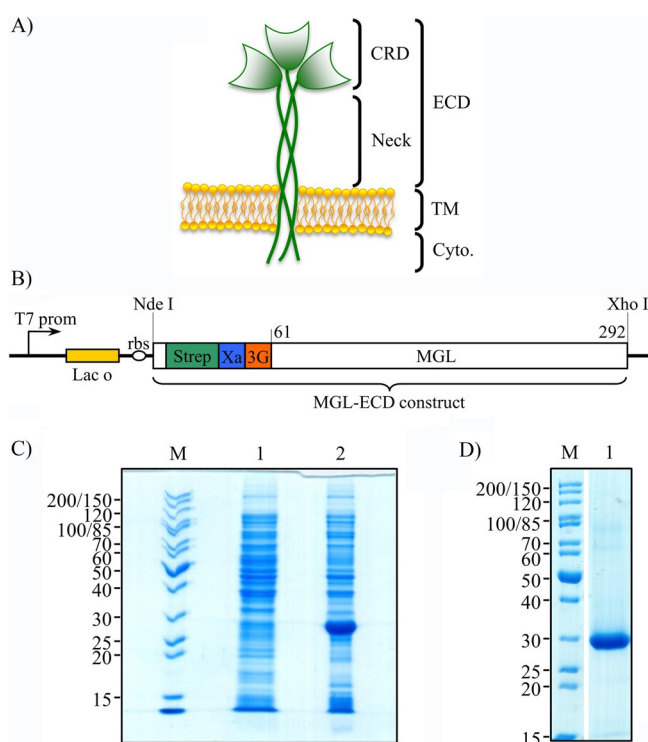
This article is part of the young researchers' issue ChemBioTalents. To view the complete issue, visit <http://chembiochem.org/chembiotalents>

belonging to both mammalian and microbial glycans.<sup>[11]</sup> However, although it is acknowledged that the recognition of pathogen-associated glycostructures by human CTLs represents a key point of host immune defence,<sup>[9,12,13]</sup> more insights into the molecular basis of the binding mechanisms between these pattern recognition receptors and their bacterial ligands, including the LPS, are still needed.

Within this framework, we have set up a proof of concept study and investigated the ability of the macrophage galactose-type lectin (MGL; also referred to as CLEC10 A or CD301),<sup>[14]</sup> to recognise the carbohydrate moiety of bacterial LPS. In contrast to several other CTLs, including dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN) and the macrophage mannose (MR) receptors, which exhibit a preference for mannose-type sugars, human MGL is characterised by an exquisite specificity for terminal GalNAc/Gal residues; thus making it potentially an attractive target for the galactose-containing LOS structure of *E. coli* R1.<sup>[15]</sup>

Core OS recognition by MGL has been investigated by means of saturation transfer difference (STD) NMR spectroscopy analysis combined with computational studies.

Herein, we used the sequence of isoform 2 of human MGL, which is shorter than that of MGL isoform-1 (also called DC-ASGPR), for cloning and recombinant production. Initially, two groups cloned these two MGL isoforms, firstly from macrophages (this CTL was therefore named macrophage galactose lectin) and then from DCs, leading to the name DC-ASGPR due to similarity to liver ASGPR with regard to selectivity.<sup>[14,16]</sup> DC-ASGPR is a longer variant of MGL that results from alternative splicing with an insertion of 27 amino acids and a deletion of 3 amino acids in the neck domain. Both MGL and DC-ASGPR possess identical CRDs, which suggests that there is no difference in their carbohydrate recognition specificities, mainly characterised on MGL up to now. The MGL extracellular domain (MGL-ECD; Figure 1 A, B) has been cloned and overexpressed at high levels in *E. coli* as inclusion bodies (Figure 1 C). Inclusion bodies were isolated and refolded as described herein. Folded functional MGLs have an oligomeric state that allows tight binding, through an avidity-based mechanism, on



**Figure 1.** Cloning and purification of MGL-ECD. A) Domain organisation of MGL. TM: transmembrane domain. Cyto.: cytoplasmic domain. B) Construct for the overproduction of MGL-ECD. Strep: StrepTag II. Xa: factor Xa protease cleavage site. 3G: tri-glycine linker. C) SDS-PAGE analysis of MGL-ECD overexpression. Lane M: PageRuler unstained protein ladder (Fermentas); lane 1: total proteins before induction; lane 2: total proteins after induction. A band at about 28 kDa corresponding to MGL-ECD MW is overexpressed. D) SDS-PAGE analysis of purified MGL-ECD. Lane 1: A unique band at about 28 kDa corresponding to pure MGL-ECD is observed.

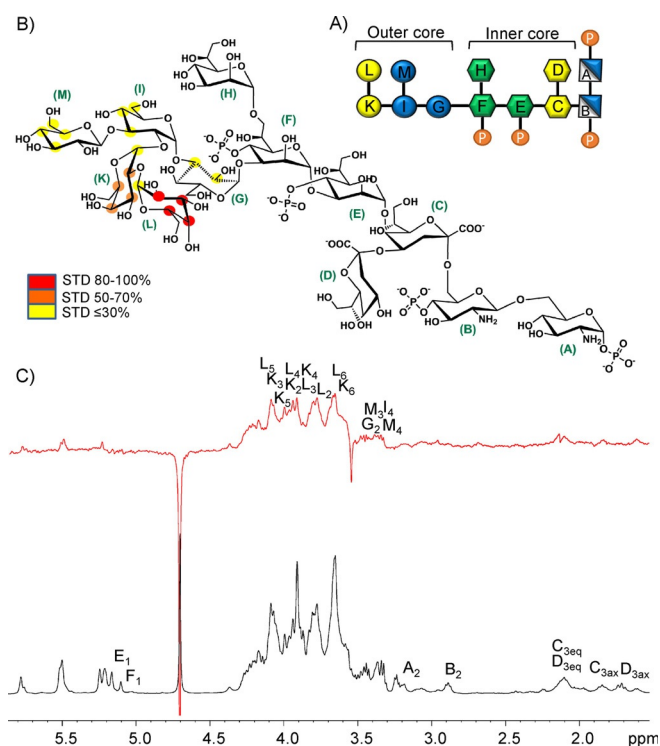
a GalNAc-agarose column.<sup>[17]</sup> Elution is performed thanks to the use of ethylenediaminetetraacetic acid (EDTA), thus demonstrating that binding is  $\text{Ca}^{2+}$  dependent. This first affinity purification step ensures the correct sorting of functionally folded MGL-ECDs, which are reloaded with  $\text{Ca}^{2+}$  in a second purification step of size-exclusion chromatography. MGL-ECD oligomers, reported as trimers,<sup>[17]</sup> are then pure and ready to use for biophysical studies (Figure 1 D).

Firstly, the LOS of *E. coli* R1 ( $\text{OS}_{\text{R1}}$ ) was extracted, purified and de-acylated to derive a soluble OS, which was, in turn, further characterised by means of NMR spectroscopy (Figure 2 A, B and see the Supporting Information for more details). Once the NMR resonances were assigned, the ability of the MGL-ECD to recognise and bind the core OS was assessed by means of STD NMR; a method that allows the measurement of the transfer of saturation from the receptor to the ligand, and thus, provides a reliable mapping of the ligand epitope in the bound state.<sup>[18,19]</sup>

The observation of STD enhancements in the STD spectrum of the mixture of MGL-ECD/ $\text{OS}_{\text{R1}}$  clearly revealed the protein–ligand interaction (Figure 2 B, C). The great overlap between the NMR resonances of sugar protons hampered an extensive and quantitative study of the STD effects. However, analysis of signals resonating in isolated regions of the spectrum, togeth-

Roberta Marchetti studied chemistry at the University of Naples “Federico II”. In 2013, she received her PhD on NMR studies of molecular recognition processes. Since 2016, she has been a researcher at the Department of Chemical Sciences, University of Naples “Federico II”. She is specialised in organic chemistry, with a special focus on the isolation, purification and characterisation of bacterial cell wall glycoconjugates. Her research interests are mainly focused on the comprehension of glycoconjugate–protein interactions through the application of NMR spectroscopy methods.





**Figure 2.** STD NMR spectroscopy analysis of the MGL–OS<sub>R1</sub> mixture. A) Schematic structure of OS<sub>R1</sub> derived from the LOS of *E. coli* R1. Symbol nomenclature for glycans has been used.<sup>[23]</sup> Phosphate groups are depicted as white letter P in orange circles. B) STD-derived epitope mapping of the MGL:OS<sub>R1</sub> interaction, with colour coding from the highest (red) to lowest (yellow) observed STD effects. C) <sup>1</sup>H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the 1:100 mixture of MGL–OS<sub>R1</sub>. Some key proton resonances are labelled. STD NMR spectroscopy analysis indicated that the inner core pointed farther from the surface of the MGL protein, and was mainly involved in the interaction with the terminal region of the OS.

er with the strong differences in multiplicity and relative intensities of STD signals with respect to the corresponding reference, suggested that the outer core was the moiety more engaged in binding with the lectin. Notably, several STD signals belonged to the outer core region. In detail, the strongest STD effects derived from the terminal galactose units (L and K) and, in particular, protons H-2–H-5 of residue L (see the discussion on the computational analysis below). This observation was consistent with the previously reported ability of the human MGL to accommodate not only the *N*-acetylgalactosamine, but also the galactose moiety, in its binding site.<sup>[20]</sup> Saturation transfer to the glucose residues, I, G and M, adjacent to the terminal galactose disaccharide, was also detected, although to a lesser extent; thus suggesting lesser involvement in the interaction with MGL. On the contrary, the residues of the inner core region did not seem to be involved in the binding interface, as suggested, for instance, by the absence of STD effects for protons at position 3 of KDO moieties and those of protons at position 2 of the two glucosamine units of lipid A. To further demonstrate the pivotal role of the terminal galactose moiety in the recognition and interaction process, we investigated the binding of MGL with the deacylated core OS

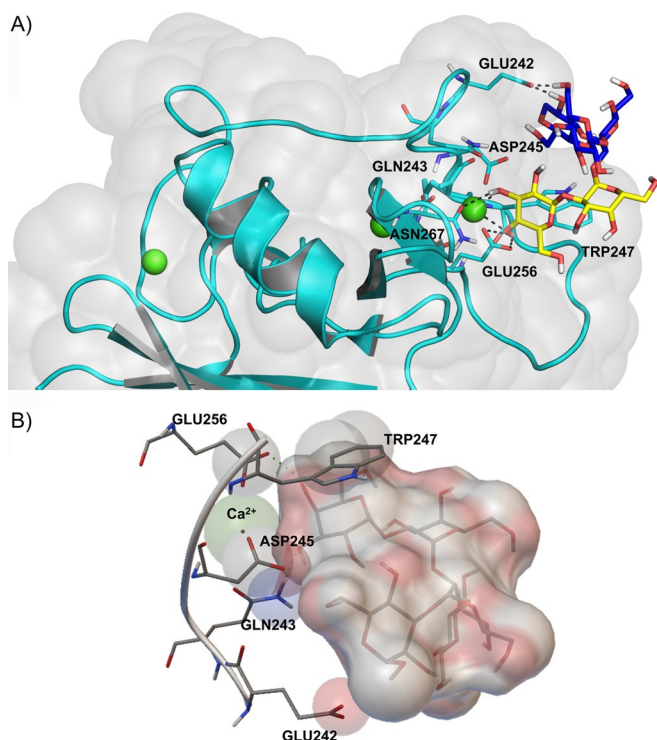
from *E. coli* R3, the structure of which contains the same inner core as that of OS<sub>R1</sub>, but differs in the composition of the outer core by the presence of glucose instead of galactose residues in the terminal region (Figure S4A). In particular, OS<sub>R3</sub> is composed of an  $\alpha$ -Glc-(1–2)- $\alpha$ -Glc-(1–2)-[ $\alpha$ -GlcNAC-(1–3)]- $\alpha$ -Gal-(1–3)- $\alpha$ -Glc branched pentasaccharide connected to the inner core. Interestingly, the lack of terminal galactose units totally destroyed the binding with MGL, which resulted in the absence of STD NMR signals (Figure S4B).

The interaction of MGL-ECD with LOS<sub>R1</sub> was further characterised by using molecular modelling and computational studies to gain a 3D perspective of the molecular recognition mechanism. Because the crystal structure of the protein has not yet been elucidated, a homology model of the MGL CRD (Cys157–Leu284) was built by using the protein structure characterised by the highest sequence identity (74%), namely, the CRD of the mammalian asialoglycoprotein receptor (PDB ID: 1DV8; see also Figure S4) as a template.<sup>[21]</sup> Then, docking calculations of OS<sub>R1</sub> within the modelled structure were performed.

As a first step toward understanding the binding mode of the MGL in the interaction with the bacterial OS, the terminal disaccharide Gal $\alpha$ (1,2)Gal was docked into the primary binding site of the MGL; thus allowing us to confirm the binding specificity of the protein for the galactose. In accordance with data previously reported by Marcelo et al. on the interaction of MGL with a single galactose unit,<sup>[21]</sup> our results revealed that the terminal galactose residue was the major determinant of the binding process. Indeed, although the second galactose moiety was predicted to be close to the binding site of the protein, it seemed to play a secondary role in the interaction with the MGL (see the Supporting Information).

Considering that, from the results of the STD NMR spectroscopy analysis, the outer core of LOS was the moiety in closest proximity to the MGL binding surface, the galactose-containing branched pentasaccharide, which composed the terminal part of the LOS, was then used for docking calculations (Figure 3). The resulting 3D model of the complex revealed that the terminal Gal moiety was located inside the Ca<sup>2+</sup>-containing sugar binding site of the MGL CRD. Crucial contacts between OH-3 and OH-4 of the galactoside ring and the Ca<sup>2+</sup> ion were indeed observed and confirmed the specificity of the receptor toward galactoside residues (Figure 2). From the analysis of the receptor/ligand complex, relevant polar interactions were found between the terminal Gal OH-3 and OH-4 and the side chains of Asp245, Asn268, Asp269 and Glu256 residues. In addition, hydrogen-bond interactions with the carboxylate group of Glu242 were observed, namely, with OH-2 of the glucose residue (G) and OH-6 of the branched Glc moiety (M; Figure 2A). Interestingly, in contrast with the results obtained on the disaccharide Gal $\alpha$ (1,2)Gal (see the Supporting Information), only one possible binding mode has been predicted for the pentasaccharide docked in the protein binding pocket (Figure S7B,C). The selection of this binding mode, in which the polar contacts were markedly reinforced by stacking interactions between the aromatic system of the conserved Trp247 and protons H-1 and H-2 of the galactose moiety, could be ascribable to steric hindrance of the pentasaccharide branched





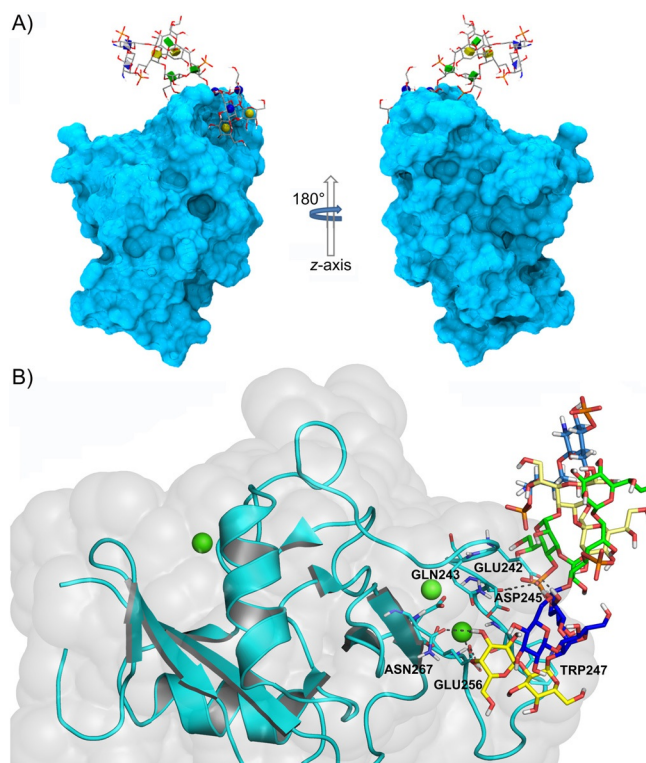
**Figure 3.** A 3D model of the MGL–pentasaccharide complex. A) Pentasaccharide binding at the MGL binding site. The main amino acid residues involved in binding are labelled. Galactose and glucose residues are depicted in yellow and blue, respectively. B) The binding pocket of MGL in the presence of the pentasaccharide in the bound form, as derived from the AutoDock program. Only the amino acid residues involved in the binding process and the  $\text{Ca}^{2+}$  ion are depicted. Green dotted lines represent intermolecular hydrogen bonds.

structure, which permitted only one orientation of the galactose residue (Figure S7B,C).

The described MGL/pentasaccharide complex was used as a starting point to manually dock the whole  $\text{OS}_{\text{R1}}$  structure onto the CRD of MGL (Figure 4). The resulting 3D complex clearly indicated that the inner core region of the LOS was not implicated in the interaction with MGL because it pointed far away from the binding site of the receptor model.

Overall, these observations were in good accordance with the previously discussed STD data; thus further suggesting that significant interactions were mainly established between the protein and sugar residues of the outer core region, whereas the inner core did not seem to play a key role in the interaction with the protein.

In conclusion, we have demonstrated that human CTLs were potential binders of bacterial lipopolysaccharide in solution and that, as expected, they bound the LPS carbohydrate moiety. In particular, herein we have provided evidence that the human MGL was able to detect not only tumour-related carbohydrate structures present in mucin-like glycopeptides, but also glycan structures exposed on the bacterial surface. Whereas a wide range of infectious microorganisms were identified as pathogenic ligands of CTLs, such as DC-SIGN and MR,<sup>[24,25]</sup> glycan microarray profiling on MGL revealed its interaction only with filoviruses and the helminth parasite *Schisto-*



**Figure 4.** A 3D model of the MGL– $\text{OS}_{\text{R1}}$  complex. A) Binding pose of the OS from *E. coli* R1,  $\text{OS}_{\text{R1}}$ , docked into the binding pocket of the MGL starting from the 3D model of the MGL–pentasaccharide complex and based on NMR spectroscopy data. The 3D symbol nomenclature for glycans has been used.<sup>[23]</sup> B) Specific interactions between binding-site residues and the OS are depicted. Sugar residues are coloured according to the symbol nomenclature for glycans. Yellow dotted lines represent polar contacts.

*soma mansonii*.<sup>[26]</sup> Moreover, only a few papers have been published, so far, attempting to prove the potential role of the MGL in detecting LPS structures derived from *Bordetella pertussis* and *Campylobacter jejunii*.<sup>[27,28]</sup> Indeed, we provided molecular insights into the structure of the MGL–LOS complex. We proved, by means of STD NMR spectroscopy, that the MGL was able to recognise and bind the sugar part of the LOS of *E. coli* R1, the major glycoform of which exhibited a terminal Gal- $\alpha(1,2)$ Gal epitope. Further details of the binding mode of MGL to LOS were obtained by homology modelling and docking calculations. Our results improve, at a molecular level, knowledge of the interaction between human CTLs and bacterial glycans and may help to better understand the role of MGL in shaping immunity upon bacterial recognition. MGL has already been reported as a good candidate receptor for DC-based cancer immunotherapy, given its well-known ability to endocytose specific tumour-related antigens.<sup>[29]</sup> The discovery of the ability of MGL to detect an endotoxin component, LOS, of the outer membrane of the Gram-negative bacterium *E. coli* might offer great potential for the future development of therapeutics for bacterial disease intervention. Further studies will be necessary to investigate if and how structural variations between different LOSs could affect recognition and binding process, thus tailoring the outcome of the host–pathogen interplay.

## Acknowledgements

For MGL-ECD production, this work used the Multistep Protein Purification Platform (MP3) of the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI Grant ANR-10-INSB-05-02 and GRAL Grant ANR-10-LABX-49-01 within the Grenoble Partnership for Structural Biology. We thank the French Agence Nationale de la Recherche (ANR) PIA for Glyco@Alps (ANR-15-IDEX-02) for their support of F.B. and M.M. We wish to thank the European Union (FSE, PON Ricerca e Innovazione 2014–2020, Azione I.1 “Dottorati Innovativi con caratterizzazione Industriale”) for funding the Ph.D. grant to R.E.F. Finally, we acknowledge Prof. Sven Muller-Loennies for providing the *E. coli* strains used herein (R1 and R3 core types).

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** carbohydrates · lipopolysaccharides · molecular recognition · NMR spectroscopy · proteins

- [1] S. Kusumoto, K. Fukase, T. Shiba, *Proc. Jpn. Acad. Ser. B* **2010**, *86*, 322–337.
- [2] H. L. Tytgat, S. Lebeer, *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 372–417.
- [3] F. Di Lorenzo, C. De Castro, R. Lanzetta, M. Parrilli, A. Silipo, A. Molinaro, *RSC Drug Discovery Series* **2015**, *43*, 38–63.
- [4] S. D. Gamage, C. M. McGannon, A. A. Weiss, *J. Bacteriol.* **2004**, *186*, 5506–5512.
- [5] A. Silipo, A. Molinaro in *Endotoxins: Structure, Function and Recognition, Sub-Cellular Biochemistry, Vol. 53* (Eds.: X. Wang, P. J. Quinn), Springer, Dordrecht, **2010**, pp. 69–99.
- [6] A. Molinaro, O. Holst, F. Di Lorenzo, M. Callaghan, A. Nurisso, G. D’Errico, A. Zamyatina, F. Peri, R. Berisio, R. Jerala, J. Jiménez-Barbero, A. Silipo, S. Martín-Santamaría, *Chem. Eur. J.* **2015**, *21*, 500–519.
- [7] J. Shi, Y. Zhao, Y. Wang, W. Gao, J. Ding, P. Li, L. Hu, F. Shao, *Nature* **2014**, *514*, 187–192.
- [8] D. A. Wesener, A. Dugan, L. L. Kiessling, *Curr. Opin. Struct. Biol.* **2017**, *44*, 168–178.
- [9] C. A. Orozco, N. Martinez-Bosch, P. E. Guerrero, J. Vinaixa, T. Dalotto-Moreno, M. Iglesias, M. Moreno, M. Djurec, F. Poirier, H. J. Gabius, M. E. Fernandez-Zapico, R. F. Hwang, C. Guerra, G. A. Rabinovich, P. Navarro, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3769–E3778.
- [10] C. Napolitano, I. G. Zizzari, A. Rughetti, H. Rahimi, T. Irimura, H. Clausen, H. H. Wandall, F. Belleudi, F. Bellati, L. Pierelli, L. Frati, M. Nuti, *Eur. J. Immunol.* **2012**, *42*, 936–945.
- [11] S. R. Stowell, C. M. Arthur, R. McBride, O. Berger, N. Razi, J. Heimbürg-Molinaro, L. C. Rodrigues, J. P. Gourdine, A. Noll, S. Von Gunten, D. F. Smith, Y. A. Knirel, J. C. Paulson, R. D. Cummings, *Nat. Chem. Biol.* **2014**, *10*, 470–476.
- [12] K. Denda-Nagai, T. Irimura in *C-Type Lectin Receptors in Immunity* (Ed.: S. Yamasaki), Springer Japan, Tokyo, **2016**, pp. 165–178.
- [13] A. Geissner, A. Reinhardt, C. Rademacher, T. Johannsen, J. Monteiro, B. Lepenies, M. Thépaut, F. Fieschi, J. Mrázková, M. Wimmerova, F. Schuhmacher, S. Götze, D. Grünstein, X. Guo, H. Sik Hahm, J. Kandasamy, D. Leonori, C. E. Martin, S. G. Parameswarappa, S. Pasari, M. K. Schlegel, H. Tanaka, G. Xiao, Y. Yang, C. L. Pereira, C. Anish, P. H. Seeberger, *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 1958–1967.
- [14] N. Suzuki, K. Yamamoto, S. Toyoshima, T. Osawa, T. Irimura, *J. Immunol.* **1996**, *156*, 128–135.
- [15] I. M. Dambuza, G. D. Brown, *Curr. Opin. Immunol.* **2015**, *32*, 21–27.
- [16] J. Valladeau, V. Duvert-Frances, J.-J. Pin, M. J. Kleijmeer, S. Ait-Yahia, O. Ravel, C. Vincent, F. Vega, Jr., A. Helms, D. Gorman, S. M. Zurawski, G. Zurawski, J. Ford, S. Saeland, *J. Immunol.* **2001**, *167*, 5767–5774.
- [17] S. A. F. Jégouzo, A. Quintero-Martínez, X. Ouyang, A. dos Santos, M. E. Taylor, K. Drickamer, *Glycobiology* **2013**, *23*, 853–864.
- [18] M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788; *Angew. Chem.* **1999**, *111*, 1902–1906.
- [19] R. Marchetti, S. Perez, A. Arda, A. Imberty, J. J. Barbero, A. Silipo, *ChemistryOpen* **2016**, *5*, 274–296.
- [20] A. Varki, R. D. Cummings, M. Aebi, N. H. Packer, P. H. Seeberger, J. D. Esko, P. Stanley, G. Hart, A. Darvill, T. Kinoshita, J. J. Prestegard, R. L. Schnaar, H. H. Freeze, J. D. Marth, C. R. Bertozzi, M. E. Etzler, M. Frank, J. F. Vliegthart, T. Lütteke, S. Perez, E. Bolton, P. Rudd, J. Paulson, M. Kanehisa, P. Toukach, K. F. Aoki-Kinoshita, A. Dell, H. Narimatsu, W. York, N. Taniguchi, S. Kornfeld, *Glycobiology* **2015**, *25*, 1323–1324.
- [21] F. Marcelo, F. Garcia-Martin, T. Matsushita, J. Sardinha, H. Coelho, A. Oude-Vrielink, C. Koller, S. André, E. J. Cabrita, H. J. Gabius, S. Nishimura, J. Jiménez-Barbero, F. J. Canada, *Chem. Eur. J.* **2014**, *20*, 16147–16155.
- [22] M. Meier, M. D. Bider, V. N. Malashkevich, M. Spiess, P. Burkhard, *J. Mol. Biol.* **2000**, *300*, 857–865.
- [23] D. F. Thieker, J. A. Hadden, K. Schulten, R. J. Woods, *Glycobiology* **2016**, *26*, 786–787.
- [24] A. Cambi, M. Koopman, C. G. Figdor, *Cell. Microbiol.* **2005**, *7*, 481–488.
- [25] L. Steeghs, S. J. van Vliet, H. Uronen-Hansson, A. van Mourik, A. Engering, M. Sanchez-Hernandez, M. Klein, R. Callard, J. P. van Putten, P. van der Ley, Y. van Kooyk, J. G. van de Winkel, *Cell. Microbiol.* **2006**, *8*, 316–325.
- [26] S. J. van Vliet, E. Saeland, Y. van Kooyk, *Trends Immunol.* **2008**, *29*, 83–90.
- [27] K. V. Vukman, A. Ravidà, A. M. Aldridge, S. M. O’Neill, *J. Leukocyte Biol.* **2013**, *94*, 439–448.
- [28] N. M. van Sorge, N. M. Bleumink, S. J. van Vliet, E. Saeland, W. L. van der Pol, Y. van Kooyk, J. P. van Putten, *Cell. Microbiol.* **2009**, *11*, 1768–1781.
- [29] I. G. Zizzari, P. Martufi, F. Battisti, H. Rahimi, S. Caponnetto, F. Bellati, M. Nuti, A. Rughetti, C. Napolitano, *PLoS One* **2015**, *10*, e0132617.

Manuscript received: February 8, 2019

Accepted manuscript online: March 28, 2019

Version of record online: June 24, 2019

## Supporting Information

### **Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of *Escherichia coli* Lipooligosaccharide**

Meriem Maalej<sup>+, [a, b]</sup> Rosa Ester Forgione<sup>+, [a]</sup> Roberta Marchetti,<sup>\*[a]</sup> François Bulteau,<sup>[b]</sup>  
Michel Thépaut,<sup>[b]</sup> Rosa Lanzetta,<sup>[a]</sup> Cedric Laguri,<sup>[b]</sup> Jean-Pierre Simorre,<sup>[b]</sup> Franck Fieschi,<sup>[b]</sup>  
Antonio Molinaro,<sup>[a]</sup> and Alba Silipo<sup>[a]</sup>

cbic\_201900087\_sm\_miscellaneous\_information.pdf

## Supplementary Information

### NMR structural characterization of OS<sub>R1</sub>.

The structure of the fully deacylated LOS from *E. coli* (OS<sub>R1</sub>) was elucidated by NMR analysis. A complete set of homo- and hetero-nuclear 2D NMR experiments (DQF-COSY, TOCSY, ROESY, NOESY, <sup>13</sup>C–<sup>1</sup>H HSQC and <sup>13</sup>C–<sup>1</sup>H HMBC) was recorded to assign all the spin systems and the monosaccharide sequence.

Briefly, in the anomeric region of the <sup>1</sup>H NMR and <sup>13</sup>C–<sup>1</sup>H HSQC spectra, nine anomeric signals were identified as composing the major glycoform of the oligosaccharide. Moreover, two pairs of signals at 1.85/2.05 and 1.68/2.06 ppm were assigned to the diastereotopic H-3 methylene protons of two Kdo moieties, namely residues **C** and **D**.

*Intra*-residual NOE contacts, together with the values of <sup>3</sup>J<sub>H1,H2</sub> coupling constants, were used to identify the anomeric configuration of each residue, whereas the relative configuration of each monosaccharide unit was obtained by the values of the vicinal <sup>3</sup>J<sub>H,H</sub> coupling constants.

The residues **A** and **B** were identified as α- and β- GlcN residues, respectively, composing the disaccharide backbone of the lipid A. Their amino sugar nature was testified by the correlation, in the HSQC spectrum, of the protons at position 2 with a carbon linked to a nitrogen atom (54.4/54.2 ppm). The high field shift of proton resonances of H-2 was indicative of the absence of acylation at these positions. Spin systems **E**, **F** and **H** were identified as α-heptose residues of the inner core, as confirmed by NOE cross-peaks between the respective H-1 - H-2 intra-residue pairs. Residues **I**, **G** and **M** were all assigned as *gluco*-configured sugars. The large <sup>3</sup>J<sub>H-1,H-2</sub> values, together with the NOE contacts of H-1 with H-3 and H-5, were diagnostic of the β-anomeric configurations of spin system **M**; whereas residues **I** and **G** were recognized as α-glucose given the small <sup>3</sup>J<sub>H-1,H-2</sub> value, confirmed by the intra-residual NOE contact of H-1 with H-2 and by H-5/C-5 chemical shift values. Finally, spin system **K** and **L** were attributed to α-*galacto*-configured sugar residues.

The substitution pattern of all residues was established by *inter*-residual NOE contacts and heteronuclear multiple bond correlations together with the downfield shift of the carbon resonances of the glycosylated positions.

Thus, NMR spectroscopic data allowed us to assign all the resonances belonging to the main glycoform of the LOS of the *E. coli* R1, reported in Figure S1. It was in good agreement with previously published data.<sup>1</sup>

### STD NMR analysis

With the aim to avoid eventual artefacts due to the direct irradiation of the ligand, control experiments were performed. As shown in the Figure S3, the on-resonance irradiation of the protein at 7.5 ppm did not affect the ligand signals. No significant STD enhancements were indeed observed in the corresponding STD spectrum of the ligand alone in solution.

### Homology modelling of MGL

The homology modeling of the MGL CRD was built by using asialoglycoprotein receptor (PDB code: 1DV8) as template (see Supplementary Material and Methods). The model retained the overall protein fold of the template, with a CRD constituted by core region of β- strands and by two α-helices located on the core sides. The structure displayed three calcium ions coordinated by the protein residues, in three different calcium binding sites. The putative sugar-binding site, which was centred around the second calcium ion (Figure S4) was formed by Gln243, Asp245, Glu256, Asn268, Asp269. These residues were also involved in the coordination of the calcium ion. The top edge of the binding was constituted by a curved loop that contained the conserved Trp247 residue, which represents a feature shared by most the Gal-specific proteins characterized so far.<sup>2,3</sup>

MGL - disaccharide complex: From the analysis of the docking calculations on the MGL – disaccharide complex, two main different docking poses, with similar theoretical binding energies



and differing for a rotation of 180°, were predicted (Figure S6, Table S1). In both cases, the terminal residue of galactose was coordinated by the calcium ion through the hydroxyl groups in position three (OH-3) and four (OH-4). OH-3 and OH-4 established also hydrogen bonds with Asp245, Glu256, Asn268 and Asp269. In full agreement with several X-ray structures of other C-type lectins bound to galactoside derivatives<sup>4,5,6,7</sup>, depending on the orientation of the ligand in the binding pocket, a different face of the galactoside ring displays hydrophobic interactions with the tryptophan residue Trp247, engaging in CH- $\pi$  interactions protons H-3, H-4, H-5, and H-6<sub>a,b</sub> in one binding mode and protons H-1 and H-2 in the other (Figure S7).

Furthermore, the two binding modes also differed for the contacts established by protein with the second galactose unit, due to the different orientation of the sugar residue in the binding site of the protein, in mode a and b, respectively. In detail, beside the overall weak contacts for both binding modes, stacking interactions were observed between the H-3 proton of the second galactose and Trp247 only in the binding mode a. On the contrary, in mode b, a hydrogen bond between the OH in position 6 and Glu242 was observed.

## Materials and Methods

### Bacterial growth and LOS isolation

Bacterial *E. coli* R1 and R3 strains were grown, starting from glycerol stock, in Luria Bertani (LB) broth medium at 37°C and the cells were harvested by centrifugation. The freeze-dried cells were extracted according to the petroleum ether–chloroform–phenol (PCP) as described.<sup>8</sup> For each strain, the LOS fraction was precipitated and analysed through DOC-polyacrylamide gel electrophoresis 13,5% followed by silver nitrate staining<sup>9</sup> highlighting the presence of lipooligosaccharide material.

An aliquot of LOS was treated with anhydrous hydrazine (1 ml for 20 mg of sample), stirred at 37°C for 90 min, cooled, poured into ice-cold acetone, and allowed to precipitate. The precipitate was then centrifuged, washed twice with ice-cold acetone, dried, and dissolved in water and lyophilized.<sup>10</sup> The O-deacylated sample was subsequently N-deacylated with a strong alkaline treatment (KOH 4 M, 120 °C). After desalting using a column of Sephadex G-10, the resulting oligosaccharide fraction was further separated utilizing size exclusion chromatography on Biogel P2, from which oligosaccharide OS was obtained (OS<sub>R1</sub>).<sup>11</sup>

### Cloning and Protein expression and purification of MGL-ECD.

MGL amino acids 61 to 292 corresponding to extracellular domain (MGL-ECD) (Figure 1a), were cloned downstream to StrepTagII, Factor Xa protease cleavage site and 3 glycines in between Nde I and Xho I restriction site of pET-30 multiple cloning site region (Figure 1b). MGL-ECD was expressed in *E. coli* BL21(DE3) in 1 L of LB medium supplemented with 50 µg/mL kanamycin at 37°C. Expression was induced for 3 h by addition of 1 mM isopropyl 1-thio-D-galactopyranoside when the culture had reached an OD<sub>600nm</sub> of 0.8 (Figure 1c). The protein was expressed in the cytoplasm as inclusion bodies. Cells were pelleted by a 20 min centrifugation at 4000 g at 4°C.

### Refolding and purification of MGL-ECD

The pellet was re-suspended in 30 mL of buffer 150 mM NaCl, 25 mM Tris-HCl pH 8, 4 mM CaCl<sub>2</sub> (calcium buffer) and one anti-protease tablet (Complete EDTA free, Roche), using a Potter-Elvehjem. Bacteria were disrupted by sonication and cell debris were eliminated by centrifugation at 100000 g for 30 min at 4°C. The pellet was successively washed in 30 mL of buffer 2 M Urea, 150 mM NaCl, 25mM Tris-HCl pH 8 and 1 % triton X-100, centrifuged at 100000 g, washed in 30 mL of calcium buffer, centrifuged at 100000 g. Inclusion bodies were finally solubilized in 30 mL of buffer 6 M guanidine, 150 mM NaCl, 25 mM tris-HCl and 0,01 %  $\beta$ -mercaptoethanol (guanidine buffer) and insoluble material was removed by a centrifugation of 30 min at 100000 g at 4°C. Based on calculated MGL-ECD  $\epsilon(280)=69440$  M<sup>-1</sup>cm<sup>-1</sup>, the protein concentration of supernatant was adjusted to 2 mg/mL with guanidine buffer. The refolding of the protein was performed by a drop-by-drop dilution in a buffer 150 mM NaCl, 100 mM Tris-HCl pH 8 and 25 mM CaCl<sub>2</sub>. Product of refolding was dialyzed 3 times against calcium buffer 25 mM Tris-HCl pH 8, 150 mM NaCl and 4 mM CaCl<sub>2</sub>. Insoluble material was eliminated by a final centrifugation step at 100000 g for 1 h at 4°C. The protein was purified using a two steps protocol on GalNAc-Agarose affinity column (Sigma), eluted with buffer 150 mM NaCl, 25 mM tris-HCl and 5 mM EDTA. The eluted sample is loaded onto a Toyopearl HW-50S gel filtration column

(Tosoh Bioscience), equilibrated with the calcium buffer. After analysis by SDS-PAGE gel fractions containing MGL-ECD were pooled and concentrated to 5.6 mg/mL by ultrafiltration using a Vivaspin 20 PES 10 kDa MWCO (Figure 1d).

## NMR spectroscopy

All NMR experiments were recorded at 298 K on Bruker 600 DRX equipped with a cryo probe. The samples were solved in Tris D<sub>2</sub>O 25 mM, NaCl 150mM, CaCl<sub>2</sub> 4 mM. For all homo-nuclear experiments, data sets were acquired with ( $t_1 \times t_2$ ) of 4096  $\times$  512 points. TOCSY experiments were performed with spin lock times of 100 ms. ROESY and NOESY spectra were acquired with mixing times of 300 ms. In all homonuclear experiments the data matrix was zero-filled in both dimensions to give a matrix of 4 k  $\times$  2 k points and a cosine-bell function before Fourier transformation was used to enhance the resolution. Coupling constants were determined by 2D phase sensitive DQF-COSY. HSQC and HMBC experiments were recorded in the <sup>1</sup>H-detected mode via single quantum coherence with proton decoupling in the <sup>13</sup>C domain, using data sets of 2048  $\times$  256 points. Experiments were performed in the phase-sensitive mode as described previously.<sup>12</sup> A 60 ms delay was used for the evolution of long-range connectivities in the HMBC experiment. In all heteronuclear experiments the data matrix was extended to 2048  $\times$  1024 points using forward linear prediction extrapolation.

STD NMR experiments were acquired by using 32 k data points then zero filled up to 64 k data points prior to processing; in order to increase the S/N ratio the FIDs were multiplied with an exponential function ( $I_b = 1-2$  Hz). Control STD NMR experiments were performed on both protein and ligand in the free state at different irradiation frequencies, and by using the same experimental conditions set for the analysis of the mixture, in order to avoid artefacts due to false-positive STD signals. The on-resonance frequency was set at 7.5 ppm since no STD ligand signals were observed in the corresponding reference spectrum, while 100 ppm was set as the off-resonance pulse frequency. The protein - ligand molar ratio was 1:100 and the saturation time was set at 2 s. A 40 Gauss pulse with a length of 50 ms and an attenuation of 60 db was used for protein saturation and the unwanted broad resonance signals of the protein were avoided by using a spin lock pulse of 50 ms.

The STD effects were calculated by  $(I_0 - I_{sat})/I_0$ , where  $(I_0 - I_{sat})$  is the intensity of the signal in the STD NMR spectrum and  $I_0$  is the peak intensity of an unsaturated reference spectrum (off-resonance). The STD signal with the highest intensity was set to 100% and the others were normalized to this peak. Data acquisition and processing were performed with TOPSPIN 3.1 software.

## Homology modeling of MGL

Human macrophage galactose-type lectin (MGL, isoform 2) encoding sequence was extracted from NCBI (access. No. NP\_006335.2). The carbohydrate recognition domain was identified by submitting the query sequence to CD-search<sup>13</sup>. As result, the region between the residues 157-284 of the protein query was found to match to the C-type lectin-like domain (access. No. cd03590). Then, the sequence interval was aligned against several reference sequences with known three-dimensional structure to identify potential templates for the homology modeling of MGL. The structure of the CRD (aa Cys154 – Leu280) of the mammalian asialoglycoprotein (PDB code: 1DV8) was selected as the template because of the highest identity (74%) and query cover (100%). The homology model was generated by submitting the target template alignment to SWISS-MODEL server<sup>14</sup>. Missing hydrogen atoms were added, zero-order bond to metals were created and protonation state of ionisable groups was computed by using Maestro Protein Preparation Wizard.<sup>15</sup> The structure was submitted to 100 000 steps of steepest descent minimization with MacroModel,<sup>16</sup> using OPLS\_2005 force field as integrated in MAESTRO suite of programs.

The quality of the homology models was assessed with the program PROCHECK,<sup>17</sup> and Ramachandran plot analysis for MGL showed that approximately 88% of all residues were within the most favoured regions and no residues were in the disallowed regions.

## Docking calculations

The 3D coordinates of Gal $\alpha$ (1-2)Gal, of the core pentasaccharide region and of the LOS were built by means of MAESTRO. The bonds were parametrised and the geometries of each ligand were optimized by 100 000 step of steepest descent minimization with OPLS\_2005 force field by using MacroModel. Ligands were prepared for docking calculations using AutoDockTools,<sup>18</sup> setting all rotatable bonds free to move, except for the glycosidic linkages, during the docking calculations.

Docking calculations of all compounds were performed by using AutoDock 4.2.2..<sup>19</sup> Analysis of the docking poses was performed with AutoDockTools. The docking protocol was validated by carrying out the docking of asialoglycoprotein receptor in complex with  $\alpha$ -lactose (PDB-ID: 5JPV).<sup>20</sup>

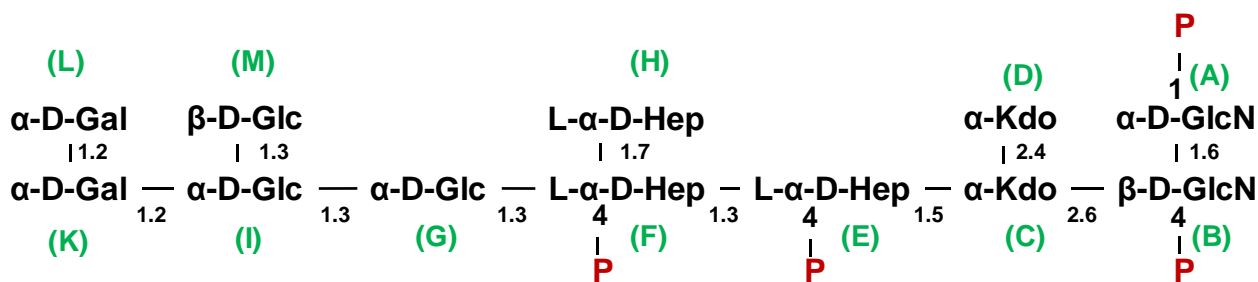
The grid point spacing was set at 0.375 Å, and a hexahedral box was built with x, y, z dimensions: 40 Å, 42 Å, 40 Å centered in the centroid position among residues Gln243, Asp245, Glu256, Asn268, Asp269 of MGL model and Gln239, Asp241, Glu252 and Asn264 of asialoglycoprotein receptor. A total of 200 runs using Lamarckian Genetic algorithm was performed, with a population size of 100, and 250000 energy evaluations. The selected docked complexes were submitted to 100 000 steps of steepest descent minimization with MacroModel,<sup>42</sup> using OPLS\_2005 force field.

## Supporting Figures and Tables

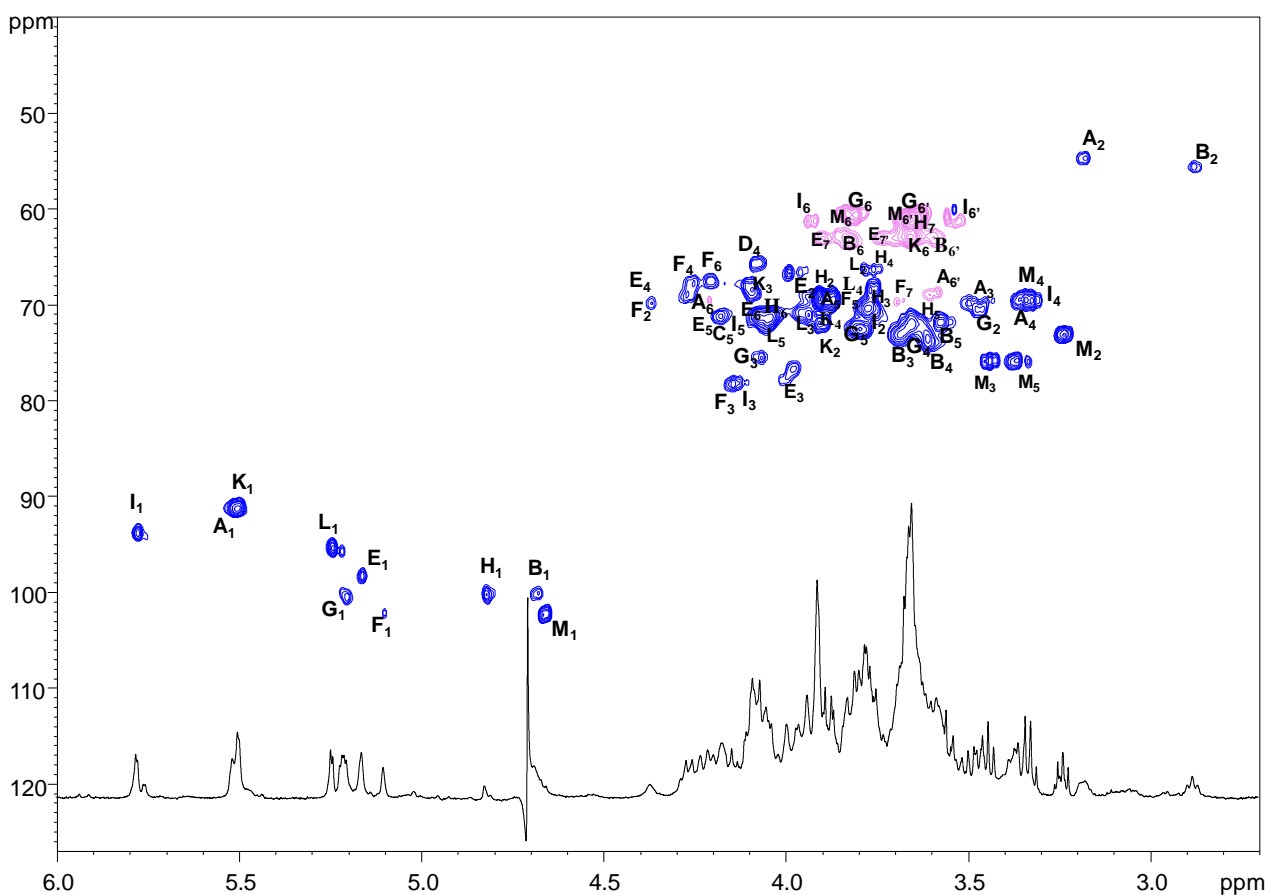
**Table S1** Cluster rank, Cluster population, Computed binding energy and RMSD (Root Mean Square Deviation) for the molecular docking (AutoDock) of Gal $\alpha$ (1,2)Gal disaccharide in the binding pocket of MGL. The values are referred to the lowest energy docked conformation from the highest ranked clusters 1 and 2, corresponding to mode A and mode B respectively.

Docking pose	Cluster Rank	Number of conformations in this cluster	Estimated Free Energy of Binding kcal/mol	RMSD from reference structure
<b>Cluster 1 (Mode A)</b>	1	97	- 1.23	15.934 Å
<b>Cluster 2 (Mode B)</b>	2	56	-1.08	13.401 Å

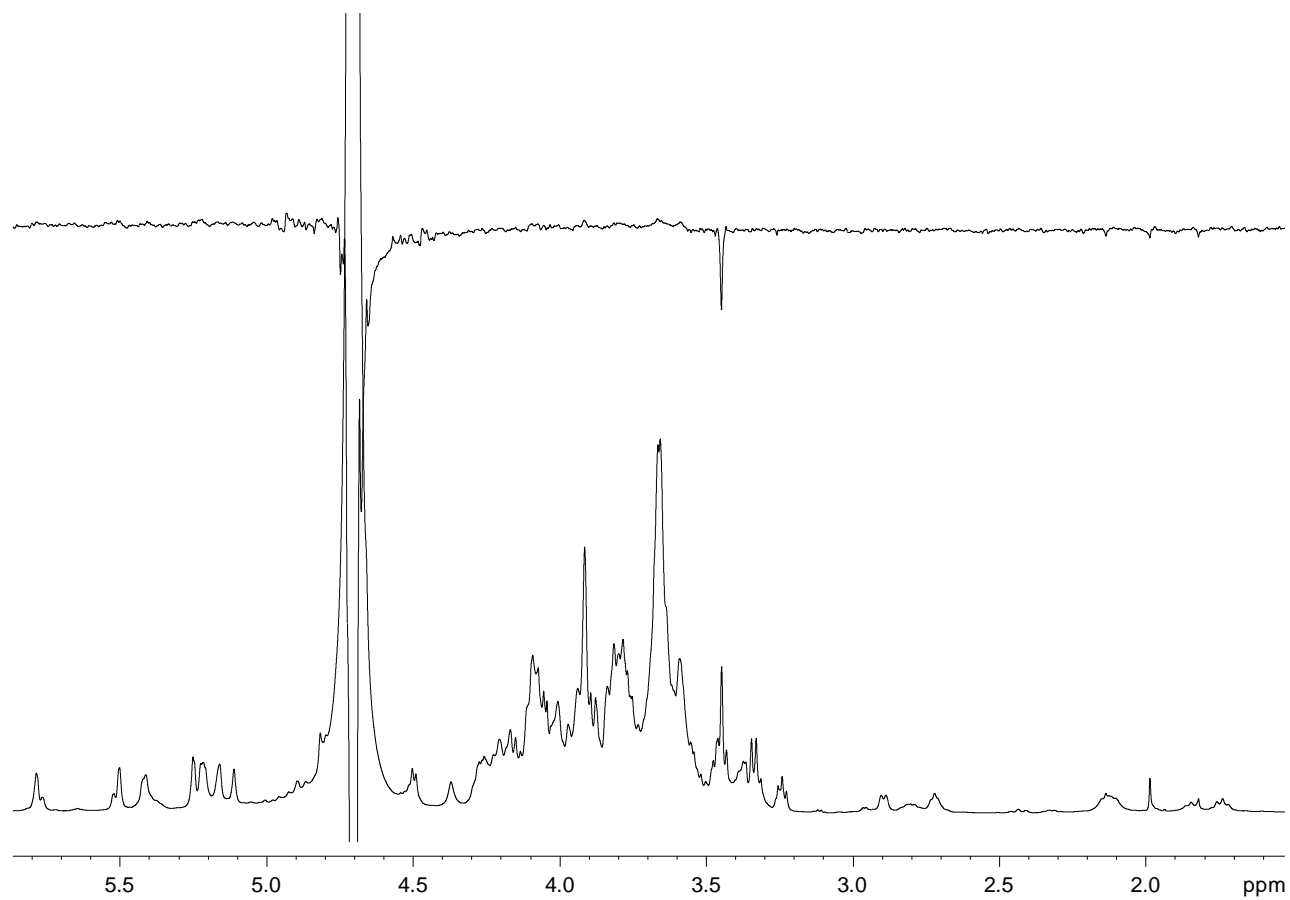
**Figure S1.** Structural assessment for the core OS from *E. coli* R1 (OS<sub>R1</sub>) derived from combination of results on fully deacylated product.



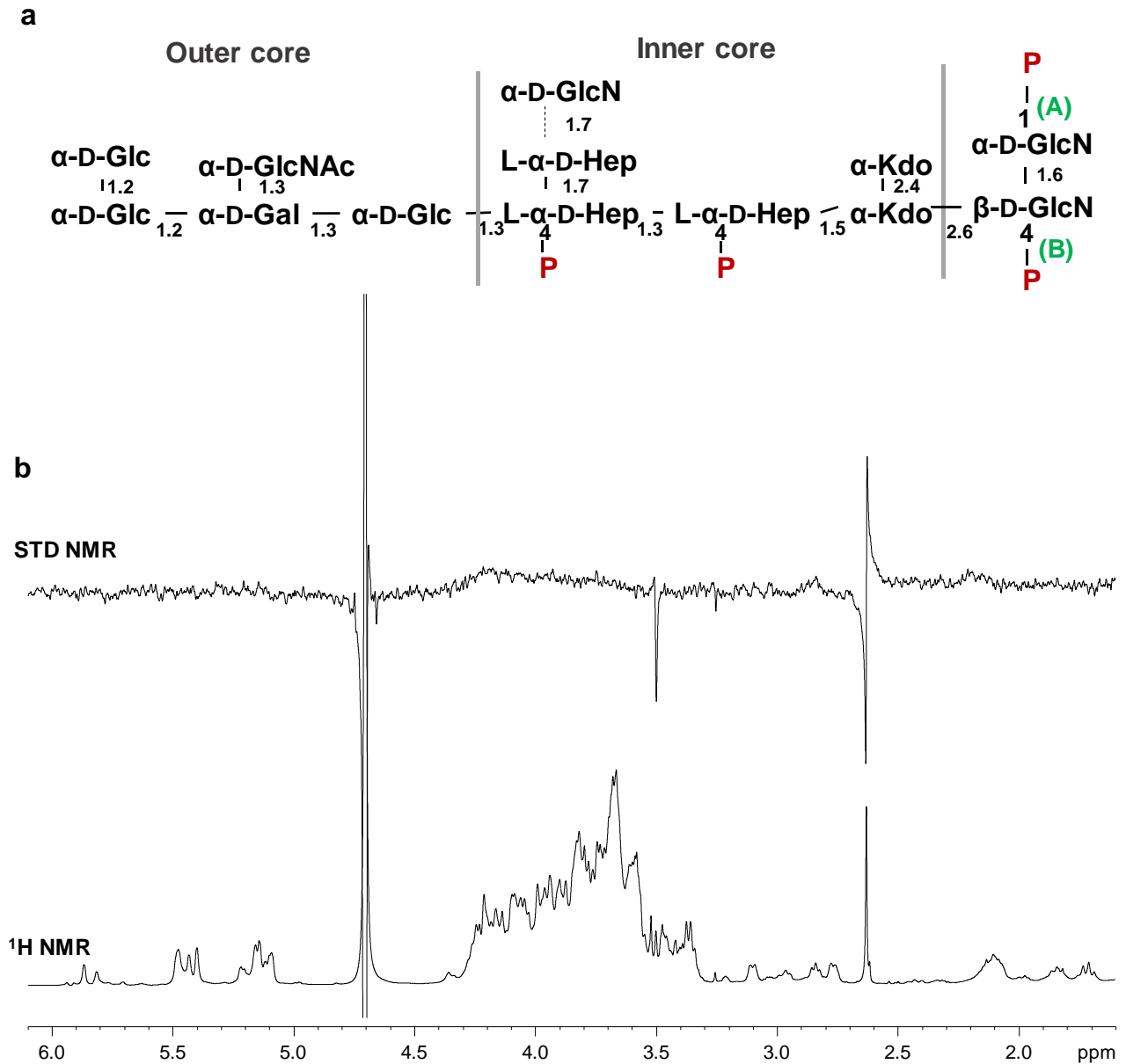
**Figure S2.** Zoom of the  $^{13}\text{C}$ ,  $^1\text{H}$  HSQC spectrum of OS<sub>R1</sub>. The most relevant heteronuclear correlations are reported.



**Figure S3.** Control STD NMR spectrum of the OS<sub>R1</sub> in the free state. The on-resonance irradiation frequency was set at 7.5ppm and the saturation time was 2 seconds.



**Figure S4.** a. Structure of the oligosaccharide OS (OS<sub>R3</sub>) derived from the lipooligosaccharide of *E. coli* R3. b. <sup>1</sup>H NMR reference spectrum (bottom) and 1D STD NMR spectrum (up) of the mixture MGL – OS<sub>R3</sub> 1:100. The absence of signals into STD NMR spectrum indicated that the core-OS<sub>R3</sub> was not recognized by the protein.



**Figure S5.**

**a.** Multiple alignment of human MGL with the homologous protein ASGR. Alignment was performed with the CLUSTAL Omega program and minimally adjusted manually. The sequence intervals of the CRD domain of human-MGL (Cys157-Leu284) and of the template used for the homology modeling, namely ASGR (Cys154-Leu181) are highlighted in grey. The residues important for galactoside binding are highlighted in cyan.

**b.** Superimposition of MGL-CRD homology-based model (in blue) and the X-ray crystal structure from ASGR (in magenta).

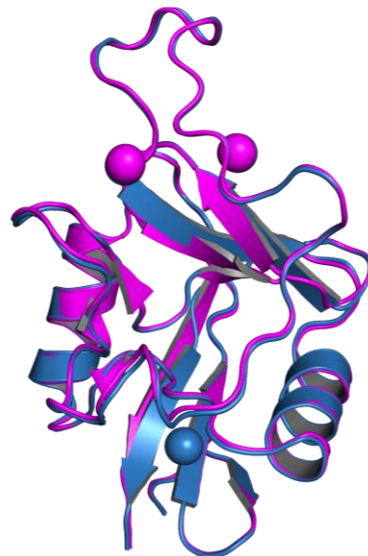
**a** CLUSTAL O(1.2.4) multiple sequence alignment

NP_006335.2*	MTRTYENFQYLENKV-KVQGFKNGPLPLQSLQRLCSGPHLLLSLGLGLLLLVIICVVG	59
BAB83508.1**	MTKEYQDLQHLDNNEESDHHQLRKGPPLQRLCSGPRLLLLSLGLSLLLLLVVVCVIG	60
	** : *::*:*:*: . : :::* * * ***** *****.******:***:	
NP_006335.2*	FQNSKFQRDVLTLRDFSNFTSNTVAEIQALTSQGSSEETIASLKAIEGFKQERQAVH	119
BAB83508.1**	SQNSQLQEELRGLRETFSNFTASTEAVKGLSTQGGNVGRKMKSLQLEKQKDLSEDH	120
	* ** :*:*. * * *****.* *::*:*:**.. : : *::*: * : : . *	
NP_006335.2*	SEMLLRVQQLVQDLKLTLCQVATLNNNGEEASTEGTCCPVNWVEHQDSCYWFSGMSWA	179
BAB83508.1**	SSLLLHVQFVSDLRSLSCQMAALQNGS----ERTCCPVNWVEHERSCYWFSRSGKAWA	176
	*.:***:*. *.*. *.:***:*. *.*. * *****.* *****.* * : **	
NP_006335.2*	EAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYTWMGLSDPEGAWKWVDGTDYATGFQNW	239
BAB83508.1**	DADNYCRLEDAHLVVVTSWEEQKQFVQHHIGPVNTWMGLHDQNGPWKWDGTDYETGFKNW	236
	:*::**:*:***:*. * * **:*:***:*. * ***** * : * ***** * **:*:	
NP_006335.2*	KPGQPDDWQGHGLGGGEDCAHFHPDGRWNDDVCQRPYHWVCEAGLGQTSQESH--	292
BAB83508.1**	RPEQPDDWYGHGLGGGEDCAHFHTDDGRWNDDVCQRPYRWVCELDKASQEPPLL	291
	: * ***** ***** * ***** *****.* ***** : *::**	

\* C-Type lectin domain family 10 member A, isoform 2

\*\* Asialoglycoprotein receptor 1

**b**

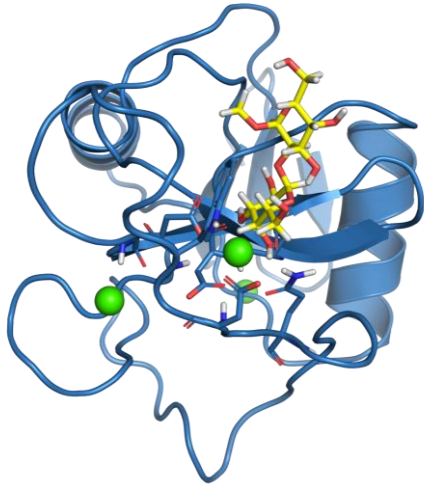




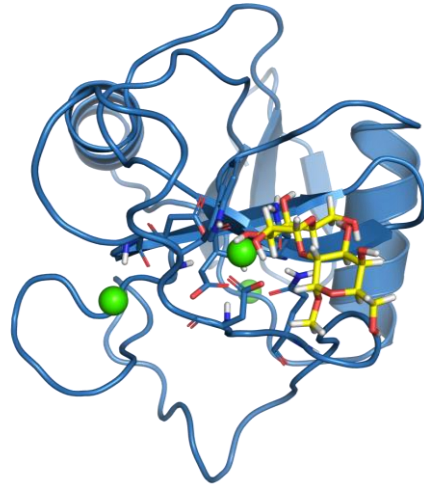
**FigureS6. 3D models of MGL - Gal- $\alpha$ -(1-2)-Gal complex**

Complexes of MGL, obtained from docking calculation, showing the two main binding modes (**a** and **b**) of the galactose disaccharide when bound to MGL CRD. The two complexes mainly differ for the orientation of the terminal galactose moiety in the binding site of the protein.

**a**



**b**



**Figure S7.**

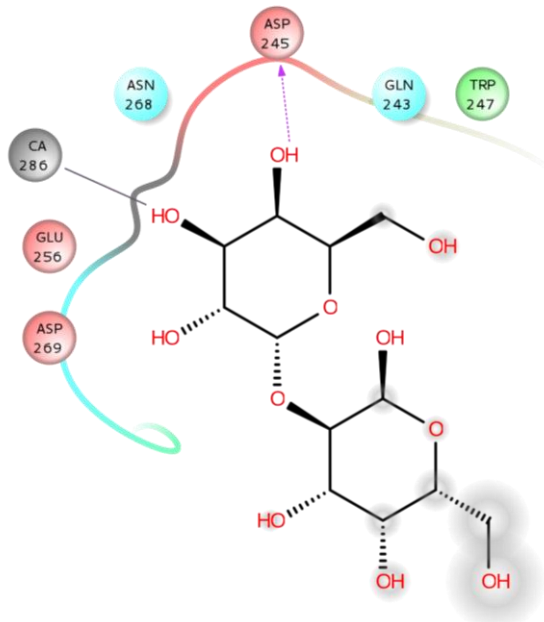
**a.** Gal $\alpha$ (1,2)Gal disaccharide binding at the MGL binding site. The main protein: ligand interactions of the two binding modes, A (on the left), B (on the right) are shown.

**b.** Expansion of the protein binding site showing the stacking interaction between the residue Trp247 and the galactoside ring of Gal $\alpha$ (1,2)Gal, according to the binding mode A (left) and B (right).

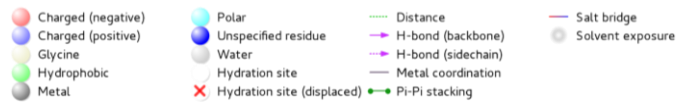
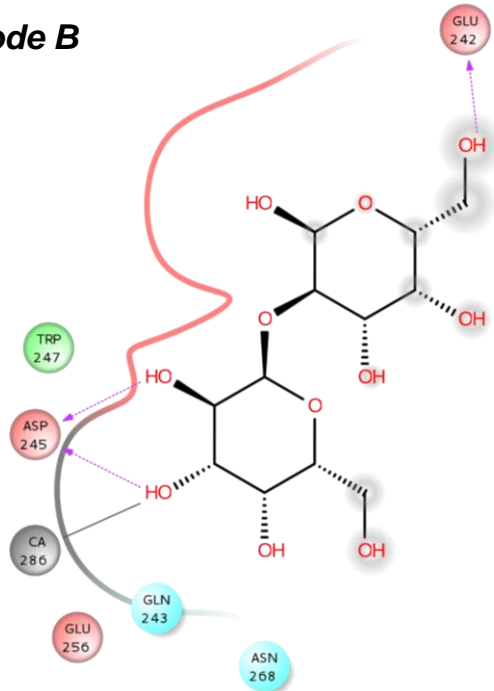
**c.** Expansion of the protein binding site showing the stacking interaction between the tryptophan, Trp247, and the terminal galactose unit of the pentasaccharide.

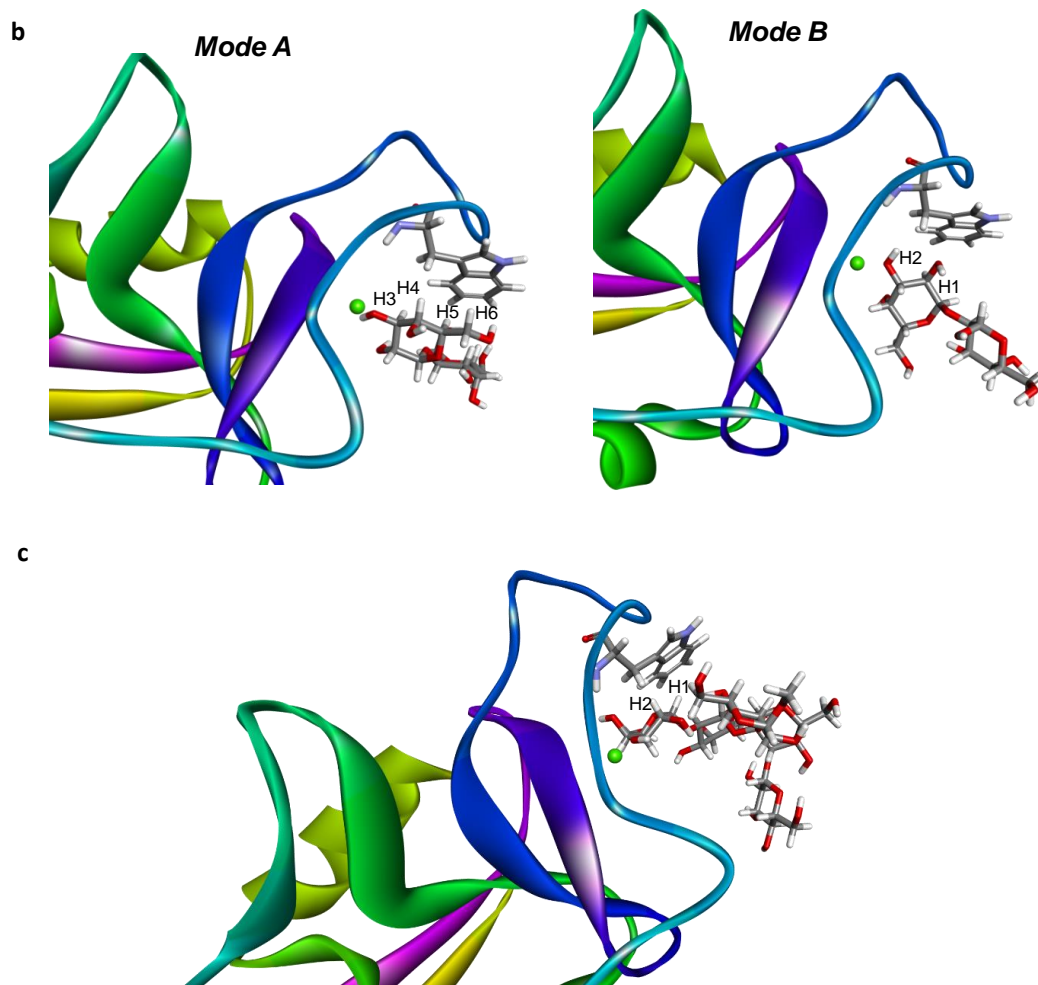
**a**

**Mode A**



**Mode B**





## References

- <sup>1</sup> S. Müller-Loennies, B. Lindner, H. Brade. *Eur J Biochem* **2002**, 269(23), 5982-91
- <sup>2</sup> AR. Kolatkar, AK. Leung, R. Isecke, R. Brossmer, K. Drickamer, WI. Weis. *J. Biol. Chem.* **1998**, 273, 19502–19508
- <sup>3</sup> H. Sugawara, M. Kusunoki, G. Kurisu, T. Fujimoto, H. Aoyagi, T. Hatakeyama, *J. Biol. Chem.* **2004**, 279, 45219–45225
- <sup>4</sup> AR. Kolatkar, AK. Leung, R. Isecke, R. Brossmer, K. Drickamer, WI. Weis. *J. Biol. Chem.* **1998**, 273, 19502–19508
- <sup>5</sup> H. Sugawara, M. Kusunoki, G. Kurisu, T. Fujimoto, H. Aoyagi, T. Hatakeyama, *J. Biol. Chem.* **2004**, 279, 45219–45225
- <sup>6</sup> T. Hatakeyama, T. Kamiya, M. Kusunoki, S. Nakamura-Tsuruta, J. Hirabayashi, S. Goda, H. Unno, *J. Biol. Chem.* **2011**, 286, 10305 – 10315
- <sup>7</sup> H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer, W. I. Weis, *J. Mol. Biol.* **2011**, 405, 1027 –1039
- <sup>8</sup> S. Leone, A. Silipo, EL. Nazarenko, R. Lanzetta, M. Parrilli, A. Molinaro. *Marine Drugs* **2007**, 5, 85–112.
- <sup>9</sup> R. Kittelberger, F. Hilbink. *J Biochem Biophys Methods* **1993**, 26(1), 81–6
- <sup>10</sup> O. Holst. *Methods Mol Biol* **2000**, 145, 345–53
- <sup>11</sup> O. Holst, S. Muller-Loennies, B. Lindner, H. Brade. *Eur J Biochem* **1993**, 214, 695–701
- <sup>12</sup> DJ. States, RA. Haberkorn, DJ. Ruben. *J. Magn. Reson* **1982**, 48, 286-292
- <sup>13</sup> A. Marchler-Bauer, S. H. Bryant. *Nucleic Acids Res* **2004**, 32, W327–W331.
- <sup>14</sup> K. Arnold, L. Bordoli, J. Kopp, T. Schwede. *Bioinformatics* **2006**, 22, 195-201
- <sup>15</sup> Schrodinger. 2012. Epik version 2.3, Schrodinger, LLC, New York, NY.
- <sup>16</sup> MacroModel, Schrödinger, LLC, New York, NY, 2018
- <sup>17</sup> RA. Laskowski, MW. MacArthur, DS. Moss, JM. Thornton. *J. App. Cryst.* **1993**, 26, 283-291
- <sup>18</sup> GM. Morris, R. Huey, W. Lindstrom, MF. Sanner, RK. Belew, DS. Goodsell, AJ. Olson. *J. Computational Chemistry* **2009**, 16, 2785-91

---

<sup>19</sup> GM. Morris, DS. Goodsell, RS. Halliday, R. Huey, WE. Hart, RK. Belew AJ. Olson. *Journal of Computational Chemistry* **1998**, *19*, 1639-1662

<sup>20</sup> CA. Sanhueza, MM. Baksh, B. Thuma, MD. Roy, S. Dutta, C. Prévile, BA. Chrnyk, K. Beaumont, R. Dullea, M. Ammirati, S. Liu, D. Gebhard, JE. Finley, CT. Salatto, A. King-Ahmad, I. Stock, K. Atkinson, B. Reidich, W. Lin, R. Kumar, M. Tu, E. Menhaji-Klotz, DA. Price, S. Liras, MG. Finn, V. Mascitti. *J. Am. Chem. Soc.* **2017**, *139* (9), 3528–3536

## ***Chapter X.***

## X. Conclusions and perspectives

### X-1. Conclusions

As a C-type lectin receptor (CLR) of Macrophages and Dendritic Cells, MGL mediates contacts with other cells being an adhesion molecule and a Pathogen Recognition Receptor (PRR) during innate immune response. But many questions still need to be resolved regarding its function as a PRR of pathogenic bacteria. The reasons behind these interrogations are related to the lack of structural information about this lectin and, on the other hand, to the structure complexity of potential bacterial ligands (i.e., PAMPs like LPS).

In this PhD work, we succeeded to cope with such complex and large interaction system using a combination of scientific approaches. From the lectin point of view, we studied two domains: The Carbohydrate Recognition Domain (**CRD**) and the Extracellular Domain (**ECD**) which allowed us to figure out how both MGL constructs could interact with the studied ligands, considering the specificity and the affinity exhibited by these two domains.

#### X-1-1. Qualitative control of the isolated glycoconjugates

Lipopolysaccharides (LPS) are PAMPs exposed at the outer membrane of Gram-negative bacteria, that mediate interactions with the innate immune system by driving contacts with immune cells receptors during the first step of pathogen recognition. However, the mechanism of LPS interactions is still limited to the sole TLR system which comprises a complex of receptors that come into play when LPS aggregates are present.

The advantage related to this shared PhD is about the opportunity of studying, by NMR in Grenoble, different versions of LPS glycoconjugates, considering laboratory's expertise on LPS investigations in Naples for glycoconjugates extractions from natural sources. LPS glycoconjugates could be recognized by CLRs through molecular interactions with the carbohydrate region, such bindings have been observed for DC-SIGN lectin<sup>325</sup>. CLRs binding to carbohydrates composing glycoconjugates is specific and depends on the exposed sugar motifs. To understand, in a first step, the diversity of the carbohydrate codes exhibited by LPS, we studied four different LPS glycoconjugates produced by wild-type and mutants of *E. coli* and *B. cenocepacia* strains. For that, subsequent approaches were carried out from the bacterial growth of these strains to the identification of LPSs compositions.

The nature of the isolated glycoconjugates was determined through the analyses of electrophoresis profiles of **LPS/LOS** biomolecules. Later, the presence of the carbohydrate

residues was assessed using GC-MS analyses for all the isolated LOSs. We particularly inspected NMR spectra of **OS R1** which confirmed our previous spectrometry results.

### **X-1-2. The CRD of human MGL harbors a potential secondary binding surface for *E. coli* glycoconjugates**

When we considered the monomeric CRD MGL for interaction studies, we characterized the binding between the three *E. coli* glycoconjugates OS R1, OS R3 and LPS O157:H7, at the molecular level. The analyses of these “close-up” versions of ligands (c.f. MGL CRD) resulted in determining binding surfaces on CRD MGL using NMR titrations and the K<sub>d</sub> value was estimated for LPS O157:H7 to be in the millimolar range. As reported for other CLR domains mainly presented in chapter II, CRD domains usually exhibit low affinity for terminal sugars of oligosaccharides or monovalent ligands.

We observed, for all *E. coli* ligands, a secondary binding surface, in addition to the Ca<sup>2+</sup> binding site. According to our NMR titration data including GalNAc provoked chemical shifts (CSPs), we managed to rapidly spot residues from the new binding surface and to compare binding profiles between the three ligands. The existence of more than one binding site has been frequently observed for CRDs of CLR domains. Accordingly, we think that the relative locations of binding sites in CRD MGL may accommodate its interaction with more than one ligand or with a multivalent ligand, as it is the case for human Mannose Binding Lectin (MBL) for example, where the binding sites positioning is a key determinant<sup>315</sup>. Two possible binding scenarios could be suggested for human **MGL CRD** interaction with LPS structures where (i) both binding sites (the first Ca<sup>2+</sup> binding site and the putative secondary binding surface) are involved, with which one stabilizes the other, or (ii) contacts take place mainly at the second binding surface. Calcium binding sites implication is not easy to confirm from biomolecular and biophysical essays, with lack of complementary atomic level insights.

### **X-1-3. The ECD of MGL exhibits high affinity for *E. coli* R1 glycoconjugates**

In addition to CRD MGL, the binding of ECD MGL to *E. coli* glycoconjugates was studied by a combination of approaches that consider the high molecular weight of the lectin and the different degrees of solubilities of LPS versions. Here, we characterized the binding between *E. coli* LOS (OS R1 via terminal di-Gal) and **ECD MGL**. We found out that ECD MGL is extremely selective and recognized specifically the OS R1 epitope over OS R3 despite the high structural similarities between the two ligands and, accordingly the interaction could occur via avidity-based mechanism.

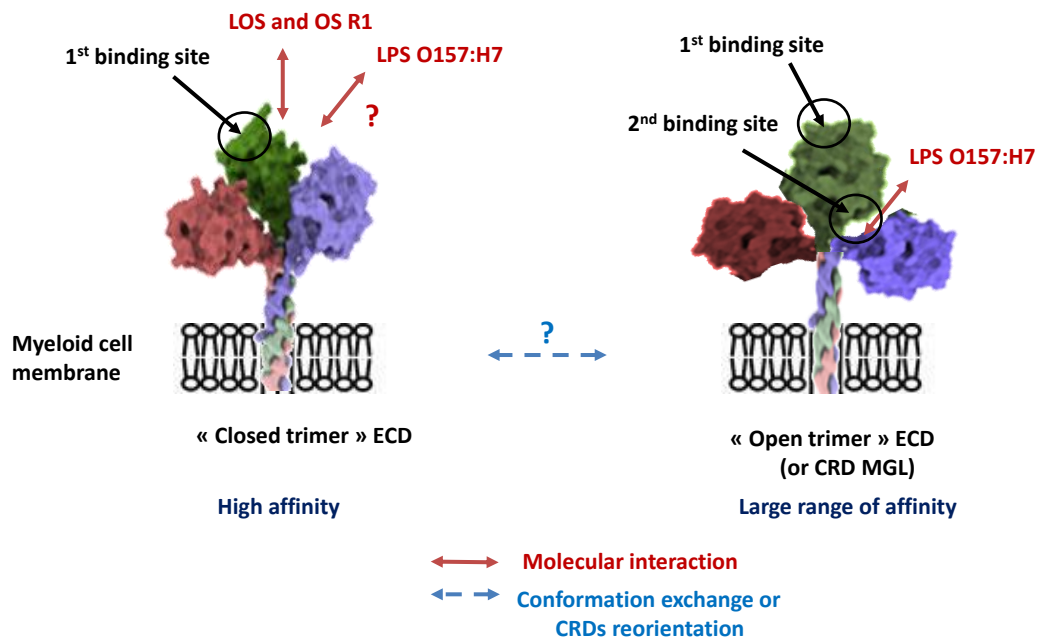
This binding was further confirmed to be strong based on our BLI measurements together with TEM analyses. Our results support the idea of **multivalent binding** ensured by ECD MGL trimeric conformation, comprising the three CRDs. The extended and branched structure of LOS R1 would perhaps contribute to strengthen the binding to ECD MGL by favouring multivalent interactions favored by such ligand arrangement<sup>370</sup>. LOS R1 binds to ECD MGL with a Kd in the nanomolar range in agreement with the avidity improvement as we assumed it to be. A Kd of ~300 nM supports a strong binding that might involve many regions of the core OS such as distal ends (outer core and perhaps inner core). Still, this binding mode must be demonstrated eventually by mapping binding epitopes either on isotopically labeled <sup>15</sup>N-<sup>13</sup>C-LOS R1 or <sup>15</sup>N-<sup>13</sup>C-ECD MGL using NMR spectroscopy experiments.

### **X-1-3-1. Human MGL structural organisation on Myeloid cell membrane, an overview**

Our results regarding the strong binding affinity within MGL-LOS R1 interaction system and the weak binding observed for LPS O157:H7, incited us to wonder if there is a potential conformation exchange of MGL, amidst these interactions, for instance between closed and open flower bouquet structures, in nature, like for DC-SIGN and DC-SIGNR CLR<sup>s</sup><sup>167</sup>, having no idea about its large ligand bound-structure. MGL conformational structure is flexible<sup>350</sup> and, its plasticity and dynamics would perhaps depend on the ligand structure that may be recruited or not by ECD MGL.

Figure 104 shows an imagined view of MGL structural organisation during molecular interactions with the studied ligands, considering NMR titrations and biomolecular analyses established on both the CRD and the ECD of MGL, respectively. Is there a conformational change exhibited by MGL (by moving the three CRDs) and presumably induced by the ligand being faced?





**Figure 104. Prospective view of human ECD MGL interactions with *E. coli* glycoconjugates depending on MGL conformational organisation in immune cell membranes.** In the membrane, the view of the two biological assemblies of trimeric human ECD MGL is imagined as human MBL (pdb 1HUP) and rat MBL (pdb1KWV) or DC-SIGN (pdb1XAR) for closed and open structures, respectively. Binding sites locations are indicated by the black circles.

In addition to the relative location of binding sites in the CRD of lectins in general, the orientation of the CRDs to one another in the ECD structure is variable thus defining the carbohydrates binding sites, according to what has been observed for human and rat MBL trimers<sup>315</sup>.

As said, the high order of affinity that we observed from ECD MGL-LOS R1 (and OS) interactions may be prompted by the structural assembly of ECD MGL, by placing a group of three CRDs and this is usually observed for receptors<sup>371,372</sup>. As for LPS O157:H7 ligand, it could be that the weak binding of this glycoconjugate observed mainly on the second binding site on CRD MGL by NMR titrations, becomes even weaker until it got cancelled when the three CRDs of MGL are convened forming an ECD MGL “closed trimer” structure. Anyhow, this hypothetical LPS O157:H7 binding mode needs to be clarified yet.

Based on our experimental findings on **CRD MGL** mainly through NMR titrations, distal MGL binding sites were activated following the interaction with both LPS ligands versions (i.e., OS and LPS) however, some CRD MGL residues were either not or marginally affected then others, depending on the ligand being present. This finding suggests that MGL binding sites availabilities for the interaction could differ from one *E. coli* LPS to the other despite structure similarities of

those ligands. We predicted a weak affinity for LPS and we assumed that a long range of affinity (presumably low) could be attributed to OS R1. It would be possible that, in an imagined **ECD MGL** version “open trimer” (c.f. monomers of CRD MGL), the probable secondary binding surface which is exclusively located at the N-Terminus parts of the three CRDs, becomes more reachable by *E. coli* LPS ligands and authorises their recruitment.

Conversely, we presumed an avidity-based binding (“closed ECD trimer”) between the OS R1 and ECD MGL that was initially predicted from OS R1 mapping and STD-NMR analyses and BLI measurements together with the observation of strong disaggregation effects induced by MGL on LOS R1 vesicles as observed in TEM images (using LOS R1 as ligand). This view of MGL conformational changes, for specificity enhancement or tuning of the binding, in presence of structurally different *E. coli* ligands, is theoretical and still ill-defined and so it needs to be furthermore confirmed.

### **X-1-3-2. Human MGL interacts with *E. coli* R1 bacteria at exponential phase**

Human MGL interacted with free OS R1, LOS R1 vesicles and furthermore with glycoconjugates at the surface of intact *E. coli* R1 bacteria. However, *E. coli* R3 and B strains were not recognized by MGL despite cell-walls resemblances between these strains. Moreover, *E. coli* R1-MGL interaction occurred exclusively at exponential phase of growth. Our findings are in deep consistency with what has been found within a study about **human MBL**, better if considering that this lectin has similar binding properties as human MGL as previously discussed. The study consists in analyzing human MBL binding to several *H. alvei* strains that produce different LPSs among which *H. alvei* PCM 1209 strain which exposes, at the cell-surface, an outer core structure bearing a di-glucose motif. Human MBL specificity towards these residues of the core was first confirmed by SPR essays and, flow cytometry measurements showed that this lectin specifically binds to *H. alvei* PCM 1209. *H. alvei* LOS production was evaluated and was high during the first 3 hours of growth<sup>373</sup> (reaching exponential phase in our case). Fittingly, we could suggest that the observed MGL binding to *E. coli* R1 bacteria is specific to that strain and it mainly takes place at the outer core region of LOS R1 by assuming that MGL binds to di-galactose of OS R1 as concluded from our STD-NMR analyses.

#### **X-1-3-2-1. Improving our knowledge about LOS R1 expression at exponential phase**

Our last investigations, at the cellular level, directed us to question about cell wall composition changes in *E. coli* bacteria during growth phase transition, knowing that late

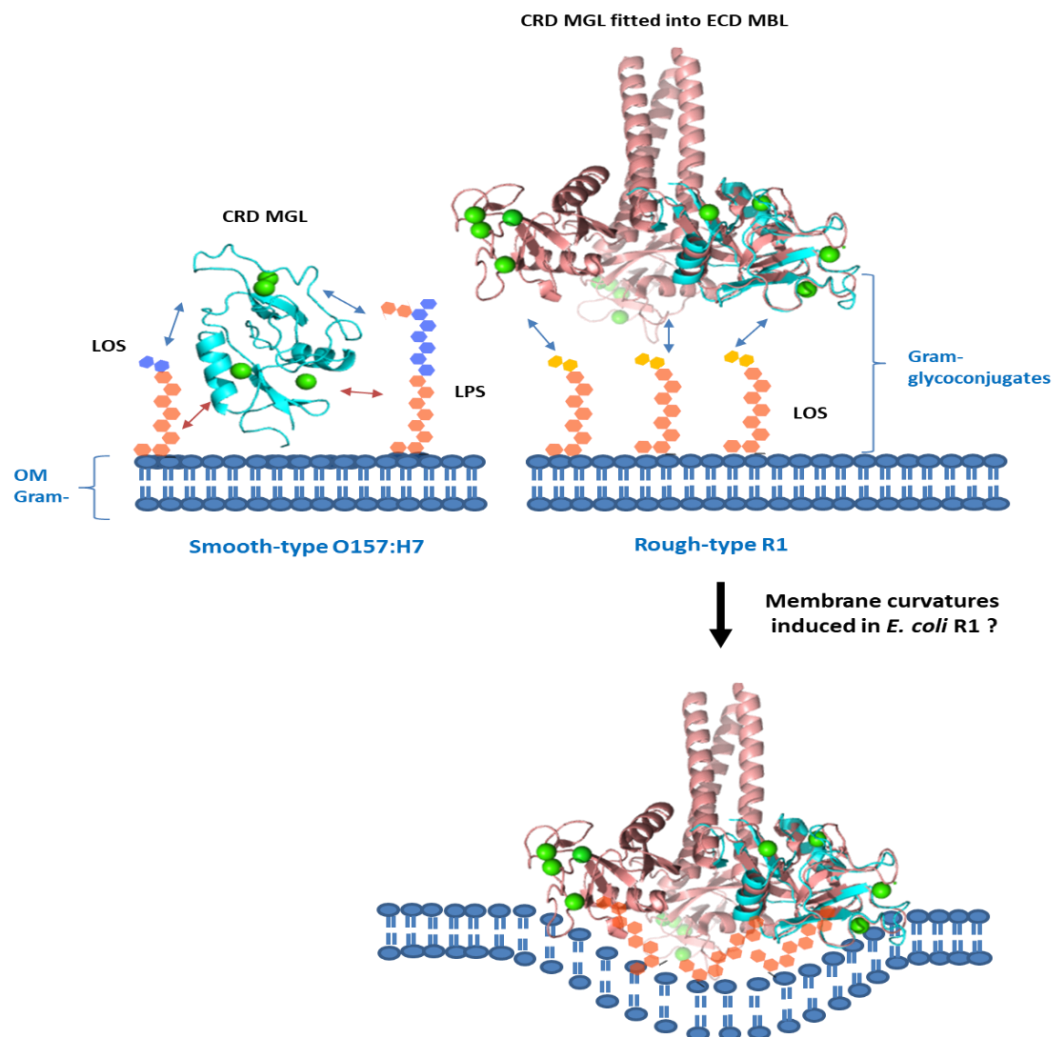
exponential phase and early stationary phase are growth phases during which infections take place in biological environment as it has been demonstrated for instance for *H. pylori* virulence gene expression to be increased during that growth stage<sup>374</sup>. We first wondered about the amount of LOS at the bacterial surface during that growth phase. On this regard, it has been found that LPS glycoconjugates (O-Antigen) could be largely expressed at stationary phase providing protection of cell membranes from the action of antimicrobial peptides (AMP)<sup>375</sup> and perhaps blocking the access of receptors to outer core regions of LOSs. Up to now, only few literatures have been published about genetical events taking place in *E. coli* R1 during bacterial growth. Do LOS R1 composition, expression, or production rates variate during *E. coli* R1 bacteria growth hence the growth phase dependency observed in the case of MGL-bound *E. coli* R1? This question is hard to answer since there are already many facts that influence LPS biosynthesis like temperature<sup>376</sup>, regulatory mechanisms<sup>377</sup> and growth medium composition (for instance in phosphate-limiting growth conditions where *E. coli* LPS composition underwent significant shifts)<sup>334</sup>.

To conclude, one may think that in such large interaction system, the outer membrane of *E. coli* R1 is decorated by a dense layer of LOS R1 exposed sugars thus favoring the strongest binding to human MGL. Endmost, are those the reasons why we observe MGL bound *E. coli* R1 bacteria exclusively at exponential phase? This question remains unanswered until having a proof of concept.

#### **X-1-4. Proposed Gram-negative membrane models: ECD and CRD MGL binding modes**

When it comes to bacterial surfaces, many facts must be considered including the high density of polysaccharides attached glycolipids. This sugar rich layer or “coating” is extended to the extracellular compartment, at the bacterial outer membrane. This would ease molecular contacts from soluble lectins (c.f. **CRD MGL**) with these glycoconjugates at different sites along the structures of LOS/LPS. During these interactions, monomeric CRD MGL might undergo clustering or self-association processes on the bound ligand favored by multiple bindings for example. Binding sites clustering is a key feature for macrophage galactose receptors and it has been studied by Jegouzo and colleagues to understand binding properties of this type of lectins<sup>378</sup>. Besides, ligand binding may be influenced either by lectin or carbohydrates clustering on biological surfaces<sup>379</sup>. On the other hand, **ECD MGL** oligomerization state stimulates the avidity of binding to glycoconjugates like LOS R1 (figure 105). The high density of sugar branched glycoconjugates may either change the apparent strong affinity to lower molar range due to steric hindrance or, it can favor increased multivalent bindings due to high local density of ECD

MGL at the OM. Consequently, the later might undergo membrane invaginations following MGL binding, similarly to what has been observed for Galectin-3 interaction with glycosphingolipids (GSLs)<sup>298</sup>.



**Figure 105. Overview of molecular interactions between CRD/ECD of human MGL and the studied LPS/LOS at the surface of Gram-negative bacteria.** The image shows essentially LPS and LOS as constituents of the bacterial outer membrane (OM) considering well-defined strains serotypes (rough- and smooth-type). The homology model of CRD MGL was fitted into the ECD of rat MBL (PDB 1KX0) given that the ECD MGL structure is unavailable. Blue and red arrows indicate glycoconjugates interactions taking place at the first and the second binding sites, respectively.

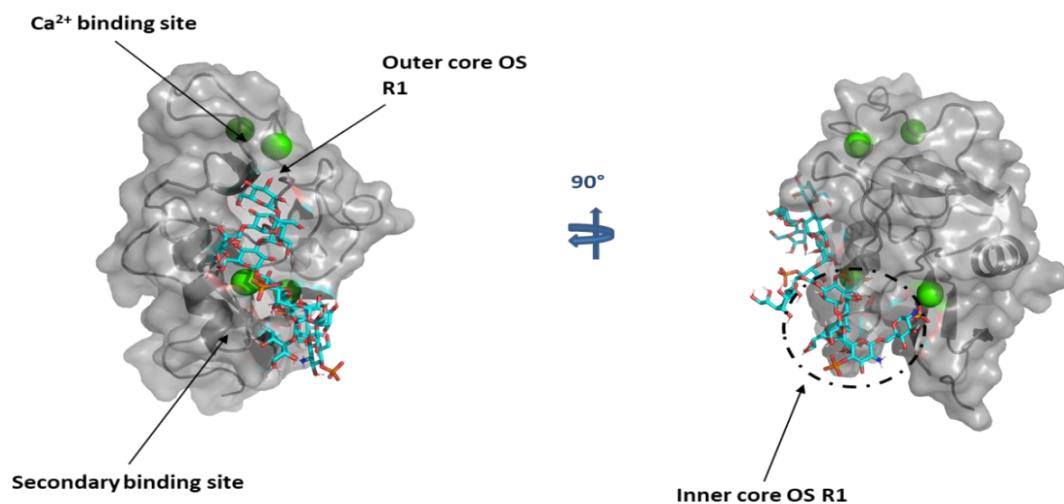
In nature, cell-membrane reorganization could occur following molecular interactions at cell surfaces given its properties to be remodelable and structurally stable at once. Here, our results showed that human MGL specifically binds to *E. coli* R1 glycoconjugates and that the binding affinity is believed to be in the nanomolar range. Again, the strength of LOS R1 binding (c.f. avidity) may be potentially correlated to an increase of MGL local density at the OM, like for DC-SIGN interaction with microbe cell-wall<sup>181</sup> and this might induce physical changes (surface

tensions) at the OM destabilizing the bacteria. If happening, membrane curvatures of *E. coli* R1 OM induced by human MGL would take place and they could be deeply assessed by using high-resolution confocal microscopy.

## X-2. Perspectives

### X-2-1. Mutagenesis studies as ensuing analyses to confirm binding sites on CRD MGL

NMR titrations on human CRD MGL allowed us to locate binding sites on the predicted model of CRD MGL and we observed a new binding surface engaged for *E. coli* ligands recognition. To gain more insights into this secondary binding site, mutagenesis analyses must be considered. To facilitate this task for future investigations, the OS R1 ligand was first docked into a model of CRD MGL (predicted from recently published X-ray structures) to better visualize binding area on the real CRD structure (figure 106). Accordingly, we prepared a preliminary list of potential mutations that would allow us to investigate the implications of each of the binding sites. We selected some residues, that are in proximity of the first binding site, to be mutated into Alanines without affecting the structural stability of the CRD, by also considering our results from CRD MGL mapping.



**Figure 106. OS R1 docking into CRD MGL structure predicted from PDB 6PY1.** Calcium ions are colored in green, OS R1 backbone in cyan with ball and stick representation and, CRD MGL in grey with surface and cartoon representation.

According to OS R1 molecular docking, it seems that there are possible binding contributions from the inner core, in addition to the outer core of OS R1. In the same line, the three tested *E. coli* ligands (OS R1, OS R3 and LOS O157: H7) share the same inner core structure and they exhibited CSPs that somehow affected the new secondary binding surface on the predicted CRD MGL structure (chapter VIII, figure 81). Binding contributions from inner core sugars like Heptose and phosphate groups have been confirmed for OS R3 interaction with mAb WN1 222-5

monoclonal antibody through ELISA, SPR and ITC analyses<sup>108,380</sup>. The authors observed a tight binding from all tested *E. coli* LPSs and they showed that binding of OS R1 outer core extremely enhanced the affinity of binding to mAb WN1 222-5 and that the fatty acids were not required for binding through testing deacetylated LOS versions.

Within this study, we succeeded to isolate, derivatize and study OS R1 and R3, in their native forms. Eventually, major and minor forms of *E. coli* OSs could be separated, via high performance chromatography techniques for example, for further analyses of LPS inner core substitutions effects on the interaction with human MGL, as a perspective study, in addition to mutagenesis studies. The involvement of *E. coli* LOS inner core in binding to the putative secondary site could be furthermore assessed by comparing data to another interaction model counting CRD MGL and BC J2315 LOS (or OS) by using NMR titrations knowing that the ability of this strain to escape from antimicrobial peptides is based on inner core substitutions<sup>381</sup>. Besides, this would aid further understanding MGL specificity to the tested *E. coli* ligands.

#### **X-2-2. Further investigations of the presumable ECD MGL conformation changes**

Our findings together with our knowledge on some CLRs binding properties, opened the question concerning human ECD MGL structural orientations on the cell membranes now that LOS R1 binding is believed to occur specifically: does this lectin adopt specific rearrangements (CRDs orientations) favouring either avidity-based mechanism or weak interactions, depending on *E. coli* glycoconjugates? To try to answer all these questions, at least partially, specific approaches, such as technique dedicated to membrane proteins, like FTIR spectroscopy based on recording vibrational spectra for probing the changes in the orientations of protein functional groups upon ligand binding, could be considered in the future<sup>382</sup>. Interestingly, this technique could be used in real-time by immobilizing the protein on the ATR (Attenuated Total Reflection) crystal and running the buffer with the ligand while recording the spectra.

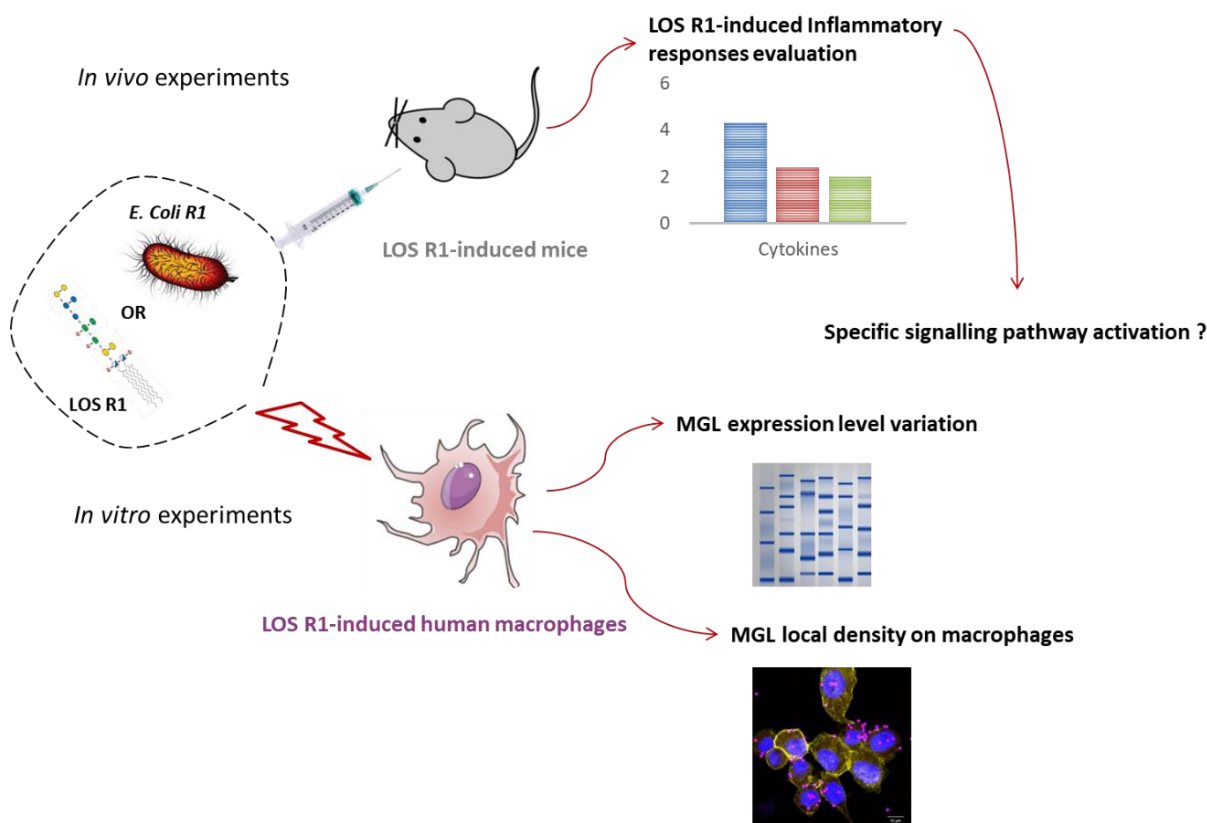
Another perspective method could be considered to verify our hypothesis regarding conformational changes of MGL upon ligands binding in solution would be to use DOSY-NMR spectroscopy<sup>383</sup>. The later would help in understanding MGL-*E. coli* ligands interactions by assessing molecular size changes (shifts from trimer to monomer for example) of the complexes, provided that human MGL structure is completely determined. In the same line, small-angle X-ray scattering (SAXS) would be a good tool for investigating lectin conformational changes provoked by large ligands binding, previously used for protein-peptides interactions<sup>384</sup>.

### **X-2-3. Further molecular investigations on LOS R1 for growth phase dependency validation**

A possible strategy that may be eventually considered for the analyses of LOS R1 expression and growth phase dependency of the interaction, would be the extraction of LOS R1 at each growth phase which will enable first their chemical analyses (including DOC-page analyses to determine the nature of glycoconjugates expressed at each stage of growth) and in parallel, their molecular interaction studies with human MGL, at the different growth phases. If this approach is successfully adapted to our scientific strategy, it would provide an in-depth set of comprehensive data regarding growth phase importance in *E. coli* pathogens recognition, from the cellular to the atomic level. As an additional experiment, *E. coli* O157:H7 bacteria recognition by human MGL could be also studied using fluorescence microscopy and cytometry.

### **X-2-4. Immunologic tests as perspectives to study immunity stimulation by *E. coli* R1**

Once the specific interaction between *E. coli* R1 bacteria (c.f. OS and LOS) and human MGL is confirmed, it is now important to enlarge the investigations to immunological tests. Our promising results would possibly open new perspectives concerning *E. coli* pathogens (R1 mutant) recognition by human MGL. In terms of cells analyses, we directed our attention particularly to the bacterial side of interaction, during the last period of the PhD project. *In vivo* studies could be considered, for the future, to evidence either pathogen recognition or immune tolerance of *E. coli* R1 by human MGL as a possible interesting target for scientific research.



**Figure 107. in vivo and in vitro experiments as future experiments for the evaluation of *E. coli* R1 immuno-stimulation.**

If applicable, immunological studies (figure 107) such as evaluating the control of expression level of human MGL by macrophages on LOS R1-induced cells and the local density of MGL at the cell surface (perhaps possible following MGL clustering) would be necessary to understand if and how MGL binding to LOS R1 is cellularly controlled.

In the same line, the expression level of DC-SIGN on LPS-induced cells as well as its local density at the surface of dendritic cells was followed within other studies<sup>181,326</sup> showing a great correlation between lectin expression and LPS (or microbial ligands) cells stimulation. Furthermore, employing immunological tests to follow cytokines and other immunity effectors production would help understanding innate immune response to LOS R1 possibly mediated by human MGL. In other words, human MGL functions to tolerate or to fight with *E. coli* R1 bacteria could be better defined eventually by performing in vitro and in vivo immunological essays.

### **X-2-5. Sugar-based drugs development as long-term perspectives**

Carbohydrate structures like OS R1 and R3 are less immunogenic than their macromolecular counterparts due to the absence of the Lipid-A portion. This would enable their use as therapeutic agents, for instance, for antibiotics to improve the efficiency of certain treatments



against bacterial infections, or as lectin inhibitors if immunological studies would prove human MGL-OSR1 involvement in innate and adaptive immunity.

In this PhD work, we showed that human MGL can recognize specifically OS R1 over OS R3 despite the fine differences in the outer core structures of these ligands. This would be very informative for further improvement of synthetic sugar-based drugs. Furthermore, if *E. coli* R1 immunogenicity will be confirmed through immunological tests, for example, our PhD work would provide many advances in the fields of bacterial infections therapeutics.

## ***Chapter XI.***

## XI. Materials and methods

### XI-1. Production, purification, and chemical analyses of LPSs samples

#### XI-1-1. Bacterial growth of *E. coli* R1 and R3 and *B. cenocepacia* J2315

*E. coli* R1(F470) and R3(F653) strains were generously provided from Prof. Sven Muller-Loennies from Research Center Borstel, Germany. Cells were grown, first as 5 mL precultures, in autoclaved Luria-Bertani (LB) medium (20 g/L) at 37°C with 90 rpm shaking overnight, for further spreading in Petri dishes. 1 colony per 5 mL of LB was then incubated overnight in the same conditions. The day after, *E. coli* R1 or R3 were inoculated at 0.5% in LB medium in the same conditions overnight. Bacteria were recovered by centrifugation 6000 rpm at 4 °C for 20 min. Then, we collected all the biomass and wash first with water (2 times) by centrifuge (8000, 20 min 4°C), then with ethanol twice and finally with acetone (5 mL) two times. Before lyophilization, bacterial pellets were well suspended in few mL of distilled water using the vortex, lyophilized, and finally dehydrated by lyophilization.

As for <sup>13</sup>C-<sup>15</sup>N LOSs (R1 and R3) production, the same protocol was used for *E. coli* cultures. LB medium was first used for pre-cultures to allow overgrowth of bacteria. Then, 10% of bacterial suspension were inoculated in autoclaved M9 minimal medium. The latter is composed of 0.5 g NaCl, 5.5 g Na<sub>2</sub>HPO<sub>4</sub> and 3 g KH<sub>2</sub>PO<sub>4</sub>, per liter of milli-Q water, supplemented with 2 g/L <sup>13</sup>C-glucose and 1 g/L <sup>15</sup>NH<sub>4</sub>Cl. Salts and vitamin mixtures (including FeCl<sub>3</sub>, MgSO<sub>4</sub>, MnCl<sub>2</sub>, CaCl<sub>2</sub>, ZnSO<sub>4</sub>, riboflavin, thiamine, biotin) were filtered through 0.2 µm filter pore and were added to the M9 medium before bacteria culture<sup>385</sup>.

*E. coli* O157:H7 Wild-type (ATCC43888) expressing LPS were grown in Prof. A. Polissi's laboratory in M9 minimum medium supplemented with 2 g/L <sup>13</sup>C-glucose.

#### XI-1-2. PCP extraction

PCP extraction was used for isotopically labelled and unlabelled LOSs R1 and R3 and LPS O157:H7 isolation from bacteria<sup>234</sup> and consists in the treatment of dried bacterial cells with phenol/chloroform/light petroleum in proportions 5:8:2 (v/v/v), respectively. 1g of dried bacteria was suspended in 20 mL of the PCP mixture at room temperature for 1 hour under the chemical hood. Mixtures were centrifuged afterwards (7000 rpm, 20 min, 4°C). This step was repeated 3 times for each bacterial strain. Supernatants were collected and subjected to rotary evaporation (at 48°C, pressure was gradually decreased from 400 to 70 mbar). The later step is useful for removal of volatile solvents such as chloroform and petroleum. The phenol phase will remain in the evaporation balloon. Phenol solution was then transferred to a falcon in ice and LOS/LPS was precipitated from pure phenol phase after adding 3-5 drops of water with gentle hand shaking.

The recovery of the isolated LPSs was then performed by centrifugation (8000 rpm, 30 min, 25°C). The sample was then washed 2 times with 85% aqueous phenol and acetone and lyophilized. In terms of yields, the PCP extraction provided 120 mg of LOS R1 per 14 Liters of *E. coli* R1, 84 mg of LOS R3 per 14 L of *E. coli* R3 and 10.8 mg of <sup>13</sup>C-LPS O157:H7 from 2 L of culture.

### **XI-1-3. Hot-phenol water extraction**

Dried cells (i.e. *E. coli* O157:H7 and *BC J2315*) were subjected to phenol water extraction<sup>236</sup>. Bacteria were suspended in the extraction mixture (90% phenol/water 1:1 (v/v)) at 65–70°C for 1 hour under stirring. The sample then was cooled at room temperature and centrifuged (7000 rpm, 4°C, 40 min) afterwards which led to the formation of three phases namely: water, milky and phenol fractions. Cell wall components are contained in the water fraction and they may be occasionally present in phenol phase. The extraction was repeated 3 times by the addition of same hot water volume and water phases were collected. Thereafter, lyophilized water and phenol phases derived from this method were extensively dialyzed (cut off 12-14 kDa) against distilled water and enzymatically treated in 100 mM Tris, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, buffer at pH 7.5, with DNase, RNase (2 mg/mL, 37°C, 5 h) and proteinase K (5 mg/mL, 56°C, 16 h) and again extensively dialyzed (cut off 12-14 kDa) against distilled water. In terms of yield, the Hot phenol extraction followed by enzymatic digestions provided 72 mg of LOS J2315 from 16L of *BC* cultures and, 55 mg of <sup>13</sup>C-LPS (phenol phase) extract from 2 Liters of *E. coli* O157:H7 culture.

### **XI-1-4. DOC-PAGE analysis**

For the detection of LPS/LOS following their extraction from bacteria, two electrophoresis methods were used<sup>239</sup>. 0.8 mg/mL (final concentration) of the isolated LOS/ LPS samples (dissolved in distilled water) were subjected to electrophoresis separation by sodium deoxycholate-polyacrylamide 13.5 % gel electrophoresis (DOC-PAGE) followed by silver nitrate staining. When the LPS/LOS migration is finished, the gel is washed using several solutions in the following order (with 50 rpm shaking in each step) : fixing solution (30 min-1h in 40% Et-OH, 5% acetic acid), oxidizing solution (30 min in 0.7% sodium meta-periodate (NaIO<sub>4</sub>) in 100 mL fixing solution), three 10 min washes in distilled water, staining solution (0.1% silver nitrate (Ag NO<sub>3</sub>) in 100 mL distilled water), one water wash and the gel is finally treated with development solution (3% Sodium carbonate Na<sub>2</sub>CO<sub>3</sub> + 0.02% 0.02% of formaldehyde ( with gentle shaking)) followed by one wash in 1% acetic acid solution.

### **XI-1-5. Methyl Glycoside and GC-MS analyses**

Sugars compositional analysis of LOS was carried upon the derivatization into Acetylated Methyl Glycosides<sup>235</sup>. 1 mg of LOS (dried under continuous vacuum) underwent Methanolysis in 500 µL

of methanoic hydrochloric acid HCl/MeOH solution (1.25 M) at 85°C for 16 h. Thereafter, fatty acids were separated by extraction in 500 µL of n-hexane solution. This step was repeated 3 times using centrifugation (1500 rpm, 10 min). The dried methanolic extract (monosaccharides) was then supplemented with MeOH solution (500 µL), 5 times with repeated amidst drying under air stream, to change the pH from acidic to neutral, required for further acetylation reaction. The dried sample underwent acetylation by adding 25 µL acetic anhydride Ac<sub>2</sub>O and 50 µL of pyridine for 30 minutes at 85°C. The resulting product was dried again under low air stream and purified by three chloroform: water extractions (1:2, v/v). Acetylated methyl glycosides and fatty acids fractions were dried afterwards and dissolved in 500 µL of pure acetone and hexane for monosaccharides and fatty acids, respectively. Thereafter, the samples were simultaneously subjected to GC-MS GLC analyses using HP gas chromatograph with Flame Ionization Detector and a column (Agilent Technologies, 30 m × 0.25 mm, film thickness 0.25 µm) of PhenylMethylSiloxane HP-5MS, with temperature program as follows: 150°C for 3 min, then 3°C per min to 280°C for 10 min.

#### **XI-1-6. Delipidation and chromatography purification of LOSs**

##### **XI-1-6-1. LOS de-*O*,de-*N*-acetylation and purification for *E. coli* glycoconjugates**

The removal of fatty acids from the isolated LOSs fractions (R1 and R3) was achieved through de-*O*-de-*N*-acetylation reactions. 68 mg of the isolated LOS (dried) underwent mild hydrazinolysis by adding 2 mL of methyl-hydrazine at 37°C under stirring for 2h. The mixture was cooled in a bath of ice and the de-*O*-acetylated LOS was precipitated by adding drops of cold acetone (1:1 v/v, methyl-hydrazine: acetone). The sample was centrifuged (30min, 4000 rpm, 4°C) and the supernatant was washed 3 times with successive centrifugations using acetone (to remove O-linked fatty acids). After that, the dried de-*O*-acetylated LOS was dissolved in 4M KOH solution at 120°C overnight under argon and then neutralized with 4M HCl solution. The resulting OS was present in the water phase, after chloroform: water (1:2, v/v) purification by centrifuging the mixture (45 min, 4000 rpm, 4°C) to remove N-linked fatty acids. Furthermore, the fully de-acetylated product (water phase) was purified by gel permeation chromatography to separate OSs from salts generated following the deacetylation reaction. The sample was loaded into G10 column in degassed distilled water and the resulting OS fraction (14.9 mg) was lyophilized and stored at -20°C.

##### **XI-1-6-2. Lipid-A removal and purification of *BC J2315* glycoconjugate**

72 mg of the isolated LOS J2315 sample were first loaded into Sephacryl S-300 column for removal of bacterial lipoproteins that may be co-extracted with LOS during the phenol/water extraction, into eluent (degas distilled water). The eluted LOS fraction (40 mg) underwent 1% acetic acid treatment at 100 °C for 5h. The sample is then centrifuged and the supernatant containing the OS was then washed twice with water (30 min, 4°C, 8000 rpm) and lyophilized. This reaction yielded 35.5 mg of OS and 4.5 mg of Lipid-A. The OS J2315 was loaded into P4 followed by P2 Biogel columns for additional purification. The final fraction of OS (2.9 mg) was solubilized in distilled water, lyophilized, and stored at -20°C.

<sup>1</sup>H-1D NMR experiments were recorded afterward to verify the purification level of the isolated and deacylated LOSs, in a cryoprobe-equipped Bruker 600 DRX spectrometer. 5 mg of each lyophilized OS sample (R1, R3 and J2315) were dissolved in 500 µL of D<sub>2</sub>O+2µL of Trimethyl Silyl Propanoate (TMS, as reference) and transferred into a 5 mm NMR tube.

#### **XI-1-7. Solid state NMR experiments on LPS O157:H7**

Before running solid state NMR experiments, 2 mg of lyophilized <sup>13</sup>C LPS O157:H7 (2.2 mg for <sup>15</sup>N <sup>13</sup>C LOS R3) was suspended in buffer 25 mM Tris HCl 150 mM NaCl 4 mM CaCl<sub>2</sub> pH 8 in H<sub>2</sub>O and then sedimented into solid state NMR Bruker rotor 1.3 mm by centrifugation in Beckman coulter sw32Ti support, at 49.2 \*10<sup>3</sup> g, for 16 hours for LPS (and at 16\*10<sup>3</sup> g, 10 min for LOS R3), at 8°C. The rotor containing the sample was then introduced into a 950 MHz Bruker Avance spectrometer equipped for <sup>1</sup>H <sup>13</sup>C <sup>15</sup>N and MAS 1.3 mm hardware with a magic rotation rate of 10-50\*10<sup>3</sup> Hz employed for 1D and 2D correlation experiments. The temperature was set to 260 K (300 K sample's temperature considering the heat generated from rotor rotation). For 1D-Cross-Polarisation experiments, <sup>13</sup>C detection was employed with 32 scans and 19 ms of acquisition time and 301 ppm spectral width at 25\*10<sup>3</sup> Hz of MAS rotation. For 2D-experiments, hCCH-INEPT together with 2D-hCH-INEPT and hCC-DARR had acquisition times of 18-50 ms with 64-96 scans, registered in spectral width of 15 ppm at 50 KHz rotation frequency, with a spectral width of 250 ppm for hCC-DARR. NMR data were processed and analyzed with Topspin 3.5 and CcpNmr analysis 2.4 software, respectively.

#### **XI-2. Production of CRD and ECD of human lectins**

The extracellular domain (**ECD**) of human lectins were produced and purified by our collaborators (F. Bulteau and M. Thepaut from Prof. F. Fieschi's laboratory at IBS institute) and the protocol used for **ECD MGL** production is described in the supplementary infos of <sup>344</sup>.

To produce isotopically labelled **CRD MGL**, the protein sequence [181-316] preceded by a linker (6His-Xa-5Gly) was recombinantly expressed from pET30 plasmid which was transformed in *E. coli* BL21 (DE3) competent cells (Agilent) in 1 mL LB medium supplemented with 50 ug/mL of Kanamycin, for further spreading in Petri dishes. Preculture (220 rpm, 37°C, 6h) was initiated with 1 isolated colony in 5 mL of LB medium in the same conditions. After growth, bacteria were inoculated at 5% (5mL of centrifuged LB culture per 100 mL of M9 medium supplemented with <sup>13</sup>C-glucose and <sup>15</sup>N-NH<sub>4</sub>Cl as isotopes as described in XI-1-1) in the same growth conditions overnight. The day after, grown bacterial precultures were inoculated at 10 % (10 mL/100 mL) in 3 liters of M9 minimum medium in the same conditions. At an OD<sub>600nm</sub> of 1.6, the overexpression of <sup>15</sup>N-<sup>13</sup>C-CRD MGL was induced by the addition of 1 mM IPTG for 3h at 37°C. Parenthetically, prior expression tests (in low culture volumes) were performed to select the optimal OD for induction (0.8 and 1.6 OD), and the highest CRD MGL production was observed at OD 1.6. Bacteria were harvested from cultures by centrifugation (6000 g, 20 min, 4°C). The expression of double-labelled <sup>15</sup>N-<sup>13</sup>C-CRD MGL was verified by SDS-PAGE electrophoresis on induced and non induced cell culture aliquots showing the CRD MGL corresponding band (~17 KDa).

The protein was then extracted by our collaborators through bacteria lysis followed by inclusion bodies solubilisation and refolding to obtain soluble <sup>13</sup>C-<sup>15</sup>N-CRD MGL. Our collaborators purified the soluble protein via Histag-purification together with ion exchange chromatography to eliminate dimers followed by affinity chromatography. Soluble <sup>15</sup>N-<sup>13</sup>C-CRD MGL purification yielded 4.66 mg of purified protein per liter of culture and was kept in buffer B at -80°C.

### **XI-3. Interaction studies**

#### **XI-3-1. Buffers**

The buffers used are the following:

**Buffer A:** 25mM Tris-d<sub>11</sub>, 4mM CaCl<sub>2</sub>, 150 mM NaCl pH 7 in D<sub>2</sub>O;

**Buffer B:** 25 mM Tris HCl 150 mM NaCl 4 mM CaCl<sub>2</sub> pH 8 in H<sub>2</sub>O 10 % D<sub>2</sub>O;

**Buffer C:** 50 mM Na<sub>2</sub>HPO<sub>4</sub>, 150 mM NaCl pH8 in H<sub>2</sub>O.

#### **XI-3-1-1. Buffer exchanges for LPSs and human lectins samples**

In some experimental conditions, buffer choice is limited and so, its exchange is required either for LPS or the lectins. LPS samples were exchanged from buffer B to C, prior to Electron

Microscopy EM experiments, by using centrifugation ( $16.9 \times 10^3$  g at  $4^\circ\text{C}$  for 30 min) in Amicon Ultra-4<sup>®</sup> concentrator tubes with 10 KDa cutoff filter.

Human lectins were also exchanged in some cases through P10 column chromatography or Vivaspin<sup>®</sup> concentrator tubes (10 KDa cutoff) from buffer B to A for STD NMR analyses, by 5-10 times centrifugation of 2 min at  $4^\circ\text{C}$  at  $5.5 \times 10^3$  g.

### **XI-3-2. NMR spectroscopy**

#### **XI-3-2-1. STD NMR spectroscopy for ECD MGL-OSs interaction system**

We first assigned the sugar resonances of the OS R1 by recording mono- and bi-dimensional NMR experiments as described<sup>344</sup>. After OS resonances assignments, 1D-STD-NMR experiments were carried out on OS R1 and OS R3 ligands in presence of 50  $\mu\text{M}$  of ECD MGL in buffer A. STD NMR experiments on the ECD MGL-OS R1 at 1:100 molar ratio (and simultaneously with OS R3) mixtures were recorded at  $25^\circ\text{C}$  on Bruker 600 Avance equipped with a cryoprobe. Control STD NMR experiments were performed on both protein and ligand in the free state at different irradiation frequencies, and by using the same experimental conditions set for the analysis of the mixture, to avoid artefacts due to false-positive STD signals. The on-resonance frequency was set at 7.5 ppm. More experimental details regarding the recorded NMR experiments as well as **computational studies (homology modelling and molecular docking)** are mentioned in the supplementary informations of our published article<sup>344</sup>.

#### **XI-3-2-2. NMR titrations for CRD MGL-*E. coli* ligands system**

##### **XI-3-2-2-1. 3D NMR experiments for backbone assignments**

For backbone resonance assignment of human CRD MGL, multi-dimensional NMR experiments were recorded. First, we transferred the sample containing 200  $\mu\text{M}$  of  $^{15}\text{N}$ - $^{13}\text{C}$  CRD of human MGL in buffer B into a shigemini NMR tube.  $^1\text{H}$ - $^{15}\text{N}$ -2D BTROSY and 3D-NMR experiments HNC(O), HN(CO)CACB, HNCACB and HN(CA)CO were registered at 303K on a 950MHz spectrometer with spectral width of 15 ppm for  $^1\text{H}$  dimension and 36 ppm for  $^{15}\text{N}$  dimension and 70 ppm for  $^{13}\text{C}$  dimension. Spectral width for  $^{13}\text{C}$  dimension was set at 15 ppm for HNC(O) and HN(CA)CO experiments. The resulting NMR spectra were processed using Bruker Topspin 3.5<sup>®</sup> and CcpNmr analysis 3.0<sup>®</sup> softwares.

##### **XI-3-2-2-2. NMR titrations experiments**



NMR titration experiments were performed on samples containing 50  $\mu\text{M}$  of  $^{15}\text{N}$ - $^{13}\text{C}$ -CRD MGL in buffer B at 300-303K on 850, 700 and 600 MHz Bruker Avance spectrometers equipped with cryoprobes HCN.  $^1\text{H}$ -1D and  $^1\text{H}$ - $^{15}\text{N}$ -2D NMR experiments were recorded for five titration points for each ligand. 100 ms of acquisition time for  $^1\text{H}$ -1D experiments with a spectral width of 25 ppm.  $^1\text{H}$ - $^{15}\text{N}$ -BEST-TROSY 2D experiments were recorded with spectral width of 12 ppm for the  $^1\text{H}$  dimension and 36 ppm for the  $^{15}\text{N}$  dimension, and with a selective excited pulse centered at 8.5 ppm with 5 ppm as excitation band width with 64-96 scans and a size of FID of 1452 for  $^1\text{H}$  and 256 for  $^{15}\text{N}$  and an acquisition time of 56 ms for  $^1\text{H}$  and 41 ms for  $^{15}\text{N}$ . The resulting NMR spectra were referenced to Sodium trimethylsilyl-propane-sulfonate (DSS) and processed using Bruker Topspin<sup>®</sup> 3.5 software.

To define the threshold values, standard deviation ( $\sigma$ ) values corresponding to each of the ligands tested by NMR titration experiments were calculated. Residues for which the shift change is 2 times higher than the standard deviation were considered as significantly perturbed residues. By using the following formula<sup>308</sup>:  $[\delta\text{H}^2 + (0.2*\delta\text{N})^2]^{1/2}$ , where  $\delta\text{H}$  and  $\delta\text{N}$  are the chemical shift changes in  $^1\text{H}$  and  $^{15}\text{N}$  dimensions in ppm and 0.2 is the chosen factor for  $^{15}\text{N}$  shift weighting, we defined the threshold values to be  $2*\sigma$  for each ligand. Chemical shift perturbations (CSP) and dissociation constants ( $K_d$ ) values were calculated using the CcpNmr analysis 3.0<sup>®</sup> software.

### **XI-3-3. BLI experiments**

#### **XI-3-3-1. ECD MGL coupling with Biotin**

For coupling carboxylate groups of ECD MGL to biotin, water-soluble 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and Biotin LC hydrazide EZ-Link<sup>™</sup> were used (ThermoFisher). ECD MGL at 0.9 mg/mL final concentration in buffer C is incubated with hydrazide LC-biotin (1.25 mM) and EDC (6.5 mM) at room temperature for 2h. After that, a centrifugation step ( $14*10^3$  g, 20 min, 4°C) was performed to remove any precipitates that might be formed. The sample was extensively dialyzed against PBS buffer (in GeBaFlex-tube 8KDa MWCO, over-night with agitating at 4°C) afterwards to ensure removal of reactive and final protein concentration was 12.7  $\mu\text{M}$ . The biotin-ECD MGL samples were then stored at -80°C.

#### **XI-3-3-2. Set up of BLI essays:**

BLI data were recorded in ForteBio<sup>®</sup> Octet System, octet RED 96e, with Streptavidin tips (ForteBio, Dip and Read<sup>™</sup> Biosensors) while samples were shaking at 25°C/1000 rpm to ensure homogenization data analyzed using Data analysis HT11.1.0.25 software. Protein immobilization

buffer was HBS-tween20 (20 mM HEPES, 150 mM NaCl, 0.02% tween pH 7.4). ECD MGL-biotin was loaded in 96-well plates at different concentrations (15-178 nM) to select the best one (53 nM) in terms of lectin immobilization, surface saturation and ligand binding. After ECD MGL-biotin immobilization on BLI tips, the latter were incubated for 100s in buffer B to supplement the sample with  $\text{Ca}^{2+}$  ions and next *E. coli* ligands (LOS R1, LOS R3, LPS O157: H7) were flowed in buffer C for 600s. The final cycle consisted of tips washes (ligand dissociation) for 600s and 3 regeneration cycles using 10 mM NaOH for 12s/cycle.

#### **XI-3-4. Electron microscopy**

##### **XI-3-4-1. LOS extrusion followed by DLS measurements**

Prior to EM experiments, the isolated LOSs R1 and R3 were subjected to extrusion experiment by following the manufacturer's protocol, using Avanti Polar® Lipids extruder, to homogenize their sizes. The protocol used to extrude isolated LOSs was the following : 500  $\mu\text{L}$  of LOS at 5 mg/mL final concentration in buffer C were first sonicated (5s on/5s off, 10 min, 60% amplitude) using an ultrasonic cell disruptor (Sonic Vibra Cell™ Branson probe). Thereafter, LOS samples were loaded into the extruder syringe and passed to the opposite syringe 11 times through 4 successively exchangeable bicarbonate membranes (Nuclepore™ Whatman Track-Etched) of 0.8  $\rightarrow$  0.4  $\rightarrow$  0.2  $\rightarrow$  0.1  $\mu\text{m}$  pore diameter sizes. 0.6 mg/mL final concentration of the resulting sample underwent DLS measurements which were performed on Dynapro Nanostar (Wyatt) instrument. DLS data were analysed using Dynamics® software.

##### **XI-3-4-2. EM experiments**

The morphological behaviors of LPS and LOS samples at 0.12 mg/mL final concentration, alone and in presence of 3.5  $\mu\text{M}$  of the human lectins, were assessed by Electron Microscopy in buffer C. The experiments were performed on the Instruct-ERIC platform at the Integrated Structural Biology of Grenoble (ISBG, UMS 3518), by using negative staining electron microscopy facilities. Briefly, Mica-carbon Flotation Technique (MFT) was used and, samples were absorbed to a carbon film on mica. The staining was done using 2% Sodium Silico Tungstate (SST) ( $\text{Na}_4\text{O}^4\text{SiW}_{12}$  [pH 7.0-7.5]) or 2% Uranyl Acetate (UrAc) ( $\text{UO}_2(\text{CH}_3\text{COO})_2 \cdot 2\text{H}_2\text{O}$  [pH 4.2-4.5]), and samples were afterwards transferred to a 400-mesh copper grid. The images were taken at a magnification of 23000x using a CCD Camera Gatan Orius 1000, under low dose conditions ( $<10 \text{ e}^-/\text{\AA}^2$ ) with defocus values between 1.2 and 2.5  $\mu\text{m}$  on a Tecnai 12 LaB6 electron microscope at 120 kV accelerating voltage.

## **XI-3-5. Fluorescence Microscopy and Flow Cytometry**

### **XI-3-5-1. ECD MGL labelling with Alexa<sup>647</sup>**

The labeling of ECD MGL was established according to the manufacturer's protocol, using Alexa fluor<sup>®</sup> 647 protein labelling kit A37573 from Molecular probes. 1.13 mg of ECD MGL (200  $\mu$ L at 5.66 mg/mL) were first exchanged through dialysis (in GeBaFlex-tube 8KDa MWCO, 6h), from buffer B to C (at pH 7) because the initial Tris buffer (buffer B) contains primary amine functions that could accidentally interact with the fluorophore. Then, MGL sample volume was adjusted to 500  $\mu$ L (2.26 mg/mL final concentration). To that volume, 56.6  $\mu$ L of sodium bicarbonate (1M) were added and the sample was transferred into the vial furnished by the kit containing 100  $\mu$ g of the fluorophore (Alexa Fluor 647 reactive dye). The latter has a succinimidyl ester NHS that interacts with primary amines of the lectin thus forming labelled-MGL. This fluorophore has maximum fluorescence emission at 665 nm and maximum absorbance at 650 nm. After 1h stirring at room temperature, ECD MGL<sup>A647</sup> sample was purified, through P10 column chromatography followed by dialysis against buffer C (3h, 4°C) to remove non reacted dye. 700  $\mu$ L of the dark blue fraction contained ECD MGL<sup>A647</sup> (1.23 mg/mL final concentration). The protein aliquots were then covered from light with aluminum paper and stored at -80°C.

### **XI-3-5-2. Fluorescence microscopy and flow cytometry experiments**

#### **XI-3-5-2-1. Epifluorescence microscopy analyses**

To perform single-cell essays, *E. coli* R1, R3 and B strains were cultured in LB medium at 37 °C, 220 rpm. Bacterial growth was followed by measuring ODs at the different growth phases. For *E. coli* R1 bacteria, four OD600nm were considered i.e. 0.3, 1.2 and 4 for the establishment of growth phase dependency fluorescence experiments. As for *E. coli* R3 and B strains, lag (OD 0.3) and exponential (OD 1.2) phases were evaluated.

All microorganisms recovered from exponential phase were incubated at  $19.2 \times 10^8$  cell/mL final concentration (concentrated two times) in buffer C with 2  $\mu$ M of human ECD MGL<sup>A647</sup> and from which samples were supplemented with 2 mM CaCl<sub>2</sub>. After 15 min incubation with gentle agitation at room temperature, all samples were washed 5 times by centrifugations (3000 g, 3min, 4°C) in buffer C to remove any exceed of fluorescence. Microscopy analyses were carried out immediately after samples preparations. 5  $\mu$ L of each mixture were deposited over a glass coverslip mounted on a slide and bacteria were observed using video-microscope (V-M4D, Olympus and Quorum Technologies) composed of an inverted Olympus IX83 fully motorized microscope equipped with an incubation chamber set at 25°C for image stability. Acquisition is

achieved with a sCMOS camera (Hamamtsu Orca Flash4). For MGL fluorescence detection, the excitation light (CoolLED-P4000) was set at 635 nm (Cy5 filter set with emission at 684/24 nm). Images were visualised and analysed using Volocity® and ImageJ® softwares.

#### **XI-3-5-2.2. Measurements of MGL-bound bacterial populations by flow cytometry**

In parallel, samples containing fluorescently labelled cells (MGL-bound *E. coli* R1) were diluted 10X in buffer C and fluorescent populations were quantified by flow-cytometer (MACSQuant-VYB, Miltenyi biotech) equipped with three lasers and 8 fluorescence channels. For side scattering measurements (SSC) and Alexa647 (MGL fluorophore) the excitation was set at 561 nm and the emission signal was detected at 692/75 nm (Cy5 filter set). This is an automated instrument optimized for fluorescent proteins and covers all the emission wavelengths used with microscopy. A minimum of 100 000 events were recorded for each sample, at a rate of 10 000 events per second. By selecting representative cells through group parameter HDRT, cell distributions were analyzed on that area using MacsQuant® software using dot plot representations. For cytometric statics of each condition, three indexes were considered: mean, median and modal. Index corresponds to the percentage of measured populations multiplied by the median of fluorescence value in order to quantify the variations between samples thus discarding artifacts.

For cell viability tests, we used Syto-9 (10 µM final concentration) and propidium iodide PI (60 µM final concentration) dyes furnished in the LIVE/DEAD™ BacLight™ Kit. The bacterial samples were incubated with the dyes at room temperature in the dark for 15 min before starting cytometry measurements and microscopy analyses. The maximum excitation/emission wavelength for Syto-9 is 485/498 nm for DNA detection and the laser filter was set at 488 nm as excitation wavelength (detected at 525/50 nm). As for PI, maximum excitation/emission wavelength are 535/617 nm and the laser set was at 561 nm (detected at 692/75 nm).

#### **XI-4. Technical acknowledgments**

The scientific results presented in this manuscript are the results of a collaboration among our groups in Grenoble and Naples (J.P. Simorre and A. Molinaro's groups), the group of F. Fieschi and the cell imaging platform at IBS institute (J.P Kleman and F. Lacroix group) at IBS in Grenoble. The associated topics in which we are interested revolve around protein-carbohydrates interactions for the PhD main research groups and, biochemistry and biophysics applied to lectins for Fieschi's group. The PhD project has received funding from the French Agence Nationale de

la Recherche (ANR) PIA for Glyco@- Alps (ANR-15-IDEX-02) and Ministero dell'Istruzione, dell'Università e della Ricerca Università degli Studi di Napoli Federico II.

This work used the platforms of the Grenoble Instruct-ERIC center (ISBG ; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). We thank Aline Le Roy and Christine Ebel, for assistance with access to the Protein Analysis On Line (PAOL) platform for DLS measurements, the SPR/BLI platform personal, Jean-Baptiste REISER Ph.D for help and assistance. The electron microscope facility is supported by the Rhône-Alpes Region, the Fondation Recherche Medicale (FRM), the fonds FEDER, the Centre National de la Recherche Scientifique (CNRS), the CEA, the University of Grenoble, EMBL, and the GIS-Infrastructures en Biologie Sante et Agronomie (IBISA). We thank D. Fenel and G. Schoehn, from the Electron Microscopy platform of the Integrated Structural Biology of Grenoble (ISBG, UMI3265).

## References

1. Roberts, C. A. & Buikstra, J. E. Bacterial infections. in *Ortner's Identification of Pathological Conditions in Human Skeletal Remains* (2019). doi:10.1016/B978-0-12-809738-0.00011-9
2. Galanos, C. & Freudenberg, M. A. Bacterial endotoxins: biological properties and mechanisms of action. *Mediators Inflamm.* (1993). doi:10.1155/S0962935193000687
3. Varki, A. *et al.* Chapter 39 Bacterial and Viral Infections. *Essentials Glycobiol.* (2009).
4. Rietschel, E. T. *et al.* Bacterial endotoxin: molecular relationships of structure to activity and function. *FASEB J.* (1994). doi:10.1096/fasebj.8.2.8119492
5. Kumar, S., Ingle, H., Prasad, D. V. R. & Kumar, H. Recognition of bacterial infection by innate immune sensors. *Critical Reviews in Microbiology* (2013). doi:10.3109/1040841X.2012.706249
6. Christian, G. Ueber die isolierte farbung der Schizomyceten in Schnitt und Trockenpreparaten (in German). *Fortschr Med* (1884).
7. Silhavy, T. J., Kahne, D. & Walker, S. The Bacterial Cell Envelope<sup>1</sup> T. J. Silhavy, D. Kahne and S. Walker, . *Cold Spring Harb Perspect Biol* (2010).
8. Kotani, S., Watanabe, Y., Shimono, T., Narita, T. & Kato, K. Immunoadjuvant activities of cell walls, their water soluble fractions and peptidoglycan subunits, prepared from various gram positive bacteria, and of synthetic N acetylmuramyl peptides. *Zeitschrift fur Immunitatsforsch.* (1975).
9. Viala, J. *et al.* Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nat. Immunol.* (2004). doi:10.1038/ni1131
10. Watanabe, T. *et al.* NOD1 contributes to mouse host defense against Helicobacter pylori via induction of type I IFN and activation of the ISGF3 signaling pathway. *J. Clin. Invest.* (2010). doi:10.1172/JCI39481
11. Krutzik, S. R. *et al.* Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat. Med.* (2003). doi:10.1038/nm864
12. Takeuchi, O., Hoshino, K. & Akira, S. Cutting Edge: TLR2-Deficient and MyD88-Deficient Mice Are Highly Susceptible to Staphylococcus aureus Infection . *J. Immunol.* (2000). doi:10.4049/jimmunol.165.10.5392
13. Clarke, T. B., Francella, N., Huegel, A. & Weiser, J. N. Invasive bacterial pathogens exploit TLR-mediated downregulation of tight junction components to facilitate translocation across the epithelium. *Cell Host Microbe* (2011). doi:10.1016/j.chom.2011.04.012
14. Ballou, S. P. & Lozanski, G. Induction of inflammatory cytokine release from cultured human monocytes by C-reactive protein. *Cytokine* (1992). doi:10.1016/1043-4666(92)90079-7
15. Yamada, Y. *et al.* Cytokines, acute phase proteins, and tissue injury. C-reactive protein opsonizes dead cells for debridement and stimulates cytokine production. in *Annals of the New York Academy of Sciences* (1990). doi:10.1111/j.1749-6632.1990.tb00176.x
16. Volanakis, J. E. & Kaplan, M. H. Specificity of C-Reactive Protein for Choline Phosphate Residues of Pneumococcal C-Polysaccharide (35323). *Proc. Soc. Exp. Biol. Med.* (1971). doi:10.3181/00379727-136-35323
17. Lehotzkya, R. E. *et al.* Molecular basis for peptidoglycan recognition by a bactericidal lectin. *Proc. Natl. Acad. Sci. U. S. A.* (2010). doi:10.1073/pnas.0909449107

18. Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* (80- ). (2006). doi:10.1126/science.1127119
19. Nagae, M. & Yamaguchi, Y. Sugar recognition and protein-protein interaction of mammalian lectins conferring diverse functions. *Current Opinion in Structural Biology* (2015). doi:10.1016/j.sbi.2015.08.005
20. Gay, N. J. & Gangloff, M. Structure and function of toll receptors and their ligands. *Annual Review of Biochemistry* (2007). doi:10.1146/annurev.biochem.76.060305.151318
21. Whitfield, C. Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annual Review of Biochemistry* (2006). doi:10.1146/annurev.biochem.75.103004.142545
22. Willis, L. M. & Whitfield, C. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydr. Res.* (2013). doi:10.1016/j.carres.2013.05.007
23. Burns, S. M. & Hull, S. I. Comparison of Loss of Serum Resistance by Defined Lipopolysaccharide Mutants and an Acapsular Mutant of Uropathogenic Escherichia coli O75:K5. *Infect. Immun.* (1998). doi:10.1128/iai.66.9.4244-4253.1998
24. Burns, S. M. & Hull, S. I. Loss of resistance to ingestion and phagocytic killing by O- and K- mutants of a uropathogenic Escherichia coli O75:K5 strain. *Infect. Immun.* (1999). doi:10.1128/iai.67.8.3757-3762.1999
25. Noel, G. J., Hoiseth, S. K. & Edelson, P. J. Type b capsule inhibits ingestion of haemophilus influenzae by murine macrophages: Studies with isogenic encapsulated and unencapsulated strains. *J. Infect. Dis.* (1992). doi:10.1093/infdis/166.1.178
26. Abdelhameed, A. S. *et al.* A glycoconjugate of Haemophilus influenzae Type b capsular polysaccharide with tetanus toxoid protein: Hydrodynamic properties mainly influenced by the carbohydrate. *Sci. Rep.* (2016). doi:10.1038/srep22208
27. Costerton, J. W., Stewart, P. S. & Greenberg, E. P. Bacterial biofilms: A common cause of persistent infections. *Science* (1999). doi:10.1126/science.284.5418.1318
28. Pamp, S. J., Gjermansen, M., Johansen, H. K. & Tolker-Nielsen, T. Tolerance to the antimicrobial peptide colistin in Pseudomonas aeruginosa biofilms is linked to metabolically active cells, and depends on the pmr and mexAB-oprM genes. *Mol. Microbiol.* (2008). doi:10.1111/j.1365-2958.2008.06152.x
29. Russell, S. L. *et al.* Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep.* (2012). doi:10.1038/embor.2012.32
30. Galdiero, E. *et al.* Eradication of Candida albicans persister cell biofilm by the membranotropic peptide gH625. *Sci. Rep.* (2020). doi:10.1038/s41598-020-62746-w
31. Limoli, D. H., Jones, C. J. & Wozniak, D. J. Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiol. Spectr.* (2015). doi:10.1128/microbiolspec.mb-0011-2014
32. Fey, P. D. & Olson, M. E. Current concepts in biofilm formation of Staphylococcus epidermidis. *Future Microbiology* (2010). doi:10.2217/fmb.10.56
33. Zapotoczna, M., O'Neill, E. & O'Gara, J. P. Untangling the Diverse and Redundant Mechanisms of Staphylococcus aureus Biofilm Formation. *PLoS Pathogens* (2016). doi:10.1371/journal.ppat.1005671

34. Otto, M. Staphylococcal biofilms. *Current Topics in Microbiology and Immunology* (2008). doi:10.1007/978-3-540-75418-3\_10
35. Mack, D. *et al.* Characterization of transposon mutants of biofilm-producing *Staphylococcus epidermidis* impaired in the accumulative phase of biofilm production: Genetic identification of a hexosamine-containing polysaccharide intercellular adhesin. *Infect. Immun.* (1994). doi:10.1128/iai.62.8.3244-3253.1994
36. Valle, J., Solano, C., García, B., Toledo-Arana, A. & Lasa, I. Biofilm switch and immune response determinants at early stages of infection. *Trends in Microbiology* (2013). doi:10.1016/j.tim.2013.05.008
37. Byrd, M. S. *et al.* Genetic and biochemical analyses of the *Pseudomonas aeruginosa* Psl exopolysaccharide reveal overlapping roles for polysaccharide synthesis enzymes in Psl and LPS production. *Mol. Microbiol.* (2009). doi:10.1111/j.1365-2958.2009.06795.x
38. Ma, L., Lu, H., Sprinkle, A., Parsek, M. R. & Wozniak, D. J. *Pseudomonas aeruginosa* Psl is a galactose- and mannose-rich exopolysaccharide. in *Journal of Bacteriology* (2007). doi:10.1128/JB.00620-07
39. Irie, Y. *et al.* Self-produced exopolysaccharide is a signal that stimulates biofilm formation in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* (2012). doi:10.1073/pnas.1217993109
40. Billings, N. *et al.* The Extracellular Matrix Component Psl Provides Fast-Acting Antibiotic Defense in *Pseudomonas aeruginosa* Biofilms. *PLoS Pathog.* (2013). doi:10.1371/journal.ppat.1003526
41. Friedman, L. & Kolter, R. Genes involved in matrix formation in *Pseudomonas aeruginosa* PA14 biofilms. *Mol. Microbiol.* (2004). doi:10.1046/j.1365-2958.2003.03877.x
42. Zhai, W., Wu, F., Zhang, Y., Fu, Y. & Liu, Z. The immune escape mechanisms of *Mycobacterium Tuberculosis*. *International Journal of Molecular Sciences* (2019). doi:10.3390/ijms20020340
43. Harding, C. V. & Boom, W. H. Regulation of antigen presentation by *Mycobacterium tuberculosis*: A role for Toll-like receptors. *Nature Reviews Microbiology* (2010). doi:10.1038/nrmicro2321
44. Van Crevel, R., Ottenhoff, T. H. M. & Van der Meer, J. W. M. Innate immunity to *Mycobacterium tuberculosis*. *Clinical Microbiology Reviews* (2002). doi:10.1128/CMR.15.2.294-309.2002
45. Van Crevel, R., Kleinnijenhuis, J., Oosting, M., Joosten, L. A. B. & Netea, M. G. Innate immune recognition of *mycobacterium tuberculosis*. *Clinical and Developmental Immunology* (2011). doi:10.1155/2011/405310
46. Fremont, C. M. *et al.* IL-1 Receptor-Mediated Signal Is an Essential Component of MyD88-Dependent Innate Response to *Mycobacterium tuberculosis* Infection. *J. Immunol.* (2007). doi:10.4049/jimmunol.179.2.1178
47. Means, T. K. *et al.* Differential Effects of a Toll-Like Receptor Antagonist on *Mycobacterium tuberculosis*-Induced Macrophage Responses. *J. Immunol.* (2001). doi:10.4049/jimmunol.166.6.4074
48. Bafica, A. *et al.* TLR9 regulates Th1 responses and cooperates with TLR2 in mediating optimal resistance to *Mycobacterium tuberculosis*. *J. Exp. Med.* (2005). doi:10.1084/jem.20051782
49. Pompei, L. *et al.* Disparity in IL-12 Release in Dendritic Cells and Macrophages in Response to *Mycobacterium tuberculosis* Is Due to Use of Distinct TLRs. *J. Immunol.* (2007). doi:10.4049/jimmunol.178.8.5192



50. Boro, M., Singh, V. & Balaji, K. N. Mycobacterium tuberculosis-triggered Hippo pathway orchestrates CXCL1/2 expression to modulate host immune responses. *Sci. Rep.* (2016). doi:10.1038/srep37695
51. Ferwerda, G. *et al.* NOD2 and toll-like receptors are nonredundant recognition systems of Mycobacterium tuberculosis. *PLoS Pathog.* (2005). doi:10.1371/journal.ppat.0010034
52. Khan, N. *et al.* Signaling through NOD-2 and TLR-4 Bolsters the T cell Priming Capability of Dendritic cells by Inducing Autophagy. *Sci. Rep.* (2016). doi:10.1038/srep19084
53. Gutierrez, M. G. *et al.* Autophagy is a defense mechanism inhibiting BCG and Mycobacterium tuberculosis survival in infected macrophages. *Cell* (2004). doi:10.1016/j.cell.2004.11.038
54. Xu, Y. *et al.* Toll-like Receptor 4 Is a Sensor for Autophagy Associated with Innate Immunity. *Immunity* (2007). doi:10.1016/j.immuni.2007.05.022
55. Decout, A. *et al.* Deciphering the molecular basis of mycobacteria and lipoglycan recognition by the C-type lectin Dectin-2. *Sci. Rep.* (2018). doi:10.1038/s41598-018-35393-5
56. Geijtenbeek, T. B. H. *et al.* Mycobacteria target DC-SIGN to suppress dendritic cell function. *J. Exp. Med.* (2003). doi:10.1084/jem.20021229
57. Feinberg, H. *et al.* Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor mincle. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M113.497149
58. Rambaruth, N. D. S., Jégouzo, S. A. F., Marlor, H., Taylor, M. E. & Drickamer, K. Mouse mincle: Characterization as a model for human mincle and evolutionary implications. *Molecules* (2015). doi:10.3390/molecules20046670
59. Kolmos, H. J. Panum's studies on 'putrid poison' 1856. An early description of endotoxin. *Dan. Med. Bull.* (2006).
60. Shear, M. J. Chemical treatment of tumors. v. isolation of the hemorrhage-producing fraction from serratia marcescens (bacillus prodigiosus) culture filtrate. *J. Natl. Cancer Inst.* (1943). doi:10.1093/jnci/4.1.81
61. Li, L. *et al.* Hydrogen sulfide is a novel mediator of lipopolysaccharide-induced inflammation in the mouse. *FASEB J.* (2005). doi:10.1096/fj.04-3583fje
62. Ngkelo, A., Meja, K., Yeadon, M., Adcock, I. & Kirkham, P. A. LPS induced inflammatory responses in human peripheral blood mononuclear cells is mediated through NOX4 and G $\alpha$  dependent PI-3kinase signalling. *J. Inflamm.* (2012). doi:10.1186/1476-9255-9-1
63. Abu-rish, E. Y. *et al.* Pregabalin inhibits in vivo and in vitro cytokine secretion and attenuates spleen inflammation in Lipopolysaccharide/Concanavalin A -induced murine models of inflammation. *Sci. Rep.* (2020). doi:10.1038/s41598-020-61006-1
64. Liu, Z. *et al.* Protective role of endothelial calpain knockout in lipopolysaccharide-induced acute kidney injury via attenuation of the p38-iNOS pathway and NO/ROS production. *Exp. Mol. Med.* (2020). doi:10.1038/s12276-020-0426-9
65. Funakoshi-Tago, M. *et al.* Pyrocatechol, a component of coffee, suppresses LPS-induced inflammatory responses by inhibiting NF- $\kappa$ B and activating Nrf2. *Sci. Rep.* (2020). doi:10.1038/s41598-020-59380-x
66. Alexander, C. & Rietschel, E. T. Bacterial lipopolysaccharides and innate immunity. *Journal of Endotoxin Research* (2001). doi:10.1179/096805101101532675

67. Neelamegham, S. *et al.* Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* (2019). doi:10.1093/glycob/cwz045
68. Kamio, Y. & Nikaido, H. Outer Membrane of Salmonella Typhimurium: Accessibility of Phospholipid Head Groups to Phospholipase C and Cyanogen Bromide Activated Dextran in the External Medium. *Biochemistry* (1976). doi:10.1021/bi00657a012
69. Silhavy, T. J., Ruiz, N. & Kahne, D. Advances in understanding bacterial outer-membrane biogenesis. *Nature Reviews Microbiology* (2006). doi:10.1038/nrmicro1322
70. Ruiz, N., Kahne, D. & Silhavy, T. J. Transport of lipopolysaccharide across the cell envelope: The long road of discovery. *Nature Reviews Microbiology* (2009). doi:10.1038/nrmicro2184
71. Bell, R. M., Mavis, R. D., Osborn, M. J. & Roy Vagelos, P. Enzymes of phospholipid metabolism: Localization in the cytoplasmic and outer membrane of the cell envelope of Escherichia coli and Salmonella typhimurium. *BBA - Biomembr.* (1971). doi:10.1016/0005-2736(71)90144-1
72. Osborn, M. J., Gander, J. E. & Parisi, E. Mechanism of Assembly of the Outer Membrane of Salmonella typhimurium. *J. Biol. Chem.* (1972). doi:10.1016/s0021-9258(19)45128-4
73. Kadrmas, J. L. & Raetz, C. R. H. Enzymatic Synthesis of Lipopolysaccharide in Escherichia coli. *J. Biol. Chem.* (1998). doi:10.1074/jbc.273.5.2799
74. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annual Review of Biochemistry* (2002). doi:10.1146/annurev.biochem.71.110601.135414
75. Heinrichs, D. E., Yethon, J. A. & Whitfield, C. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica. *Molecular Microbiology* (1998). doi:10.1046/j.1365-2958.1998.01063.x
76. Heinrichs, D. E., Yethon, J. A., Amor, P. A. & Whitfield, C. The assembly system for the outer core portion of R1- and R4-type lipopolysaccharides of Escherichia coli: The R1 core-specific  $\beta$ -glucosyltransferase provides a novel attachment site for O-polysaccharides. *J. Biol. Chem.* (1998). doi:10.1074/jbc.273.45.29497
77. Yasuhiro, K., Yuzuru, A. & Shoshichi, N. Composition and turnover of the phospholipids in Escherichia coli. *Biochim. Biophys. Acta (BBA)/Lipids Lipid Metab.* (1967). doi:10.1016/0005-2760(67)90167-1
78. Doerrler, W. T., Reedy, M. C. & Raetz, C. R. H. An Escherichia coli Mutant Defective in Lipid Export. *J. Biol. Chem.* (2001). doi:10.1074/jbc.C100091200
79. Tefsen, B., Bos, M. P., Beckers, F., Tommassen, J. & De Cock, H. MsbA is not required for phospholipid transport in Neisseria meningitidis. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M509026200
80. Ward, A., Reyes, C. L., Yu, J., Roth, C. B. & Chang, G. Flexibility in the ABC transporter MsbA: Alternating access with a twist. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0709388104
81. Padayatti, P. S. *et al.* Structural Insights into the Lipid A Transport Pathway in MsbA. *Structure* (2019). doi:10.1016/j.str.2019.04.007
82. Wu, T. *et al.* Identification of a protein complex that assembles lipopolysaccharide in the outer membrane of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* (2006). doi:10.1073/pnas.0604744103

83. Villa, R. *et al.* The Escherichia coli lpt transenvelope protein complex for lipopolysaccharide export is assembled via conserved structurally homologous domains. *J. Bacteriol.* (2013). doi:10.1128/JB.02057-12
84. Laguri, C. *et al.* Interaction of lipopolysaccharides at intermolecular sites of the periplasmic Lpt transport assembly. *Sci. Rep.* **7**, 9715 (2017).
85. Moura, E. C. C. M. *et al.* Thanatin Impairs Lipopolysaccharide Transport Complex Assembly by Targeting LptC–LptA Interaction and Decreasing LptA Stability. *Front. Microbiol.* (2020). doi:10.3389/fmicb.2020.00909
86. MACGREGOR, D. R. & ELLIKER, P. R. A comparison of some properties of strains of *Pseudomonas aeruginosa* sensitive and resistant to quaternary ammonium compounds. *Can. J. Microbiol.* (1958). doi:10.1139/m58-054
87. LEIVE, L. A NONSPECIFIC INCREASE IN PERMEABILITY IN ESCHERICHIA COLI PRODUCED BY EDTA. *Proc. Natl. Acad. Sci. U. S. A.* (1965). doi:10.1073/pnas.53.4.745
88. Leive, L. Release of lipopolysaccharide by EDTA treatment of *E. coli*. *Biochem. Biophys. Res. Commun.* (1965). doi:10.1016/0006-291X(65)90191-9
89. Leive, L. Actinomycin sensitivity in *Escherichia coli* produced by EDTA. *Biochem. Biophys. Res. Commun.* (1965). doi:10.1016/0006-291X(65)90874-0
90. Tamaki, S., Sato, T. & Matsushashi, M. Role of lipopolysaccharides in antibiotic resistance and bacteriophage adsorption of *Escherichia coli* K-12. *J. Bacteriol.* (1971). doi:10.1128/jb.105.3.968-975.1971
91. Nikaido, H. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiol. Mol. Biol. Rev.* (2003). doi:10.1128/mmbr.67.4.593-656.2003
92. May, K. L. & Grabowicz, M. The bacterial outer membrane is an evolving antibiotic barrier. *Proceedings of the National Academy of Sciences of the United States of America* (2018). doi:10.1073/pnas.1812779115
93. Wu, E. L. *et al.* Molecular dynamics and NMR spectroscopy studies of *E. coli* lipopolysaccharide structure and dynamics. *Biophys. J.* (2013). doi:10.1016/j.bpj.2013.08.002
94. Molinaro, A. *et al.* Chemistry of lipid a: At the heart of innate immunity. *Chemistry - A European Journal* (2015). doi:10.1002/chem.201403923
95. Silipo, A. & Molinaro, A. Lipid A Structure. in *Bacterial Lipopolysaccharides* (2011). doi:10.1007/978-3-7091-0733-1\_1
96. Isshiki, Y., Kawahara, K. & Zähringer, U. Isolation and characterisation of disodium (4-amino-4-deoxy- $\beta$ -L-arabinopyranosyl)-(18)-(D-glycero- $\alpha$ -D-talo-oct-2-ulopyranosylonate)-(24)-(methyl 3-deoxy-D-manno-oct-2-ulopyranosid)onate from the lipopolysaccharide of *Burkholderia cepacia*. *Carbohydr. Res.* (1998). doi:10.1016/S0008-6215(98)00179-7
97. Schwudke, D. *et al.* The obligate predatory *Bdellovibrio bacteriovorus* possesses a neutral lipid a containing  $\alpha$ -D-mannoses that replace phosphate residues. Similarities and differences between the lipid As and the lipopolysaccharides of the wild type strain *B. bacteriovorus* H. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M303012200
98. Hamad, M. A., Di Lorenzo, F., Molinaro, A. & Valvano, M. A. Aminoarabinose is essential for lipopolysaccharide export and intrinsic antimicrobial peptide resistance in *Burkholderia cenocepacia*. *Mol. Microbiol.* (2012). doi:10.1111/j.1365-2958.2012.08154.x

99. Vinogradov, E., Perry, M. B. & Conlan, J. W. Structural analysis of Francisella tularensis lipopolysaccharide. *Eur. J. Biochem.* (2002). doi:10.1046/j.1432-1033.2002.03321.x
100. Oikawa, M. *et al.* NMR conformational analysis of biosynthetic precursor-type lipid A: Monomolecular state and supramolecular assembly. *Org. Biomol. Chem.* (2004). doi:10.1039/b410544c
101. Lagrange, B. *et al.* Human caspase-4 detects tetra-acylated LPS and cytosolic Francisella and functions differently from murine caspase-11. *Nat. Commun.* (2018). doi:10.1038/s41467-017-02682-y
102. Clifton, L. A. *et al.* The Effect of Lipopolysaccharide Core Oligosaccharide Size on the Electrostatic Binding of Antimicrobial Proteins to Models of the Gram Negative Bacterial Outer Membrane. *Langmuir* (2016). doi:10.1021/acs.langmuir.6b00240
103. Holst, O. Structure of the Lipopolysaccharide Core Region. in *Bacterial Lipopolysaccharides* (2011). doi:10.1007/978-3-7091-0733-1\_2
104. Caroff, M. & Karibian, D. Structure of bacterial lipopolysaccharides. *Carbohydrate Research* (2003). doi:10.1016/j.carres.2003.07.010
105. Stevenson, G. *et al.* Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster. *J. Bacteriol.* (1994). doi:10.1128/jb.176.13.4144-4156.1994
106. Müller-Loennies, S., Lindner, B. & Brade, H. Structural Analysis of Oligosaccharides from Lipopolysaccharide (LPS) of Escherichia coli K12 Strain W3100 Reveals a Link between Inner and Outer Core LPS Biosynthesis. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M303985200
107. Silipo, A. & Molinaro, A. The diversity of the core oligosaccharide in lipopolysaccharides. *Subcell. Biochem.* (2010). doi:10.1007/978-90-481-9078-2\_4
108. Müller-Loennies, S., Brade, L. & Brade, H. Neutralizing and cross-reactive antibodies against enterobacterial lipopolysaccharide. *International Journal of Medical Microbiology* (2007). doi:10.1016/j.ijmm.2007.04.002
109. Nilsson, I. *et al.* Molecular characterization and verification of azido-3,8-dideoxy-D-manno-oct-2-ulosonic acid incorporation into bacterial lipopolysaccharide. *J. Biol. Chem.* (2017). doi:10.1074/jbc.M117.814962
110. Isshiki, Y., Zähringer, U. & Kawahara, K. Structure of the core-oligosaccharide with a characteristic D-glycero- $\alpha$ -D-talo-oct-2-ulosylonate-(2 $\rightarrow$ 4)-3-deoxy-D-manno-oct-2-ulosonate [ $\alpha$ -Ko-(2 $\rightarrow$ 4)-Kdo] disaccharide in the lipopolysaccharide from Burkholderia cepacia. *Carbohydr. Res.* (2003). doi:10.1016/j.carres.2003.07.005
111. Vinogradov, E., Korenevsky, A. & Beveridge, T. J. The structure of the core region of the lipopolysaccharide from Shewanella algae BrY, containing 8-amino-3,8-dideoxy-D-manno-oct-2-ulosonic acid. *Carbohydr. Res.* (2004). doi:10.1016/j.carres.2003.12.010
112. Holst, O. Chemical structure of the core region of lipopolysaccharides - An update. *Trends in Glycoscience and Glycotechnology* (2002). doi:10.4052/tigg.14.87
113. Vinogradov, E. & Perry, M. B. Structural analysis of the core region of lipopolysaccharides from Proteus mirabilis serotypes 06, 048 and 057. *Eur. J. Biochem.* (2000). doi:10.1046/j.1432-1327.2000.01262.x
114. Chaput, C., Spindler, E., Gill, R. T. & Zychlinsky, A. O-Antigen Protects Gram-Negative Bacteria from Histone Killing. *PLoS One* (2013). doi:10.1371/journal.pone.0071097

115. Thirumalapura, N. R., Morton, R. J., Ramachandran, A. & Malayer, J. R. Lipopolysaccharide microarrays for the detection of antibodies. *J. Immunol. Methods* (2005). doi:10.1016/j.jim.2005.01.004
116. Sadikot, R. T., Blackwell, T. S., Christman, J. W. & Prince, A. S. Pathogen-host interactions in pseudomonas aeruginosa pneumonia. *American Journal of Respiratory and Critical Care Medicine* (2005). doi:10.1164/rccm.200408-1044SO
117. Rocchetta, H. L., Burrows, L. L. & Lam, J. S. Genetics of O-Antigen Biosynthesis in Pseudomonas aeruginosa. *Microbiol. Mol. Biol. Rev.* (1999). doi:10.1128/mubr.63.3.523-553.1999
118. Makin, S. A. & Beveridge, T. J. The influence of A-band and B-band lipopolysaccharide on the surface characteristics and adhesion of Pseudomonas aeruginosa to surfaces. *Microbiology* (1996). doi:10.1099/13500872-142-2-299
119. Janeway, C. A., Travers, P., Walport, M. & Al, E. Principles of innate and adaptive immunity. *Immunobiol. Immune Syst. Heal. Dis. 5th Ed.* (2001).
120. Hato, T. & Dagher, P. C. How the innate immune system senses trouble and causes trouble. *Clin. J. Am. Soc. Nephrol.* (2015). doi:10.2215/CJN.04680514
121. Cooper, M. D. & Alder, M. N. The evolution of adaptive immune systems. *Cell* (2006). doi:10.1016/j.cell.2006.02.001
122. Dorak, M. T. Basic Immunology: Functions and Disorders of the Immune System. *Am. J. Epidemiol.* (2002). doi:10.1093/aje/155.2.185-a
123. Kawai, T. & Akira, S. The roles of TLRs, RLRs and NLRs in pathogen recognition. *International Immunology* (2009). doi:10.1093/intimm/dxp017
124. Kumar, H., Kawai, T. & Akira, S. Pathogen recognition in the innate immune response. *Biochemical Journal* (2009). doi:10.1042/BJ20090272
125. Fischer, H., Yamamoto, M., Akira, S., Beutler, B. & Svanborg, C. Mechanism of pathogen-specific TLR4 activation in the mucosa: Fimbriae, recognition receptors and adaptor protein selection. *Eur. J. Immunol.* (2006). doi:10.1002/eji.200535149
126. Lebeer, S., Vanderleyden, J. & De Keersmaecker, S. C. J. Host interactions of probiotic bacterial surface molecules: Comparison with commensals and pathogens. *Nature Reviews Microbiology* (2010). doi:10.1038/nrmicro2297
127. Salomao, R. *et al.* Bacterial sensing, cell signaling, and modulation of the immune response during sepsis. *Shock* (2012). doi:10.1097/SHK.0b013e318262c4b0
128. Santos, S. S. *et al.* Generation of nitric oxide and reactive oxygen species by neutrophils and monocytes from septic patients and association with outcomes. *Shock* (2012). doi:10.1097/SHK.0b013e318257114e
129. Monguió-Tortajada, M., Franquesa, M., Sarrias, M. R. & Borràs, F. E. Low doses of LPS exacerbate the inflammatory response and trigger death on TLR3-primed human monocytes article. *Cell Death Dis.* (2018). doi:10.1038/s41419-018-0520-2
130. Yücel, G. *et al.* Lipopolysaccharides induced inflammatory responses and electrophysiological dysfunctions in human-induced pluripotent stem cell derived cardiomyocytes. *Sci. Rep.* (2017). doi:10.1038/s41598-017-03147-4
131. Akira, S. & Takeda, K. Toll-like receptor signalling. *Nature Reviews Immunology* (2004).

doi:10.1038/nri1391

132. Zamyatina, A. & Heine, H. Lipopolysaccharide Recognition in the Crossroads of TLR4 and Caspase-4/11 Mediated Inflammatory Pathways. *Frontiers in Immunology* (2020). doi:10.3389/fimmu.2020.585146
133. Cochet, F. & Peri, F. The role of carbohydrates in the lipopolysaccharide (LPS)/toll-like receptor 4 (TLR4) Signalling. *International Journal of Molecular Sciences* (2017). doi:10.3390/ijms18112318
134. Iijima, J. *et al.* Core fucose is critical for CD14-dependent Toll-like receptor 4 signaling. *Glycobiology* (2017). doi:10.1093/glycob/cwx075
135. Ryu, J. K. *et al.* Reconstruction of LPS Transfer Cascade Reveals Structural Determinants within LBP, CD14, and TLR4-MD2 for Efficient LPS Recognition and Transfer. *Immunity* (2017). doi:10.1016/j.immuni.2016.11.007
136. Kim, H. M. *et al.* Crystal Structure of the TLR4-MD-2 Complex with Bound Endotoxin Antagonist Eritoran. *Cell* (2007). doi:10.1016/j.cell.2007.08.002
137. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* (2009). doi:10.1038/nature07830
138. Erridge, C., Bennett-Guerrero, E. & Poxton, I. R. Structure and function of lipopolysaccharides. *Microbes and Infection* (2002). doi:10.1016/S1286-4579(02)01604-0
139. Gioannini, T. L. *et al.* Isolation of an endotoxin-MD-2 complex that produces Toll-like receptor 4-dependent cell activation at picomolar concentrations. *Proc. Natl. Acad. Sci. U. S. A.* (2004). doi:10.1073/pnas.0306906101
140. Hirschfeld, M., Ma, Y., Weis, J. H., Vogel, S. N. & Weis, J. J. Cutting Edge: Repurification of Lipopolysaccharide Eliminates Signaling Through Both Human and Murine Toll-Like Receptor 2. *J. Immunol.* (2000). doi:10.4049/jimmunol.165.2.618
141. Werts, C. *et al.* Leptospiral lipopolysaccharide activates cells through a TLR2-dependent mechanism. *Nat. Immunol.* (2001). doi:10.1038/86354
142. Schumann, R. R. *et al.* Lipopolysaccharide activates caspase-1 (interleukin-1-converting enzyme) in cultured monocytic and endothelial cells. *Blood* (1998). doi:10.1182/blood.v91.2.577
143. Hagar, J. A., Powell, D. A., Aachoui, Y., Ernst, R. K. & Miao, E. A. Cytoplasmic LPS activates caspase-11: Implications in TLR4-independent endotoxic shock. *Science* (80-. ). (2013). doi:10.1126/science.1240988
144. Kayagaki, N. *et al.* Noncanonical inflammasome activation by intracellular LPS independent of TLR4. *Science* (80-. ). (2013). doi:10.1126/science.1240248
145. Shi, J. *et al.* Inflammatory caspases are innate immune receptors for intracellular LPS. *Nature* (2014). doi:10.1038/nature13683
146. Viganò, E. *et al.* Human caspase-4 and caspase-5 regulate the one-step non-canonical inflammasome activation in monocytes. *Nat. Commun.* (2015). doi:10.1038/ncomms9761
147. An, J. *et al.* Caspase-4 disaggregates lipopolysaccharide micelles via LPS-CARD interaction. *Sci. Rep.* (2019). doi:10.1038/s41598-018-36811-4
148. Yi, Y. S. Caspase-11 non-canonical inflammasome: a critical sensor of intracellular lipopolysaccharide in macrophage-mediated inflammatory responses. *Immunology* (2017). doi:10.1111/imm.12787

149. Boyd, W. C. & Shapleigh, E. Specific precipitating activity of plant agglutinins (lectins). *Science* (80- ). (1954). doi:10.1126/science.119.3091.419
150. Barondes, S. H. Soluble lectins: A new class of extracellular proteins. *Science* (1984). doi:10.1126/science.6367039
151. Varki A, Cummings RD, E. J. Glycosylation changes in cancer, *Essentials of Glycobiology. Essentials of Glycobiology* (2009).
152. Brudner, M. *et al.* Lectin-Dependent Enhancement of Ebola Virus Infection via Soluble and Transmembrane C-type Lectin Receptors. *PLoS One* (2013). doi:10.1371/journal.pone.0060838
153. Teillet, F. *et al.* The Two Major Oligomeric Forms of Human Mannan-Binding Lectin: Chemical Characterization, Carbohydrate-Binding Properties, and Interaction with MBL-Associated Serine Proteases. *J. Immunol.* (2005). doi:10.4049/jimmunol.174.5.2870
154. Boldt, D. H. & Dorsey, S. A. WHEAT GERM AGGLUTININ DECREASES EXPRESSION OF THE t8 ANTIGEN ON HUMAN PERIPHERAL MONONUCLEAR CELLS. *Immunol. Invest.* (1984). doi:10.3109/08820138409061303
155. Drickamer, K. & Taylor, M. E. Recent insights into structures and functions of C-type lectins in the immune system. *Current Opinion in Structural Biology* (2015). doi:10.1016/j.sbi.2015.06.003
156. Mnich, M. E., van Dalen, R. & van Sorge, N. M. C-Type Lectin Receptors in Host Defense Against Bacterial Pathogens. *Frontiers in Cellular and Infection Microbiology* (2020). doi:10.3389/fcimb.2020.00309
157. Dong, M. *et al.* Conformational Changes in Mannan-Binding Lectin Bound to Ligand Surfaces. *J. Immunol.* (2007). doi:10.4049/jimmunol.178.5.3016
158. Brown, J. *et al.* Structure of the fungal  $\beta$ -glucan-binding immune receptor dectin-1: Implications for function. *Protein Sci.* (2007). doi:10.1110/ps.072791207
159. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* (2000). doi:10.1016/S0092-8674(00)00138-0
160. Neumann, K. *et al.* Clec12a is an inhibitory receptor for uric acid crystals that regulates inflammation in response to cell death. *Immunity* (2014). doi:10.1016/j.immuni.2013.12.015
161. Zelensky, A. N. & Gready, J. E. The C-type lectin-like domain superfamily. *FEBS Journal* (2005). doi:10.1111/j.1742-4658.2005.05031.x
162. Drickamer, K. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* (1992). doi:10.1038/360183a0
163. Powlesland, A. S. *et al.* Widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins. *J. Biol. Chem.* (2006). doi:10.1074/jbc.M601925200
164. de Witte, L., Abt, M., Schneider-Schaulies, S., van Kooyk, Y. & Geijtenbeek, T. B. H. Measles Virus Targets DC-SIGN To Enhance Dendritic Cell Infection. *J. Virol.* (2006). doi:10.1128/jvi.80.7.3477-3486.2006
165. Brown, G. D., Willment, J. A. & Whitehead, L. C-type lectins in immunity and homeostasis. *Nature Reviews Immunology* (2018). doi:10.1038/s41577-018-0004-8
166. Soilleux, E. J. *et al.* Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations in situ and in vitro. *J. Leukoc. Biol.* (2002). doi:10.1189/jlb.1106655

167. Pöhlmann, S. *et al.* DC-SIGNR, a DC-SIGN homologue expressed in endothelial cells, binds to human and simian immunodeficiency viruses and activates infection in trans. *Proc. Natl. Acad. Sci. U. S. A.* (2001). doi:10.1073/pnas.051631398
168. Tabarani, G. *et al.* DC-SIGN neck domain is a pH-sensor controlling oligomerization. Sxns and hydrodynamic studies of extracellular domain. *J. Biol. Chem.* (2009). doi:10.1074/jbc.M109.021204
169. Mitchell, D. A., Fadden, A. J. & Drickamer, K. A novel mechanism of carbohydrate recognition by the C-type lectins DC-SIGN and DC-SIGNR. Subunit organization and binding to multivalent ligands. *J. Biol. Chem.* **276**, 28939–28945 (2001).
170. Van Kooyk, Y. & Geijtenbeek, T. B. H. DC-SIGN: Escape mechanism for pathogens. *Nature Reviews Immunology* (2003). doi:10.1038/nri1182
171. Geijtenbeek, T. B. H. *et al.* Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. *Cell* (2000). doi:10.1016/S0092-8674(00)80693-5
172. Khoo, U. S., Chan, K. Y. K., Chan, V. S. F. & Lin, C. L. S. DC-SIGN and L-SIGN: The SIGNs for infection. *Journal of Molecular Medicine* (2008). doi:10.1007/s00109-008-0350-2
173. García-Vallejo, J. J. *et al.* DC-SIGN mediates adhesion and rolling of dendritic cells on primary human umbilical vein endothelial cells through LewisY antigen expressed on ICAM-2. *Mol. Immunol.* (2008). doi:10.1016/j.molimm.2007.11.001
174. Geurtsen, J., Driessen, N. N. & Appelmelk, B. J. Mannose-Fucose Recognition by DC-SIGN. in *Microbial Glycobiology* (2010). doi:10.1016/B978-0-12-374546-0.00034-1
175. Yuriev, E., Farrugia, W., Scott, A. M. & Ramsland, P. A. Three-dimensional structures of carbohydrate determinants of Lewis system antigens: Implications for effective antibody targeting of cancer. *Immunology and Cell Biology* (2005). doi:10.1111/j.1440-1711.2005.01374.x
176. Pederson, K., Mitchell, D. A. & Prestegard, J. H. Structural characterization of the DC-SIGN-LewisX complex. *Biochemistry* (2014). doi:10.1021/bi5005014
177. Guo, Y. *et al.* Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR. *Nat. Struct. Mol. Biol.* (2004). doi:10.1038/nsmb784
178. Reina, J. J. *et al.* Docking, synthesis, and NMR studies of mannosyl trisaccharide ligands for DC-SIGN lectin. *Org. Biomol. Chem.* (2008). doi:10.1039/b802144a
179. Daniel Martínez, J. *et al.* Unraveling sugar binding modes to DC-SIGN by employing fluorinated carbohydrates. *Molecules* (2019). doi:10.3390/molecules24122337
180. Pitarque, S. *et al.* Deciphering the molecular bases of Mycobacterium tuberculosis binding to the lectin DC-SIGN reveals an underestimated complexity. *Biochem. J.* (2005). doi:10.1042/BJ20050709
181. Itano, M. S. *et al.* Super-resolution imaging of C-type lectin spatial rearrangement within the dendritic cell plasma membrane at fungal microbe contact sites. *Front. Phys.* (2014). doi:10.3389/fphy.2014.00046
182. Geijtenbeek, T. B. H. *et al.* Identification of different binding sites in the dendritic cell-specific receptor DC-SIGN for intercellular adhesion molecule 3 and HIV-1. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M111532200



183. Simmons, G. *et al.* DC-SIGN and DC-SIGNR bind Ebola glycoproteins and enhance infection of macrophages and endothelial cells. *Virology* (2003). doi:10.1006/viro.2002.1730
184. Yang, Z.-Y. *et al.* pH-Dependent Entry of Severe Acute Respiratory Syndrome Coronavirus Is Mediated by the Spike Glycoprotein and Enhanced by Dendritic Cell Transfer through DC-SIGN. *J. Virol.* (2004). doi:10.1128/jvi.78.11.5642-5650.2004
185. Lozach, P. Y. *et al.* C-type lectins L-SIGN and DC-SIGN capture and transmit infectious hepatitis C virus pseudotype particles. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M402296200
186. Mason, C. P. & Tarr, A. W. Human lectins and their roles in viral infections. *Molecules* (2015). doi:10.3390/molecules20022229
187. Alvarez, C. P. *et al.* C-Type Lectins DC-SIGN and L-SIGN Mediate Cellular Entry by Ebola Virus in cis and in trans. *J. Virol.* (2002). doi:10.1128/jvi.76.13.6841-6844.2002
188. Valverde, P. *et al.* Molecular Insights into DC-SIGN Binding to Self-Antigens: The Interaction with the Blood Group A/B Antigens. *ACS Chem. Biol.* **14**, 1660–1671 (2019).
189. Gringhuis, S. I. *et al.* C-Type Lectin DC-SIGN Modulates Toll-like Receptor Signaling via Raf-1 Kinase-Dependent Acetylation of Transcription Factor NF- $\kappa$ B. *Immunity* (2007). doi:10.1016/j.immuni.2007.03.012
190. Raes, G. *et al.* Macrophage galactose-type C-type lectins as novel markers for alternatively activated macrophages elicited by parasitic infections and allergic airway inflammation. *J. Leukoc. Biol.* (2005). doi:10.1189/jlb.0304212
191. van Vliet, S. J., Paessens, L. C., Broks-van den Berg, V. C. M., Geijtenbeek, T. B. H. & van Kooyk, Y. The C-Type Lectin Macrophage Galactose-Type Lectin Impedes Migration of Immature APCs. *J. Immunol.* (2008). doi:10.4049/jimmunol.181.5.3148
192. Sato, K. *et al.* Lack of antigen-specific tissue remodeling in mice deficient in the macrophage galactose-type calcium-type lectin 1/CD301a. *Blood* (2005). doi:10.1182/blood-2004-12-4943
193. Sato, K. *et al.* Redistributions of macrophages expressing the macrophage galactose-type C-type lectin (MGL) during antigen-induced chronic granulation tissue formation. *Int. Immunol.* (2005). doi:10.1093/intimm/dxh235
194. Higashi, N. *et al.* The macrophage C-type lectin specific for galactose/N-acetylgalactosamine is an endocytic receptor expressed on monocyte-derived immature dendritic cells. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M202104200
195. Valladeau, J. *et al.* Immature Human Dendritic Cells Express Asialoglycoprotein Receptor Isoforms for Efficient Receptor-Mediated Endocytosis. *J. Immunol.* (2001). doi:10.4049/jimmunol.167.10.5767
196. Iida, S. I., Yamamoto, K. & Irimura, T. Interaction of human macrophage C-type lectin with O-linked N-acetylgalactosamine residues on mucin glycopeptides. *J. Biol. Chem.* (1999). doi:10.1074/jbc.274.16.10697
197. Nagae, M. & Yamaguchi, Y. Structural Aspects of Carbohydrate Recognition Mechanisms of C-Type Lectins. in *Current Topics in Microbiology and Immunology* (2020). doi:10.1007/82\_2019\_181
198. Tsuiji, M. *et al.* Molecular cloning and characterization of a novel mouse macrophage C-type lectin, mMGL2, which has a distinct carbohydrate specificity from mMGL1. *J. Biol. Chem.* (2002). doi:10.1074/jbc.M203774200

199. Kenneth Hooper, J. Asgr1 and its enigmatic relative, CLEC10A. *International Journal of Molecular Sciences* (2020). doi:10.3390/ijms21144818
200. Suzuki, N., Yamamoto, K., Toyoshima, S., Osawa, T. & Irimura, T. Molecular cloning and expression of cDNA encoding human macrophage C-type lectin. Its unique carbohydrate binding specificity for Tn antigen. *J. Immunol.* (1996).
201. van Sorge, N. M. *et al.* N-glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL. *Cell. Microbiol.* (2009). doi:10.1111/j.1462-5822.2009.01370.x
202. Van Vliet, S. J. *et al.* Variation of *Neisseria gonorrhoeae* lipooligosaccharide directs dendritic cell-induced T helper responses. *PLoS Pathog.* (2009). doi:10.1371/journal.ppat.1000625
203. Marcelo, F. *et al.* Delineating binding modes of Gal/GalNAc and structural elements of the molecular recognition of tumor-associated mucin glycopeptides by the human macrophage galactose-type lectin. *Chem. - A Eur. J.* (2014). doi:10.1002/chem.201404566
204. Marcelo, F. *et al.* Identification of a secondary binding site in human macrophage galactose-type lectin by microarray studies: Implications for the molecular recognition of its ligands. *J. Biol. Chem.* (2019). doi:10.1074/jbc.RA118.004957
205. van Vliet, S. J., Gringhuis, S. I., Geijtenbeek, T. B. H. & van Kooyk, Y. Regulation of effector T cells by antigen-presenting cells via interaction of the C-type lectin MGL with CD45. *Nat. Immunol.* (2006). doi:10.1038/ni1390
206. Napoletano, C. *et al.* Tumor-associated Tn-MUC1 glycoform is internalized through the macrophage galactose-type C-type lectin and delivered to the HLA class I and II compartments in dendritic cells. *Cancer Res.* (2007). doi:10.1158/0008-5472.CAN-07-1035
207. van Dalen, R. *et al.* Langerhans cells sense *Staphylococcus aureus* wall teichoic acid through langerin to induce inflammatory responses. *MBio* (2019). doi:10.1128/mBio.00330-19
208. Muñoz-García, J. C. *et al.* Langerin-heparin interaction: Two binding sites for small and large ligands as revealed by a combination of NMR spectroscopy and cross-linking mapping experiments. *J. Am. Chem. Soc.* (2015). doi:10.1021/ja511529x
209. Feinberg, H. *et al.* Structural basis for langerin recognition of diverse pathogen and mammalian glycans through a single binding site. *J. Mol. Biol.* (2011). doi:10.1016/j.jmb.2010.11.039
210. Dulal, H. P., Adachi, Y., Ohno, N. & Yamaguchi, Y.  $\beta$ -Glucan-induced cooperative oligomerization of Dectin-1 C-type lectin-like domain. *Glycobiology* (2018). doi:10.1093/glycob/cwy039
211. Bugarcic, A. *et al.* Human and mouse macrophage-inducible C-type lectin (Mincle) bind *Candida albicans*. *Glycobiology* (2008). doi:10.1093/glycob/cwn046
212. Ito, T. *et al.* A CD1a+/CD11c+ subset of human blood dendritic cells is a direct precursor of Langerhans cells. *J. Immunol.* (1999).
213. De Witte, L. *et al.* Langerin is a natural barrier to HIV-1 transmission by Langerhans cells. *Nat. Med.* (2007). doi:10.1038/nm1541
214. Yang, K. *et al.* Host Langerin (CD207) is a receptor for *Yersinia pestis* phagocytosis and promotes dissemination. *Immunol. Cell Biol.* (2015). doi:10.1038/icb.2015.46
215. de Jong, M. A. W. P. *et al.* C-type lectin Langerin is a  $\beta$ -glucan receptor on human Langerhans cells that recognizes opportunistic and pathogenic fungi. *Mol. Immunol.* (2010).

doi:10.1016/j.molimm.2009.12.016

216. Valverde, P., Martínez, J. D., Cañada, F. J., Ardá, A. & Jiménez-Barbero, J. Molecular Recognition in C-Type Lectins: The Cases of DC-SIGN, Langerin, MGL, and L-Sectin. *ChemBioChem* (2020). doi:10.1002/cbic.202000238
217. BLOCH, H., SORKIN, E. & ERLNMEYER, H. A toxic lipid component of the tubercle bacillus (cord factor). I. Isolation from petroleum ether extracts of young bacterial cultures. *Am. Rev. Tuberc.* (1953). doi:10.1164/art.1953.67.5.629
218. Hunter, R. L., Olsen, M. R., Jagannath, C. & Actor, J. K. Multiple roles of cord factor in the pathogenesis of primary, secondary, and cavitary tuberculosis, including a revised description of the pathology of secondary disease. *Annals of Clinical and Laboratory Science* (2006).
219. Chabrol, E. *et al.* Glycosaminoglycans Are Interactants of Langerin: Comparison with gp120 Highlights an Unexpected Calcium-Independent Binding Mode. *PLoS One* (2012). doi:10.1371/journal.pone.0050722
220. Gantner, B. N., Simmons, R. M. & Underhill, D. M. Dectin-1 mediates macrophage recognition of *Candida albicans* yeast but not filaments. *EMBO J.* (2005). doi:10.1038/sj.emboj.7600594
221. Marotte, K. *et al.* X-ray structures and thermodynamics of the interaction of PA-IIL from *Pseudomonas aeruginosa* with disaccharide derivatives. *ChemMedChem* (2007). doi:10.1002/cmdc.200700100
222. Imberty, A., Chabre, Y. M. & Roy, R. Glycomimetics and glycodendrimers as high affinity microbial anti-adhesins. in *Chemistry - A European Journal* (2008). doi:10.1002/chem.200800700
223. Ernst, B. & Magnani, J. L. From carbohydrate leads to glycomimetic drugs. *Nature Reviews Drug Discovery* (2009). doi:10.1038/nrd2852
224. Gurney, K. B. *et al.* Binding and Transfer of Human Immunodeficiency Virus by DC-SIGN+ Cells in Human Rectal Mucosa. *J. Virol.* (2005). doi:10.1128/jvi.79.9.5762-5773.2005
225. Belkaid, Y. & Hand, T. W. Role of the microbiota in immunity and inflammation. *Cell* (2014). doi:10.1016/j.cell.2014.03.011
226. Porkolab, V. *et al.* Rational-Differential Design of Highly Specific Glycomimetic Ligands: Targeting DC-SIGN and Excluding Langerin Recognition. *ACS Chem. Biol.* (2018). doi:10.1021/acscchembio.7b00958
227. Valverde, P., Ardá, A., Reichardt, N. C., Jiménez-Barbero, J. & Gimeno, A. Glycans in drug discovery. *MedChemComm* (2019). doi:10.1039/c9md00292h
228. Porkolab, V. *et al.* Development of C-type lectin-oriented surfaces for high avidity glycoconjugates: Towards mimicking multivalent interactions on the cell surface. *Org. Biomol. Chem.* (2020). doi:10.1039/d0ob00781a
229. Telen, M. J. *et al.* Randomized phase 2 study of GMI-1070 in SCD: Reduction in time to resolution of vaso-occlusive events and decreased opioid use. *Blood* (2015). doi:10.1182/blood-2014-06-583351
230. Tvaroška, I., Selvaraj, C. & Koča, J. Selectins-The two Dr. Jekyll and Mr. Hyde faces of adhesion molecules-A review. *Molecules* (2020). doi:10.3390/molecules25122835
231. Watz, H. *et al.* Inhaled pan-selectin antagonist Bimosiamose attenuates airway inflammation in COPD. *Pulm. Pharmacol. Ther.* (2013). doi:10.1016/j.pupt.2012.12.003

232. Gray, C. J. *et al.* Advancing Solutions to the Carbohydrate Sequencing Challenge. *Journal of the American Chemical Society* (2019). doi:10.1021/jacs.9b06406
233. Wang, X., Zhang, C., Shi, F. & Hu, X. Purification and Characterization of Lipopolysaccharides. *Subcell. Biochem.* (2010). doi:10.1007/978-90-481-9078-2\_2
234. Galanos, C., Lüderitz, O. & Westphal, O. A New Method for the Extraction of R Lipopolysaccharides. *Eur. J. Biochem.* (1969). doi:10.1111/j.1432-1033.1969.tb00601.x
235. De Castro, C., Parrilli, M., Holst, O. & Molinaro, A. Microbe-associated molecular patterns in innate immunity. Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. in *Methods in Enzymology* (2010). doi:10.1016/S0076-6879(10)80005-9
236. Westphal, O. & Jann, K. Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedures. . *Methods Carbohydr. Chem.* (1965).
237. Rezanian, S. *et al.* Extraction, purification and characterization of lipopolysaccharide from *Escherichia coli* and *Salmonella typhi*. *Avicenna J. Med. Biotechnol.* (2011).
238. Yi, E. C. & Hackett, M. Rapid isolation method for lipopolysaccharide and lipid A from Gram-negative bacteria. *Analyst* (2000). doi:10.1039/b000368i
239. Tsai, C. M. & Frasch, C. E. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Anal. Biochem.* **119**, 115–119 (1982).
240. Komuro, T. & Galanos, C. Analysis of salmonella lipopolysaccharides by sodium deoxycholate-polyacrylamide gel electrophoresis. *J. Chromatogr. A* (1988). doi:10.1016/S0021-9673(01)83593-7
241. Steinberg, T. H. *et al.* Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics* (2001). doi:10.1002/1615-9861(200107)1:7<841::AID-PROT841>3.0.CO;2-E
242. Holst, O. Deacylation of lipopolysaccharides and isolation of oligosaccharide phosphates. *Methods Mol. Biol.* (2000). doi:10.1385/1-59259-052-7:345
243. Leontein, K., Lindberg, B. & Lonngren, J. Assignment of absolute configuration of sugars by g.l.c. of their acetylated glycosides formed from chiral alcohols. *Carbohydr. Res.* (1978). doi:10.1016/S0008-6215(00)80882-4
244. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* (1984). doi:10.1016/0008-6215(84)85242-8
245. BLIGH, E. G. & DYER, W. J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* (1959). doi:10.1139/o59-099
246. Burton, A. J. & Carter, H. E. Purification and Characterization of the Lipid A Component of the Lipopolysaccharides from *Escherichia coli*. *Biochemistry* (1964). doi:10.1021/bi00891a018
247. Dell, A. & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* (2001). doi:10.1126/science.1058890
248. Daikoku, S., Widmalm, G. & Kanie, O. Analysis of a series of isomeric oligosaccharides by energy-resolved mass spectrometry: A challenge on homobranched trisaccharides. *Rapid Commun. Mass Spectrom.* (2009). doi:10.1002/rcm.4303
249. Harvey, D. J. Fragmentation of negative ions from carbohydrates: Part 1. Use of nitrate and other anionic adducts for the production of negative ion electrospray spectra from N-linked

- carbohydrates. *J. Am. Soc. Mass Spectrom.* (2005). doi:10.1016/j.jasms.2005.01.004
250. Costello, C. E. & Vath, J. E. Tandem mass spectrometry of glycolipids. *Methods Enzymol.* (1990). doi:10.1016/0076-6879(90)93448-T
251. Boué, S. M. & Cole, R. B. Confirmation of the structure of lipid A from *Enterobacter agglomerans* by electrospray ionization tandem mass spectrometry. *J. Mass Spectrom.* (2000). doi:10.1002/(SICI)1096-9888(200003)35:3<361::AID-JMS943>3.0.CO;2-D
252. Kussak, A. & Weintraub, A. Quadrupole ion-trap mass spectrometry to locate fatty acids on lipid A from Gram-negative bacteria. *Anal. Biochem.* (2002). doi:10.1016/S0003-2697(02)00004-0
253. Kilár, A., Dörnyei, Á. & Kocsis, B. Structural characterization of bacterial lipopolysaccharides with mass spectrometry and on- and off-line separation techniques. *Mass Spectrometry Reviews* (2013). doi:10.1002/mas.21352
254. Rockwood, A. L., Kushnir, M. M. & Clarke, N. J. Mass spectrometry. in *Principles and Applications of Clinical Mass Spectrometry: Small Molecules, Peptides, and Pathogens* (2018). doi:10.1016/B978-0-12-816063-3.00002-5
255. Linnerborg, M., Weintraub, A. & Widmalm, G. Structural studies of the O-antigen polysaccharide from the enteroinvasive *Escherichia coli* O164 cross-reacting with *Shigella dysenteriae* type 3. *Eur. J. Biochem.* (1999). doi:10.1046/j.1432-1327.1999.00878.x
256. Sturiale, L. *et al.* New conditions for matrix-assisted laser desorption/ionization mass spectrometry of native bacterial R-type lipopolysaccharides. *Rapid Commun. Mass Spectrom.* (2005). doi:10.1002/rcm.1994
257. Silipo, A. *et al.* The complete structure of the lipooligosaccharide from the halophilic bacterium *Pseudoalteromonas issachenkonii* KMM 3549T. *Carbohydr. Res.* (2004). doi:10.1016/j.carres.2004.05.008
258. Ieranò, T. *et al.* Structural elucidation of a novel *B. cenocepacia* ET-12 lipooligosaccharide isolated from a cystic fibrosis patient after lung transplantation. *European J. Org. Chem.* (2010). doi:10.1002/ejoc.200901200
259. Silipo, A. *et al.* The complete structure and pro-inflammatory activity of the lipooligosaccharide of the highly epidemic and virulent gram-negative bacterium *Burkholderia cenocepacia* ET-12 (strain J2315). *Chem. - A Eur. J.* (2007). doi:10.1002/chem.200601406
260. Levitt, M. H. *Spin dynamics: basics of nuclear magnetic resonance. Magnetic Resonance in Chemistry* (2002).
261. Claridge, T. D. W. High-resolution NMR techniques in organic chemistry. *Tetrahedron Org. Chem. Ser.* ; (1999).
262. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angewandte Chemie - International Edition* (2003). doi:10.1002/anie.200390233
263. Davies, D. B. The nuclear overhauser effect in structural and conformational analysis. *Spectrochim. Acta Part A Mol. Spectrosc.* (1991). doi:10.1016/0584-8539(91)80264-j
264. Keeler, J. Chapter 7. Two-dimensional NMR. *Underst. NMR Spectrosc.* (2010).
265. Melinda J.Duer. *Solid-State NMR Spectroscopy Principles and Applications. Solid-State NMR Spectroscopy Principles and Applications* (2007). doi:10.1002/9780470999394

266. Hennel, J. W. & Klinowski, J. Magic Angle Spinning: A Historical Perspective. *ChemInform* (2005). doi:10.1002/chin.200535326
267. Strandberg, E., Bentz, D., Wadhvani, P. & Ulrich, A. S. Chiral supramolecular architecture of stable transmembrane pores formed by an  $\alpha$ -helical antibiotic peptide in the presence of lysolipids. *Sci. Rep.* (2020). doi:10.1038/s41598-020-61526-w
268. Laguri, C. *et al.* Solid State NMR Studies of Intact Lipopolysaccharide Endotoxin. *ACS Chem. Biol.* (2018). doi:10.1021/acscchembio.8b00271
269. Schanda, P. *et al.* Atomic model of a cell-wall cross-linking enzyme in complex with an intact bacterial peptidoglycan. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja5105987
270. Ehren, H. L. *et al.* Characterization of the cell wall of a mushroom forming fungus at atomic resolution using solid-state NMR spectroscopy. *Cell Surf.* (2020). doi:10.1016/j.tcs.2020.100046
271. Siemer, A. B. Advances in studying protein disorder with solid-state NMR. *Solid State Nuclear Magnetic Resonance* (2020). doi:10.1016/j.ssnmr.2020.101643
272. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* (2014). doi:10.1016/j.jmr.2013.11.011
273. Duus, J., Gotfredsen, C. H. & Bock, K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem. Rev.* (2000). doi:10.1021/cr990302n
274. *NMR Spectroscopy of Glycoconjugates. NMR Spectroscopy of Glycoconjugates* (2002). doi:10.1002/352760071x
275. Fontana, C., Kovacs, H. & Widmalm, G. NMR structure analysis of uniformly  $^{13}\text{C}$ -labeled carbohydrates. *J. Biomol. NMR* (2014). doi:10.1007/s10858-014-9830-6
276. Nomura, K. *et al.* Interaction of lipopolysaccharide and phospholipid in mixed membranes: Solid-state  $^{31}\text{P}$ -NMR spectroscopic and microscopic investigations. *Biophys. J.* (2008). doi:10.1529/biophysj.108.131706
277. Jachymek, W. *et al.* Structures of the O-specific polysaccharides from *Yersinia regensburgensis* (Koserella trabulsii) strains PCM 2476, 2477, 2478, and 2494: High-resolution magic-angle spinning NMR investigation of the O-specific polysaccharides in native lipopolysaccharides a. *Biochemistry* (1999). doi:10.1021/bi990673y
278. Lanne, A. B. M. *et al.* Molecular recognition of lipopolysaccharide by the lantibiotic nisin. *Biochim. Biophys. Acta - Biomembr.* (2019). doi:10.1016/j.bbamem.2018.10.006
279. Du, X. *et al.* Insights into protein–ligand interactions: Mechanisms, models, and methods. *International Journal of Molecular Sciences* (2016). doi:10.3390/ijms17020144
280. K., A. Thermodynamics of Ligand-Protein Interactions: Implications for Molecular Design. in *Thermodynamics - Interaction Studies - Solids, Liquids and Gases* (2011). doi:10.5772/19447
281. Geissner, A. *et al.* Microbe-focused glycan array screening platform. *Proc. Natl. Acad. Sci. U. S. A.* (2019). doi:10.1073/pnas.1800853116
282. Lundquist, J. J. & Toone, E. J. The cluster glycoside effect. *Chem. Rev.* (2002). doi:10.1021/cr000418f
283. Laigre, E., Goyard, D., Tiertant, C., Dejeu, J. & Renaudet, O. The study of multivalent carbohydrate-protein interactions by bio-layer interferometry. *Org. Biomol. Chem.* (2018). doi:10.1039/c8ob01664j

284. Dam, T. K. & Brewer, C. F. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* (2002). doi:10.1021/cr000401x
285. Dam, T. K., Roy, R., Das, S. K., Oscarson, S. & Brewer, C. F. Binding of multivalent carbohydrates to concanavalin A and Dioclea grandiflora lectin. Thermodynamic analysis of the 'multivalency effect'. *J. Biol. Chem.* (2000). doi:10.1074/jbc.275.19.14223
286. Sabin, C. *et al.* Binding of different monosaccharides by lectin PA-IIL from *Pseudomonas aeruginosa*: Thermodynamics data correlated with X-ray structures. *FEBS Lett.* (2006). doi:10.1016/j.febslet.2006.01.030
287. Kussrow, A. *et al.* Measurement of monovalent and polyvalent carbohydrate-lectin binding by back-scattering interferometry. *Anal. Chem.* (2009). doi:10.1021/ac900569c
288. Rief, M., Clausen-Schaumann, H. & Gaub, H. E. Sequence-dependent mechanics of single DNA molecules. *Nat. Struct. Biol.* (1999). doi:10.1038/7582
289. Hinterdorfer, P. & Dufrêne, Y. F. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature Methods* (2006). doi:10.1038/nmeth871
290. Touhami, A., Hoffmann, B., Vasella, A., Denis, F. A. & Dufrêne, Y. F. Aggregation of yeast cells: Direct measurement of discrete lectin-carbohydrate interactions. *Microbiology* (2003). doi:10.1099/mic.0.26431-0
291. Brzozowska, E. *et al.* Interactions of bacteriophage T4 adhesin with selected lipopolysaccharides studied using atomic force microscopy. *Sci. Rep.* (2018). doi:10.1038/s41598-018-29383-w
292. Giepmans, B. N. G. Bridging fluorescence microscopy and electron microscopy. *Histochemistry and Cell Biology* (2008). doi:10.1007/s00418-008-0460-5
293. Lakshminarayan, R. *et al.* Galectin-3 drives glycosphingolipid-dependent biogenesis of clathrin-independent carriers. *Nat. Cell Biol.* (2014). doi:10.1038/ncb2970
294. Aigal, S., Claudinon, J. & Römer, W. Plasma membrane reorganization: A glycolipid gateway for microbes. *Biochim. Biophys. Acta - Mol. Cell Res.* (2015). doi:10.1016/j.bbamcr.2014.11.014
295. Ahmad, N. *et al.* Galectin-3 Precipitates as a Pentamer with Synthetic Multivalent Carbohydrates and Forms Heterogeneous Cross-linked Complexes. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M312834200
296. Lepur, A., Salomonsson, E., Nilsson, U. J. & Leffler, H. Ligand induced galectin-3 protein self-association. *J. Biol. Chem.* (2012). doi:10.1074/jbc.C112.358002
297. Elola, M. T., Blidner, A. G., Ferragut, F., Bracalente, C. & Rabinovich, G. A. Assembly, organization and regulation of cell-surface receptors by lectin-glycan complexes. *Biochemical Journal* (2015). doi:10.1042/BJ20150461
298. Sych, T., Mély, Y. & Römer, W. Lipid self-assembly and lectin-induced reorganization of the plasma membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences* (2018). doi:10.1098/rstb.2017.0117
299. Richter, W. *et al.* Morphology, size distribution, and aggregate structure of lipopolysaccharide and lipid A dispersions from enterobacterial origin. *Innate Immun.* (2011). doi:10.1177/1753425910372434
300. Mueller, M. *et al.* Aggregates are the biologically active units of endotoxin. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M401231200

301. Malin-Berdel, J., Valet, G., Thiel, E., Forrester, J. A. & Gürtler, L. Flow cytometric analysis of the binding of eleven lectins to human T- and B-cells and to human T- and B-cell lines. *Cytometry* (1984). doi:10.1002/cyto.990050215
302. Mayer, M. & Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chemie - Int. Ed.* (1999). doi:10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q
303. Haselhorst, T., Lamerz, A. C. & Itzstein, M. Von. Saturation transfer difference NMR spectroscopy as a technique to investigate protein-carbohydrate interactions in solution. *Methods Mol. Biol.* (2009). doi:10.1007/978-1-59745-022-5\_26
304. Mayer, M. & Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.* (2001). doi:10.1021/ja0100120
305. Grzesiek, S., Anglister, J. & Bax, A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in <sup>13</sup>C/<sup>15</sup>N-Enriched Proteins by Isotropic Mixing of <sup>13</sup>C Magnetization. *Journal of Magnetic Resonance, Series B* (1993). doi:10.1006/jmrb.1993.1019
306. Bax, A. & Grzesiek, S. Methodological Advances in Protein NMR. in *NMR of Proteins* (1993). doi:10.1007/978-1-349-12749-8\_2
307. Grzesiek, S. & Bax, A. Amino acid type determination in the sequential assignment procedure of uniformly <sup>13</sup>C/<sup>15</sup>N-enriched proteins. *J. Biomol. NMR* (1993). doi:10.1007/BF00178261
308. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* (2013). doi:10.1016/j.pnmrs.2013.02.001
309. Becker, W., Bhattiprolu, K. C., Gubensäk, N. & Zangger, K. Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem* (2018). doi:10.1002/cphc.201701253
310. Lescop, E., Kern, T. & Brutscher, B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: Example of longitudinal-relaxation-enhanced BEST-type <sup>1</sup>H-<sup>15</sup>N correlation experiments. *J. Magn. Reson.* (2010). doi:10.1016/j.jmr.2009.12.001
311. Foster, M. P., McElroy, C. A. & Amero, C. D. Solution NMR of large molecules and assemblies. *Biochemistry* (2007). doi:10.1021/bi0621314
312. Unno, H. *et al.* Identification, Characterization, and X-ray Crystallographic Analysis of a Novel Type of Lectin AJLec from the Sea Anemone *Anthopleura japonica*. *Sci. Rep.* (2018). doi:10.1038/s41598-018-29498-0
313. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* (2018). doi:10.1016/j.neuron.2018.08.011
314. Schweinle, J. E. *et al.* Truncated forms of mannose-binding protein multimerize and bind to mannose-rich *Salmonella montevideo* but fail to activate complement in vitro. *J. Biol. Chem.* (1993). doi:10.1016/s0021-9258(18)54159-4
315. Sheriff, S., Chang, C. Y. & Ezekowitz, R. A. B. Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple  $\alpha$ -helical coiled-coil. *Nat. Struct. Biol.* (1994). doi:10.1038/nsb1194-789
316. Degn, S. E., Jensenius, J. C. & Bjerre, M. The lectin pathway and its implications in coagulation, infections and auto-immunity. *Curr. Opin. Organ Transplant.* (2011).



doi:10.1097/MOT.0b013e32834253df

317. Ihara, I., Harada, Y., Ihara, S. & Kawakami, M. A new complement-dependent bactericidal factor found in nonimmune mouse sera: Specific binding to polysaccharide of Ra chemotype Salmonella. *J. Immunol.* (1982).
318. Dumestre-Pérard, C., Doerr, E., Colomb, M. G. & Loos, M. Involvement of complement pathways in patients with bacterial septicemia. *Mol. Immunol.* (2007). doi:10.1016/j.molimm.2006.08.008
319. Kasperkiewicz, K. *et al.* Interaction of human mannose-binding lectin (MBL) with Yersinia enterocolitica lipopolysaccharide. *Int. J. Med. Microbiol.* (2015). doi:10.1016/j.ijmm.2015.07.001
320. Iobst, S. T. & Drickamer, K. Binding of sugar ligands to Ca<sup>2+</sup>-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis. *J. Biol. Chem.* (1994). doi:10.1016/s0021-9258(17)40709-5
321. Wallis, R. Interactions between mannose-binding lectin and MASPs during complement activation by the lectin pathway. *Immunobiology* (2007). doi:10.1016/j.imbio.2006.11.004
322. Bergman, M. P. *et al.* Helicobacter pylori modulates the T helper cell 1/T helper cell 2 balance through phase-variable interaction between lipopolysaccharide and DC-SIGN. *J. Exp. Med.* (2004). doi:10.1084/jem.20041061
323. Steeghs, L. *et al.* Neisseria meningitidis expressing IgtB lipopolysaccharide targets DC-SIGN and modulates dendritic cell function. *Cell. Microbiol.* (2006). doi:10.1111/j.1462-5822.2005.00623.x
324. Zhang, P. *et al.* Role of N -Acetylglucosamine within Core Lipopolysaccharide of Several Species of Gram-Negative Bacteria in Targeting the DC-SIGN (CD209) . *J. Immunol.* (2006). doi:10.4049/jimmunol.177.6.4002
325. Klena, J., Zhang, P., Schwartz, O., Hull, S. & Chen, T. The core lipopolysaccharide of Escherichia coli is a ligand for the dendritic-cell-specific intercellular adhesion molecule nonintegrin CD209 receptor. *J. Bacteriol.* (2005). doi:10.1128/JB.187.5.1710-1715.2005
326. Feng, D. *et al.* DC-SIGN reacts with TLR-4 and regulates inflammatory cytokine expression via NF- $\kappa$ B activation in renal tubular epithelial cells during acute renal injury. *Clin. Exp. Immunol.* (2018). doi:10.1111/cei.13048
327. Kaniuk, N. A., Vinogradov, E., Li, J., Monteiro, M. A. & Whitfield, C. Chromosomal and plasmid-encoded enzymes are required for assembly of the R3-type core oligosaccharide in the lipopolysaccharide of Escherichia coli O157:H7. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M401879200
328. Amor, K. *et al.* Distribution of core oligosaccharide types in lipopolysaccharides from Escherichia coli. *Infect. Immun.* (2000). doi:10.1128/IAI.68.3.1116-1124.2000
329. Vinogradov, E., Conlan, J. W. & Perry, M. B. Serological cross-reaction between the lipopolysaccharide O- polysaccharide antigens of Escherichia coli O157:H7 and strains of Citrobacter freundii and Citrobacter sedlakii. *FEMS Microbiol. Lett.* (2000). doi:10.1016/S0378-1097(00)00320-7
330. Bundle, D. R., Gerken, M. & Perry, M. B. Two-dimensional nuclear magnetic resonance at 500 MHz: The structural elucidation of a Salmonella serogroup N polysaccharide antigen. *Can. J. Chem.* (1986). doi:10.1139/v86-044
331. Conlan, J. W., KuoLee, R., Webb, A. & Perry, M. B. Salmonella landau as a live vaccine against Escherichia coli O157:H7 investigated in a mouse model of intestinal colonization. *Can. J.*

- Microbiol.* (1999). doi:10.1139/w99-059
332. Nishiuchi, Y., Doe, M., Hotta, H. & Kobayashi, K. Structure and serologic properties of O-specific polysaccharide from *Citrobacter freundii* possessing cross-reactivity with *Escherichia coli* O157:H7. *FEMS Immunol. Med. Microbiol.* (2000). doi:10.1016/S0928-8244(00)00149-8
333. Wang, L. & Reeves, P. R. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. *Infect. Immun.* (1998). doi:10.1128/iai.66.8.3545-3551.1998
334. Klein, G. *et al.* Molecular and structural basis of inner core lipopolysaccharide alterations in *Escherichia coli*: Incorporation of glucuronic acid and phosphoethanolamine in the heptose region. *J. Biol. Chem.* (2013). doi:10.1074/jbc.M112.445981
335. Mahenthalingam, E., Urban, T. A. & Goldberg, J. B. The multifarious, multireplicon *Burkholderia cepacia* complex. *Nature Reviews Microbiology* (2005). doi:10.1038/nrmicro1085
336. Müller-Loennies, S., Holst, O. & Brade, H. Chemical Structure of the Core Region of *Escherichia coli* J-5 Lipopolysaccharide. *Eur. J. Biochem.* (1994). doi:10.1111/j.1432-1033.1994.t01-1-00751.x
337. Rybka, J. & Gamian, A. Determination of endotoxin by the measurement of the acetylated methyl glycoside derivative of Kdo with gas-liquid chromatography-mass spectrometry. *J. Microbiol. Methods* (2006). doi:10.1016/j.mimet.2005.04.029
338. Szponar, B., Norin, E., Midtvedt, T. & Larsson, L. Limitations in the use of 3-hydroxy fatty acid analysis to determine endotoxin in mammalian samples. *J. Microbiol. Methods* (2002). doi:10.1016/S0167-7012(02)00038-6
339. De Santana-Filho, A. P. *et al.* GC-MS detection and quantification of lipopolysaccharides in polysaccharides through 3-O-acetyl fatty acid methyl esters. *Carbohydr. Polym.* **87**, 2730–2734 (2012).
340. Steimle, A., Autenrieth, I. B. & Frick, J. S. Structure and function: Lipid A modifications in commensals and pathogens. *International Journal of Medical Microbiology* (2016). doi:10.1016/j.ijmm.2016.03.001
341. Santos, N. C., Silva, A. C., Castanho, M. A. R. B., Martins-Silva, J. & Saldanha, C. Evaluation of Lipopolysaccharide aggregation by light scattering spectroscopy. *ChemBioChem* (2003). doi:10.1002/cbic.200390020
342. Vinogradov, E. V. *et al.* The structures of the carbohydrate backbones of the lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type). *Eur. J. Biochem.* (1999). doi:10.1046/j.1432-1327.1999.00280.x
343. Müller-Loennies, S., Lindner, B. & Brade, H. Structural analysis of deacylated lipopolysaccharide of *Escherichia coli* strains 2513 (R4 core-type) and F653 (R3 core-type). *Eur. J. Biochem.* (2002). doi:10.1046/j.1432-1033.2002.03322.x
344. Maalej, M. *et al.* Human Macrophage Galactose-Type Lectin (MGL) Recognizes the Outer Core of *Escherichia coli* Lipooligosaccharide. *ChemBioChem* **20**, 1778–1782 (2019).
345. Stevens, T. J. *et al.* A software framework for analysing solid-state MAS NMR data. *J. Biomol. NMR* (2011). doi:10.1007/s10858-011-9569-2
346. Varki, A. *et al.* Symbol nomenclature for graphical representations of glycans. *Glycobiology* (2015). doi:10.1093/glycob/cwv091
347. Mody, R., Joshi, S. H. antara. & Chaney, W. Use of lectins as diagnostic and therapeutic tools for

- cancer. *Journal of Pharmacological and Toxicological Methods* (1995). doi:10.1016/1056-8719(94)00052-6
348. Lopes, J. & Inniss, W. E. Electron microscopic study of lipopolysaccharide from an avian strain of *Escherichia coli* O18. *J. Bacteriol.* (1970). doi:10.1128/jb.103.1.238-243.1970
349. Malojčić, G. *et al.* LptE binds to and alters the physical state of LPS to catalyze its assembly at the cell surface. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1402746111
350. Diniz, A. *et al.* The Plasticity of the Carbohydrate Recognition Domain Dictates the Exquisite Mechanism of Binding of Human Macrophage Galactose-Type Lectin. *Chem. - A Eur. J.* **25**, 13945–13955 (2019).
351. van Vliet, S. J., Saeland, E. & van Kooyk, Y. Sweet preferences of MGL: carbohydrate specificity and function. *Trends in Immunology* (2008). doi:10.1016/j.it.2007.10.010
352. Skinner, S. P. *et al.* CcpNmr AnalysisAssign: a flexible platform for integrated NMR analysis. *J. Biomol. NMR* (2016). doi:10.1007/s10858-016-0060-y
353. Tanaka, J. *et al.* Specific and Differential Binding of N-Acetylgalactosamine Glycopolymers to the Human Macrophage Galactose Lectin and Asialoglycoprotein Receptor. *Biomacromolecules* (2017). doi:10.1021/acs.biomac.7b00228
354. van Vliet, S. J. *et al.* Carbohydrate profiling reveals a distinctive role for the C-type lectin MGL in the recognition of helminth parasites and tumor antigens by dendritic cells. *Int. Immunol.* (2005). doi:10.1093/intimm/dxh246
355. Bewley, C. A. & Shahzad-UI-Hussan, S. Characterizing carbohydrate-protein interactions by nuclear magnetic resonance spectroscopy. *Biopolymers* (2013). doi:10.1002/bip.22329
356. Cairo, C. W., Gestwicki, J. E., Kanai, M. & Kiessling, L. L. Control of multivalent interactions by binding epitope density. *J. Am. Chem. Soc.* (2002). doi:10.1021/ja016727k
357. Kanai, M., Mortell, K. H. & Kiessling, L. L. Varying the size of multivalent ligands: The dependence of concanavalin A binding on neoglycopolymer length. *J. Am. Chem. Soc.* (1997). doi:10.1021/ja972089n
358. Aretz, J. *et al.* Identification of Multiple Druggable Secondary Sites by Fragment Screening against DC-SIGN. *Angew. Chemie - Int. Ed.* (2017). doi:10.1002/anie.201701943
359. Hanske, J. *et al.* Bacterial polysaccharide specificity of the pattern recognition receptor langerin is highly species-dependent. *J. Biol. Chem.* **292**, (2017).
360. Li, D., Kagan, G., Hopson, R. & Williard, P. G. Formula weight prediction by internal reference diffusion-ordered NMR spectroscopy (DOSY). *J. Am. Chem. Soc.* (2009). doi:10.1021/ja810154u
361. Tobias, R. Biomolecular Binding Kinetics Assays on the Octet Platform. *ForteBio Interact. Pall Life Sciences* (2013).
362. Ribeiro, L. N. D. M., Couto, V. M., Fraceto, L. F. & De Paula, E. Use of nanoparticle concentration as a tool to understand the structural properties of colloids. *Sci. Rep.* (2018). doi:10.1038/s41598-017-18573-7
363. AvantiPolarLipids. Mini extruder Technique. *CEUR Workshop Proc.* (2015).
364. Prehm, P., Jann, B., Jann, K., Schmidt, G. & Stirm, S. On a bacteriophage T3 and T4 receptor region within the cell wall lipopolysaccharide of *Escherichia coli* B. *J. Mol. Biol.* (1976). doi:10.1016/0022-2836(76)90377-6

365. Washizaki, A., Yonesaki, T. & Otsuka, Y. Characterization of the interactions between Escherichia coli receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *Microbiologyopen* (2016). doi:10.1002/mbo3.384
366. Gauthier, J. D., Jenkins, J. A. & La Peyre, J. F. Flow cytometric analysis of lectin binding to in vitro-cultured Perkinsus marinus surface carbohydrates. *J. Parasitol.* (2004). doi:10.1645/GE-3269
367. Hendrickson, O. D., Nikitushkin, V. D., Zherdev, A. V. & Dzantiev, B. B. Lectin-based detection of Escherichia coli and Staphylococcus aureus by flow cytometry. *Arch. Microbiol.* (2019). doi:10.1007/s00203-018-1613-0
368. Pink, D. A. *et al.* Divalent Calcium Ions Inhibit the Penetration of Protamine through the Polysaccharide Brush of the Outer Membrane of Gram-Negative Bacteria. *Langmuir* (2003). doi:10.1021/la030193e
369. Lam, N. H., Ma, Z. & Ha, B. Y. Electrostatic modification of the lipopolysaccharide layer: Competing effects of divalent cations and polycationic or polyanionic molecules. *Soft Matter* (2014). doi:10.1039/c4sm01262c
370. Dam, T. K. & Brewer, C. F. Effects of clustered epitopes in multivalent ligand-receptor interactions. *Biochemistry* (2008). doi:10.1021/bi801208b
371. Latz, E. *et al.* Ligand-induced conformational changes allosterically activate Toll-like receptor 9. *Nat. Immunol.* (2007). doi:10.1038/ni1479
372. Boyé, K., Billottet, C., Pujol, N., Alves, I. D. & Bikfalvi, A. Ligand activation induces different conformational changes in CXCR3 receptor isoforms as evidenced by plasmon waveguide resonance (PWR). *Sci. Rep.* (2017). doi:10.1038/s41598-017-11151-x
373. Man-Kupisinska, A. *et al.* Interaction of mannose-binding lectin with lipopolysaccharide outer core region and its biological consequences. *Front. Immunol.* (2018). doi:10.3389/fimmu.2018.01498
374. Thompson, L. J. *et al.* Gene expression profiling of Helicobacter pylori reveals a growth-phase-dependent switch in virulence gene expression. *Infect. Immun.* (2003). doi:10.1128/IAI.71.5.2643-2655.2003
375. Agrawal, A., Rangarajan, N. & Weisshaar, J. C. Resistance of early stationary phase E. coli to membrane permeabilization by the antimicrobial peptide Cecropin A. *Biochim. Biophys. Acta - Biomembr.* (2019). doi:10.1016/j.bbamem.2019.05.012
376. Klein, G., Lindner, B., Brabetz, W., Brade, H. & Raina, S. Escherichia coli K-12 suppressor-free mutants lacking early glycosyltransferases and late acyltransferases. Minimal lipopolysaccharide structure and induction of envelope stress response. *J. Biol. Chem.* (2009). doi:10.1074/jbc.M900490200
377. Lahtinen, P., Brzezinska, A. & Skurnik, M. Temperature and growth phase regulate the transcription of the O-antigen gene cluster of Yersinia enterocolitica O:3. in *Advances in Experimental Medicine and Biology* (2003). doi:10.1007/0-306-48416-1\_55
378. Jégouzo, S. A. *et al.* Organization of the extracellular portion of the macrophage galactose receptor: A trimeric cluster of simple binding sites for N-acetylgalactosamine. *Glycobiology* (2013). doi:10.1093/glycob/cwt022
379. Munoz, E. M., Correa, J., Fernandez-Megia, E. & Riguera, R. Probing the relevance of lectin clustering for the reliable evaluation of multivalent carbohydrate recognition. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja9074826

380. Müller-Loennies, S., Brade, L., MacKenzie, C. R., Di Padova, F. E. & Brade, H. Identification of a cross-reactive epitope widely present in lipopolysaccharide from enterobacteria and recognized by the cross-protective monoclonal antibody WN1 222-5. *J. Biol. Chem.* (2003). doi:10.1074/jbc.M302904200
381. Loutet, S. A., Flannagan, R. S., Kooi, C., Sokol, P. A. & Valvano, M. A. A complete lipopolysaccharide inner core oligosaccharide is required for resistance of Burkholderia cenocepacia to antimicrobial peptides and bacterial survival in vivo. *J. Bacteriol.* (2006). doi:10.1128/JB.188.6.2073-2080.2006
382. Baenziger, J. E. & DaCosta, C. J. B. Membrane Protein Structure and Conformational Change Probed using Fourier Transform Infrared Spectroscopy. in *Biophysical Analysis of Membrane Proteins: Investigating Structure and Function* (2008). doi:10.1002/9783527621224.ch11
383. Weljie, A. M. Protein conformational changes studied by diffusion NMR spectroscopy: Application to helix-loop-helix calcium binding proteins. *Protein Sci.* (2003). doi:10.1110/ps.0226203
384. Izumi, Y., Kuwamoto, S., Jinbo, Y. & Yoshino, H. Increase in the molecular weight and radius of gyration of apocalmodulin induced by binding of target peptide: Evidence for complex formation. *FEBS Lett.* (2001). doi:10.1016/S0014-5793(01)02375-4
385. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* (1974). doi:10.1128/jb.119.3.736-747.1974

## Annexes

Figure 1. Chromatograms recorded during G10 column chromatography for delipidated LOSs R1 (A) and R3 (B)

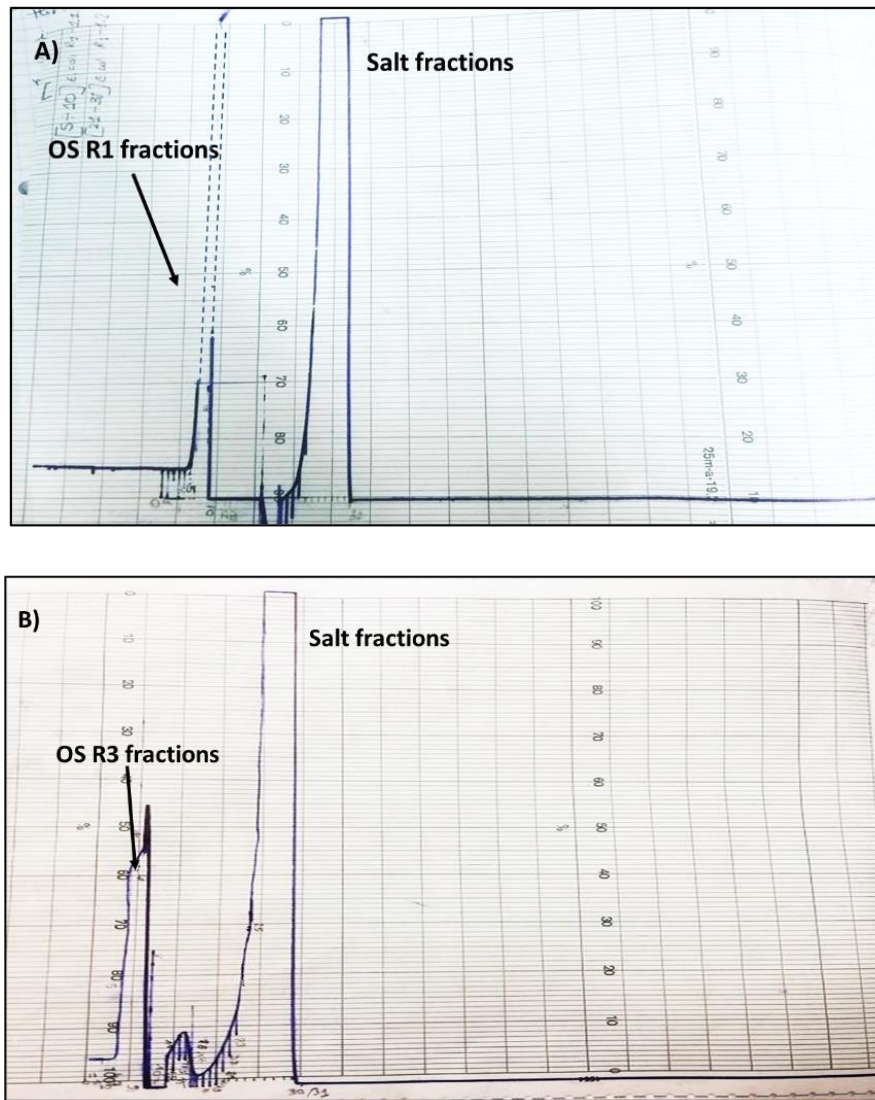
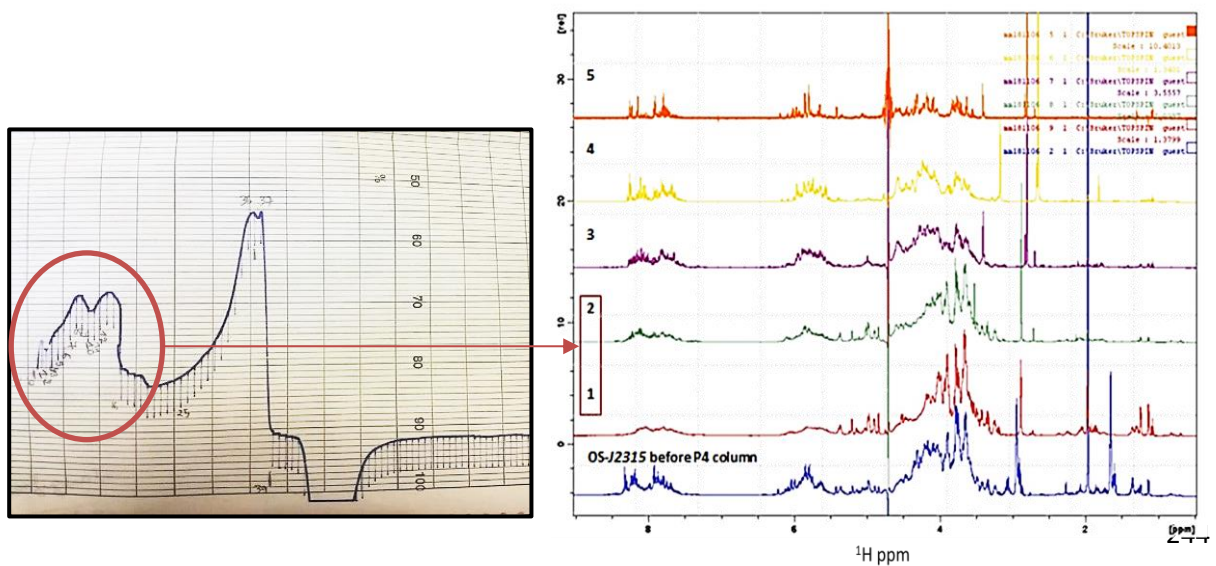
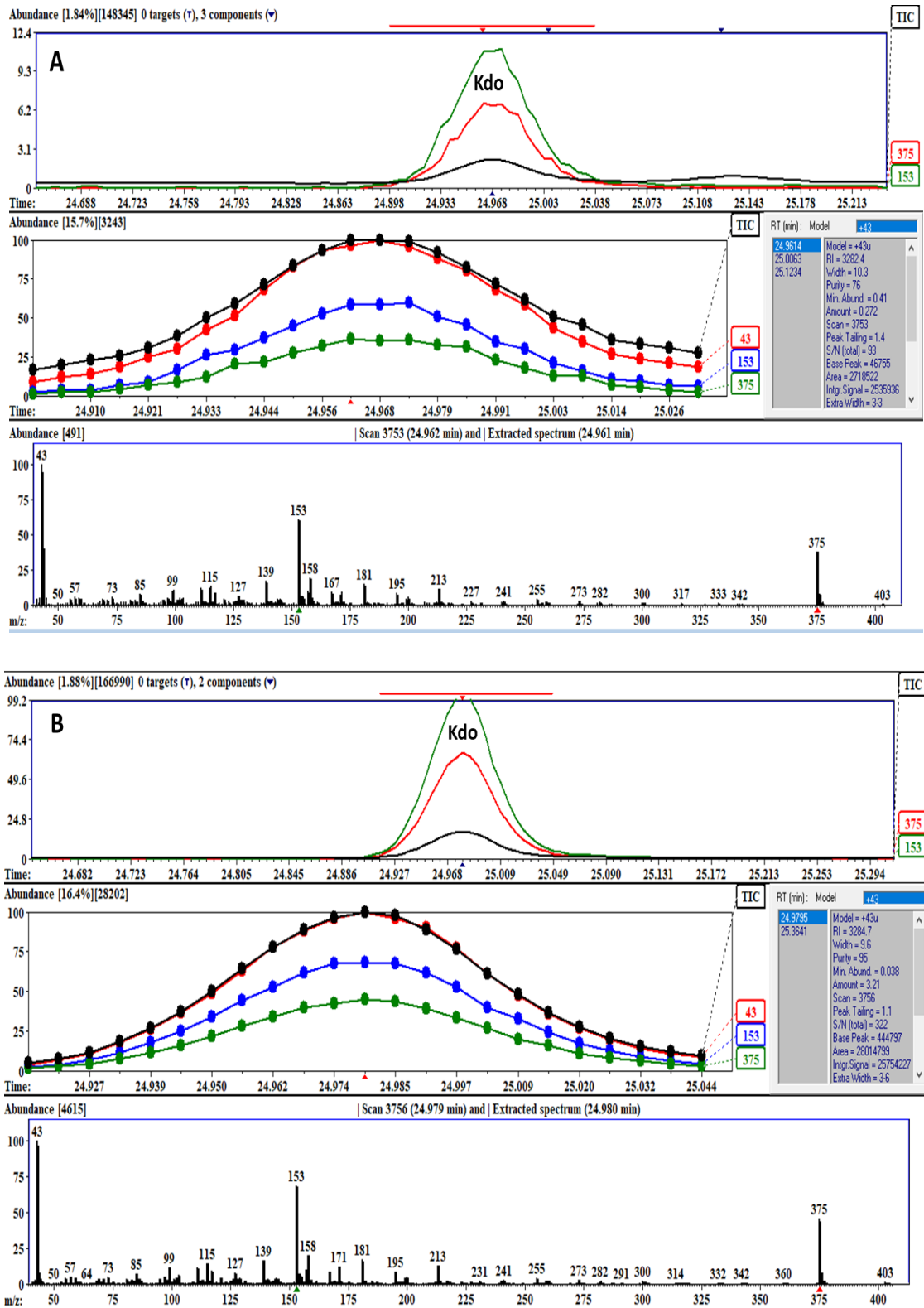


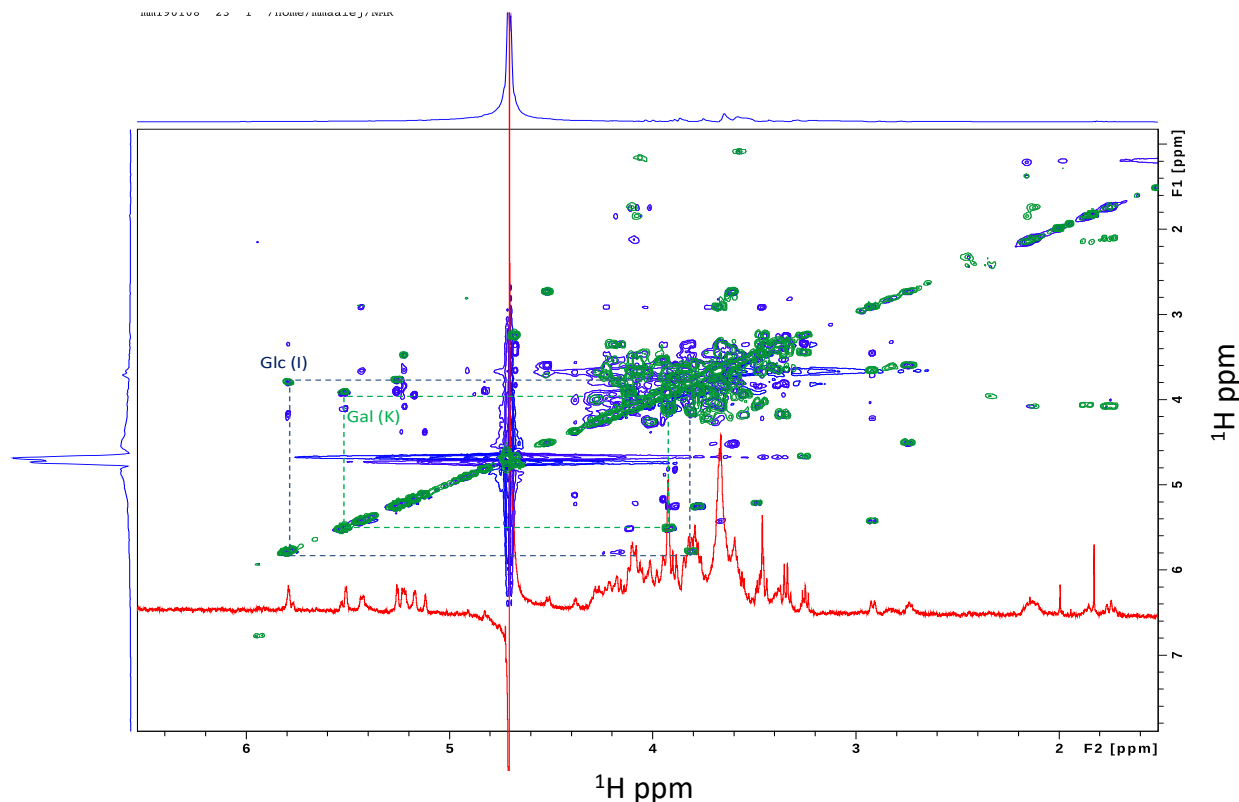
Figure 2. 600 MHz  $^1\text{H}$ -NMR spectra of OS J2315 fractions in  $\text{D}_2\text{O}$  at 298K, after P4 column chromatography.



**Figure 3. GC and MS partial chromatograms of (A) LOS R1 and (B) LOS R3 showing zoomed region of Kdo retention time peak and the occurrences of its characteristic fragmentation patterns corresponding to 43, 153 and 375 m/z reporter ions.**



**Figure 4. Superposition of 1D- $^1\text{H}$ -NMR and 2D- $^1\text{H}$ - $^1\text{H}$ -NMR COSY and TOCSY spectra of OS R1 in  $\text{D}_2\text{O}$ . Rectangular boxes indicate examples of spin systems that were attributed such as glucose (residue I) and galactose (residue K).**



**Table 1.  $^1\text{H}$ - $^{13}\text{C}$ -NMR data of deacetylated LOS *E. coli* R1 in buffer A, referenced to Acetone.**

Residue	Chemical shift of protons / carbons (ppm)							
	1	2	3 (ax/eq)	4	5	6(a/b)	7(a/b)	8(a/b)
<b>A GlcN</b>	5.51/93.9	3.25/57.71	3.5/70.23	3.35/71.9	3.92	(3.66/4.28)/72	--	--
<b>B GlcN</b>	4.68/105.3	2.82/58	3.70/73.1	3.62/75.62	3.61/74	(3.58/3.68)/65.7	--	--
<b>C Kdo</b>	--	--	(1.85/2)/nd	nd	4.18/73.5	nd	nd	nd
<b>D Kdo</b>	--	--	(1.68/2)/nd	4.09/69	nd	nd	nd	nd
<b>E Hep</b>	5.16/101.2	4.02/71.82	4.00/77.95	4.37/72.4	4.21/73.2	(4.09/nd)/73.9	(3.8/3.93)/65	--
<b>F Hep</b>	5.1/104.8	4.37/72.2	4.13/79.2	4.27/69.9	4.38/nd	4.20/70	(3.69/nd)/72	--
<b>G Glc</b>	5.20/103.6	3.5/72.5	4.08/79.7	3.66/71.3	3.85/nd	(3.63/3.77)/63	--	--
<b>H Hep</b>	4.81/103.1	3.88/71.05	3.78/70.8	3.76/69.3	3.58/nd	(4.02/nd)/73.7	3.71/63.8	--
<b>I Glc</b>	5.78/96.54	3.81/73	4.12/79	3.33/72.2	4.19/73.1	(4.01/3.52)/64	--	--
<b>K Gal</b>	5.52/93.78	3.94/74.3	4.09/71.6	3.83/71.3	4.09/nd	(3.73/nd)/65.5	--	--
<b>L Gal</b>	5.24/98.1	3.78/69	3.86/74.0	3.96/72.1	4.06/73.5	(3.82/nd/62)	--	--
<b>M Glc</b>	4.66/104.5	3.25/75.5	3.47/78.6	3.35/72.1	3.34	(3.73/3.79)/64	--	--

ND: not determined, --no corresponding peak.



Figure 5. 950MHz [ $^1\text{H}$ - $^{13}\text{C}$ ]-hCH-INEPT-NMR spectrum of 2mg of  $^{13}\text{C}$ -LPS from *E. coli* O157:H7 strain in buffer B at 270K with MAS of 50KHz.

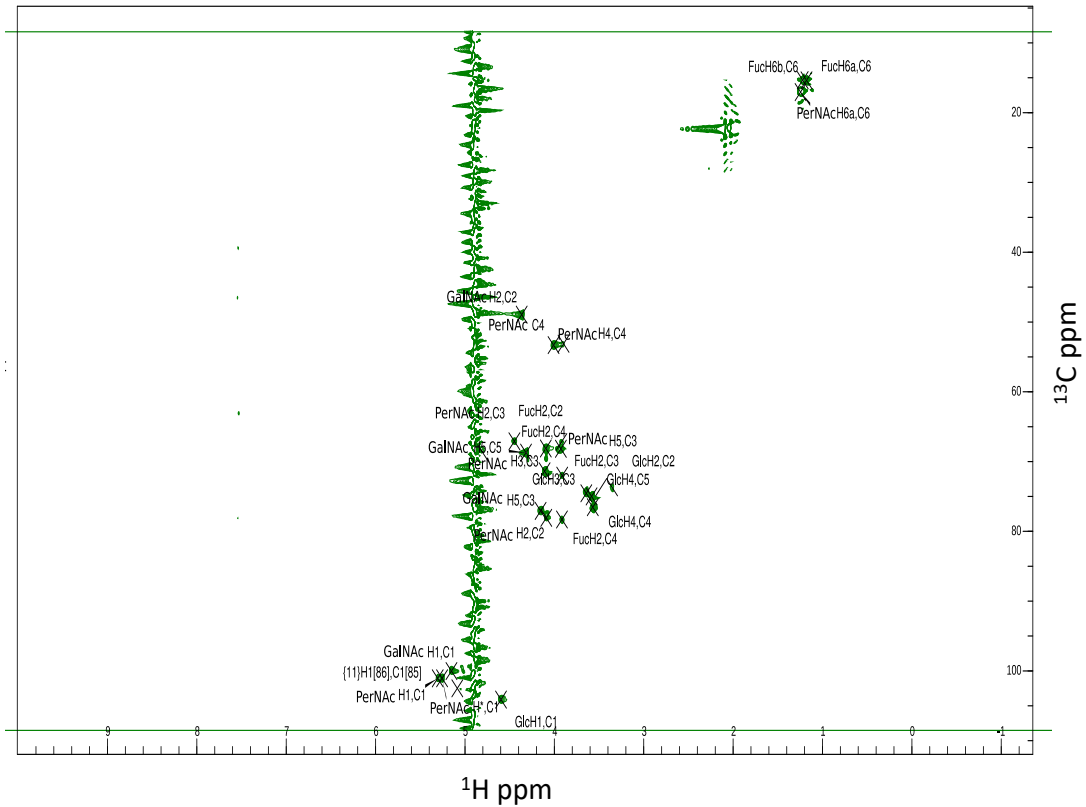
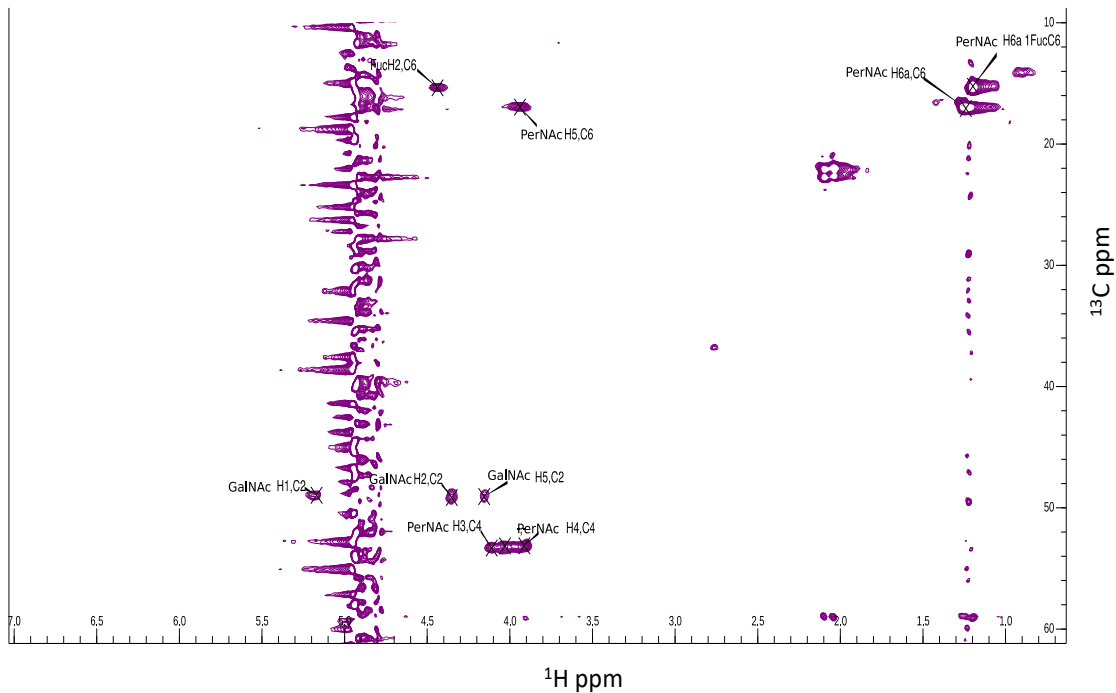


Figure 6. 950MHz [ $^1\text{H}$ - $^{13}\text{C}$ ]-hCCH-INEPT-NMR spectrum (zoomed at  $\sim 60$ -10 ppm) of 2mg of  $^{13}\text{C}$ -LPS from *E. coli* O157:H7 strain in buffer B at MAS at 270K with MAS of 50KHz.



**Table 2. hCCH-ss-NMR assignments of [<sup>1</sup>H-<sup>13</sup>C] correlations of <sup>13</sup>C-LPS O157:H7 in buffer B.**

<sup>1</sup> H position (ppm)	<sup>13</sup> C position (ppm)	Assign <sup>1</sup> H	Assign <sup>13</sup> C
5.29399	101.01838	PerNacHn2	PerNacC1
4.62205	104.18537	GlcH1	GlcC1
4.62535	73.54648	GlcH1	GlcC2
3.34363	73.52451	GlcH2	GlcC2
5.17957	99.92723	GalNacH1	GalNacC1
5.16946	48.96314	GalNacH1	GalNacC2
5.30236	77.83881	PerNacH1	PerNacC2
4.11058	77.99770	PerNacH2	PerNacC2
4.11038	68.38890	PerNacH3	None
1.24029	17.06103	PerNacH6a	PerNacC6
3.93886	16.95706	PerNacH5	PerNacC6
3.91248	53.13267	PerNacH4	PerNacC4
4.44374	67.09167	FucH2	FucC2
3.91757	67.08203	FucH2	FucC2
3.91471	72.01110	FucH*	FucC3
4.44658	72.04898	FucH2	FucC3
4.35170	49.15332	GalNacH2	GalNacC2
4.31840	68.73195	GalNacH5	C5 [60]
3.91992	78.28402	FucH1	FucC4
4.34910	77.05099	GalNacH2	GalNacC3
3.34601	104.10798	GlcH2	GlcC1
3.64537	104.10888	GlcH3	GlcC1
3.34543	74.45118	GlcH2	GlcC3
3.64401	74.52070	GlcH3	GlcC3
3.64065	73.55872	GlcH3	GlcC2
3.64451	76.56469	GlcH3	GlcC4
3.34172	76.50589	GlcH2	GlcC4
3.58610	76.59215	GlcH4	GlcC4
3.57894	74.64000	GlcH4	GlcC5
4.11032	53.32260	PerNacH3	PerNacC4
4.12124	71.43413	PerNacH3	PerNacC3
4.31747	71.44009	GalNacH2	PerNacC3
1.19811	15.23247	PerNacH6a	1FucC6
4.16198	77.09971	GalNacH5	GalNacC3
4.15504	49.06803	GalNacH5	GalNacC2
4.10445	101.08026	PerNacH3	PerNacC1
3.91972	68.18670	PerNacH5	PerNacC3
4.03302	68.30048	None	None
4.43779	15.36267	FucH2	FucC6
4.02948	53.14586	None	None

**Table 3. hCH-ss-NMR assignments of [<sup>1</sup>H-<sup>13</sup>C] correlations of <sup>13</sup>C-LPS O157:H7 in buffer B.**

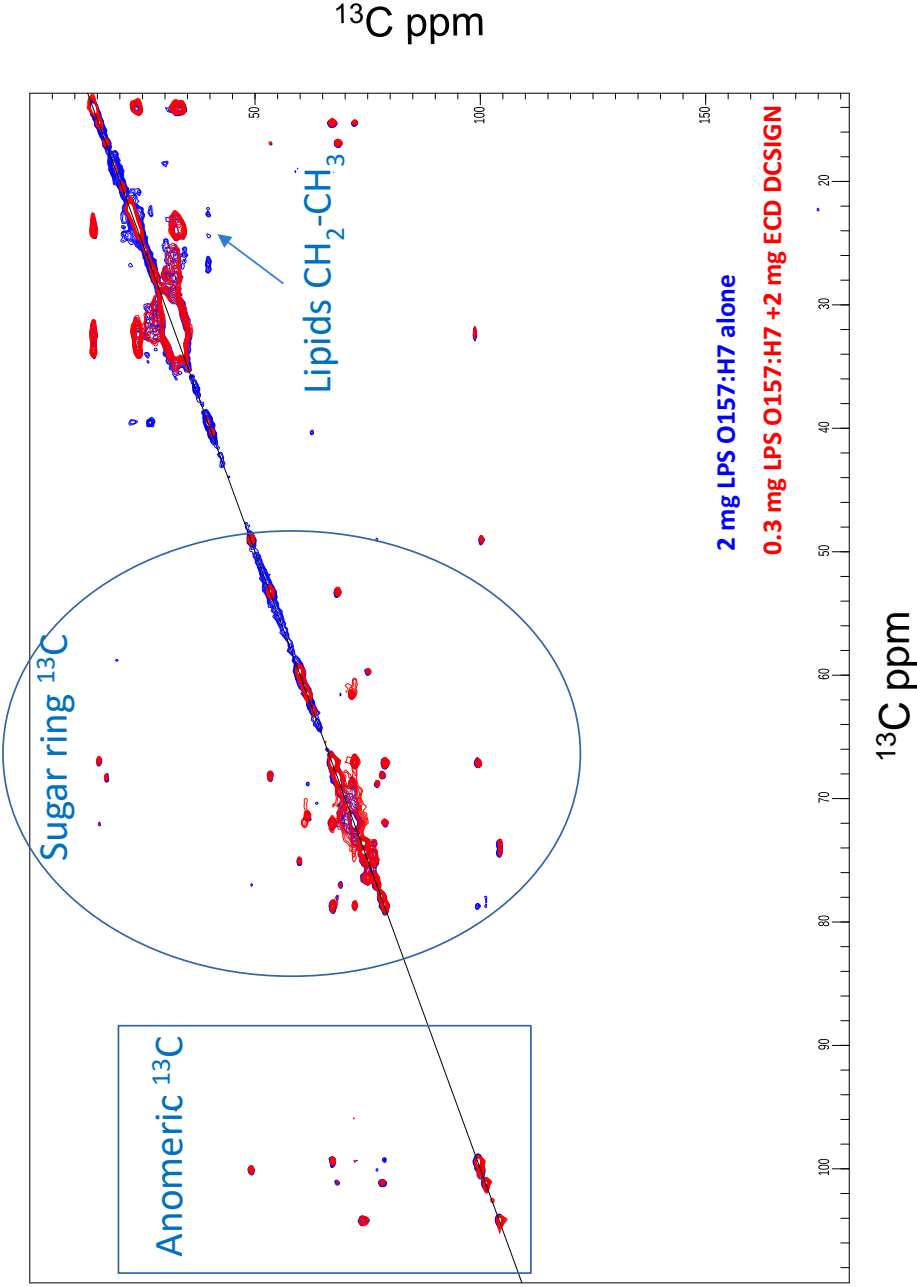
<sup>1</sup> H position (ppm)	<sup>13</sup> C position (ppm)	Assign <sup>1</sup> H	Assign <sup>13</sup> C
4.00929	53.30932	None	PerNacC4
4.36544	48.95139	GalNacH2	GalNacC2
3.89688	53.11263	PerNacH4	PerNacC4
4.44723	67.08909	FucH2	FucC4
4.31741	68.72801	GalNacH5	GalNacC5
4.10912	71.39297	PerNacH3	PerNacC3
4.09632	68.12689	PerNacH2	PerNacC3
3.92134	67.21050	FucH2	FucC2
3.91235	72.01762	FucH2	FucC3
3.92709	68.16498	PerNacH5	PerNacC3
3.64238	74.54305	GlcH3	GlcC3
3.58400	74.98944	GlcH4	GlcC5
3.57027	76.53466	GlcH4	GlcC4
3.35394	73.68460	GlcH2	GlcC2
5.30502	101.06366	PerNacH1	PerNacC1
5.25627	101.05205	PerNacH*	PerNacC1
5.15154	100.02191	GalNacH1	GalNacC1
4.59698	104.10814	GlcH1	GlcC1
4.15392	77.06730	GalNacH5	GalNacC3

4.09045	78.06280	PerNAcH2	PerNAcC2
3.91467	78.29677	FucH2	FucC4
1.24549	17.05267	PerNAcH6a	PerNAcC6
1.22139	15.33565	FucH6b	FucC6
1.17890	15.36600	FucH6a	FucC6
5.08567	102.49034	{11}H1 [86]	{11}C1 [85]

**Table 4. hCC-ss-NMR assignments of [<sup>13</sup>C-<sup>13</sup>C] correlations of <sup>13</sup>C-LPS O157:H7 in buffer B.**

<sup>13</sup> C position F1 (ppm)	<sup>13</sup> C position F2 (ppm)	Assign F1	Assign F2
72.07682	15.43109	FucC3	FucC6
66.99050	15.46456	FucC2	FucC6
68.42539	17.05685	PerNAcC3	PerNAcC6
72.01776	67.11564	FucC3	FucC2
77.00838	49.28792	GalNAcC3	GalNAcC2
78.71749	67.23817	FucC4	FucC2
59.75684	75.13906	GlcC4	GlcC5
68.15704	53.44643	PerNAcC3	PerNAcC4
104.13452	73.97305	GlcC1	GlcC2
101.13522	78.37201	PerNAcC1	PerNAcC2
68.14389	78.32094	PerNAcC3	PerNAcC2
99.43661	67.35090	FucC1	FucC2
71.94644	78.85656	FucC3	FucC4
39.54732	26.74594	None	[90]
78.02559	68.24000	PerNAcC2	PerNAcC3
100.11998	49.04158	GalNAcC1	GalNAcC2
101.11042	68.08011	PerNAcC1	PerNAcC3
99.34963	78.75489	FucC1	FucC4
100.06495	77.21420	GalNAcC1	GalNAcC3
49.07029	100.19791	GalNAcC2	GalNAcC1
53.30660	68.32156	PerNAcC4	PerNAcC3
61.59172	71.47244	C3 [64]	PerNAcC3
67.06146	72.10248	FucC4	FucC3
67.09481	78.68017	FucC2	FucC4
68.84234	77.08197	GalNAcC5	GalNAcC3
49.08164	77.03471	GalNAcC2	GalNAcC3
78.72082	72.17301	FucC4	FucC3
76.99488	68.88320	GalNAcC3	GalNAcC5
68.83744	71.51777	PerNAcC2	PerNAcC3
71.43261	61.69258	PerNAcC3	None
71.43399	68.93480	PerNAcC3	(GalNAc/PerNAc) C*
75.11774	59.68950	GlcC5	None

Figure 7. Interaction test between  $^{13}\text{C}$ -LPS O157:H7 and ECD DC-SIGN. [ $^{13}\text{C}$ - $^{13}\text{C}$ ] hCC-DARR-ss-NMR spectra.



**Table 5.  $^{15}\text{N}^{13}\text{C}$ -CRD MGL backbone resonances assignments in apo state. (chapter XI for experimental conditions). 27 spin systems, with absent cross-peaks, are not included in the table.**

<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>	<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>
6His	H	8,06462	0,00304	193Tyr	H	10,02676	0,0011
6His	N	120,16092	0,04644	193Tyr	N	121,12835	0,00835
6His	C	175,80643	0,27736	193Tyr	C	174,51667	0
6His	CA	55,87617	0	193Tyr	CA	56,80705	0
6His	CB	31,02128	0	193Tyr	CB	42,57946	0
7Ile	H	7,93738	0,00844	194Trp	H	8,3479	0,01287
7Ile	N	123,16688	0,04391	194Trp	N	122,44677	0,00917
7Ile	C	175,79064	0	194Trp	C	173,65485	0
7Ile	CA	60,7364	0,04746	194Trp	CA	56,86753	0
7Ile	CB	38,57966	0	194Trp	CB	28,27427	0
8Glu	H	8,40255	0,00715	195Phe	H	7,42132	0,00223
8Glu	N	125,54794	0,05702	195Phe	N	126,51184	6,77E-04
8Glu	C	175,6237	0	195Phe	C	173,44144	0
8Glu	CA	56,22544	0	195Phe	CB	39,29556	0
8Glu	CB	30,47539	0	195Phe	CA	55,90385	0
9Gly	H	7,94348	0,00836	196Ser	H	7,37792	0,00398
9Gly	N	116,60183	0,0333	196Ser	N	119,38959	0,02584
9Gly	C	178,85656	0	196Ser	C	172,80503	0
9Gly	CA	45,96481	0,04644	196Ser	CB	61,7887	0,04725
183Val	H	8,21491	0,00799	196Ser	CA	58,17924	0,01573
183Val	N	120,69259	0,03566	198Ser	H	6,75227	0,002
183Val	C	177,05558	0	198Ser	N	114,09423	0,0223
183Val	CA	64,37961	0	198Ser	C	171,25517	0
183Val	CB	31,55445	0	198Ser	CB	64,42884	0
184Asn	H	8,97223	0,00708	198Ser	CA	56,3202	0
184Asn	N	116,96007	0,00994	199Gly	H	8,49353	0,00779
184Asn	C	174,51131	0	199Gly	N	105,25976	0,01171
184Asn	CB	37,40699	0	199Gly	C	174,27097	0,01835
185Trp	H	8,68244	0,00105	199Gly	CA	44,48164	0
185Trp	N	121,32818	0,01086	200Met	H	9,58248	0,002
185Trp	C	176,14312	0	200Met	N	120,78928	0,0127
185Trp	CA	56,83726	0	200Met	C	173,88912	0
185Trp	CB	29,35613	0,00528	200Met	CA	56,14484	0
186Val	H	9,99502	0,00221	200Met	CB	37,32537	0
186Val	N	120,72688	0,00932	201Ser	H	8,18065	5,88E-04
186Val	C	174,99663	0	201Ser	N	114,7281	0,00841
186Val	CA	61,13791	0	201Ser	C	173,38595	0
186Val	CB	34,78306	0	201Ser	CA	58,52749	0
187Glu	H	8,68065	0,00317	201Ser	CB	64,02618	0
187Glu	N	127,97809	0,01235	202Trp	H	8,33533	0,00602
187Glu	C	176,35367	0	202Trp	N	123,60782	0,02808
187Glu	CA	55,75428	0	202Trp	C	177,68881	0
187Glu	CB	32,79756	0	202Trp	CA	62,89797	0,02572
188His	H	8,82892	1,56E-04	202Trp	CB	31,49023	0,10776
188His	N	124,64839	0,03234	202Trp	NE1	127,59733	0,00255
188His	C	173,70975	0	202Trp	HE1	10,48363	0,00309
188His	CA	58,09091	0	203Ala	H	9,11004	0,00783
188His	CB	30,67606	0	203Ala	N	116,16569	0,01496
191Ser	H	8,29602	3,42E-04	203Ala	C	181,72136	0
191Ser	N	116,88604	0,00952	203Ala	CB	18,17023	0
191Ser	C	171,29847	0	203Ala	CA	54,97432	0
191Ser	CA	57,82139	0	204Glu	H	7,25698	0,00245
191Ser	CB	68,25114	0	204Glu	N	117,95826	0,01609
192Cys	H	8,89313	0	204Glu	C	178,42593	0,01033
192Cys	N	117,9137	0	204Glu	CA	58,70526	0
192Cys	C	174,34536	0	204Glu	CB	29,51345	0
192Cys	CB	43,22345	0	205Ala	H	8,35998	0,00268
192Cys	CA	52,84875	0	205Ala	N	127,23494	0,029

<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>	<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>
205Ala	C	177,77399	0	217Val	H	8,03303	0,00152
205Ala	CA	54,48302	0,01218	217Val	N	119,50267	0,02951
205Ala	CB	16,43971	0	217Val	C	172,16463	0
206Glu	H	8,2442	0,01284	217Val	CA	62,66069	0
206Glu	N	119,2197	0,01722	217Val	CB	32,81194	0,02064
206Glu	C	179,39073	0	218Val	H	8,34508	0,00137
206Glu	CB	27,88507	0	218Val	N	128,18033	0,01182
206Glu	CA	58,87973	0	218Val	C	175,57639	0
207Lys	H	7,31909	0,00373	218Val	CA	59,70883	0
207Lys	N	118,99295	0,00335	218Val	CB	33,98431	0
207Lys	C	177,98973	0	220Asn	H	9,50592	8,50E-04
207Lys	CB	31,78147	0	220Asn	N	123,79145	0,00581
207Lys	CA	58,75683	0	220Asn	CA	55,55777	0
208Tyr	H	8,11839	0,00157	220Asn	CB	40,0049	0
208Tyr	N	120,65618	0,00852	221Ser	H	7,32432	0,00241
208Tyr	C	178,592	0	221Ser	N	108,74154	0,00557
208Tyr	CB	37,22293	0	221Ser	C	172,5269	0
208Tyr	CA	61,47692	0	221Ser	CB	66,3439	0
209Cys	H	8,48816	0,00187	221Ser	CA	56,4566	0
209Cys	N	114,75686	0,01199	223Glu	H	8,62904	0,00127
209Cys	C	177,66914	0	223Glu	N	119,3336	0,00651
209Cys	CA	55,7587	0	223Glu	C	179,0729	0
209Cys	CB	34,43609	0	223Glu	CA	60,36364	0
210Gln	H	8,36438	0,00628	223Glu	CB	28,61885	0
210Gln	N	122,5842	0,06574	224Glu	H	7,96919	9,08E-04
210Gln	C	180,648	0	224Glu	N	122,86275	0,02456
210Gln	CB	27,72263	0	224Glu	C	177,7915	0
210Gln	CA	59,62587	0	224Glu	CA	59,25988	0
211Leu	H	7,99046	9,91E-04	224Glu	CB	29,0433	0
211Leu	N	122,06301	0,01016	225Gln	H	7,84566	0,00814
211Leu	C	178,20635	0	225Gln	N	120,0649	0,02882
211Leu	CA	56,9716	0	225Gln	CB	27,77948	0
211Leu	CB	41,11703	0	225Gln	CA	58,70138	0
212Lys	H	7,26222	0,00197	226Asn	H	8,36463	0,0022
212Lys	N	117,71452	0,01684	226Asn	N	116,399	0,0058
212Lys	C	174,97449	0	226Asn	C	177,26225	0
212Lys	CA	53,31599	0	226Asn	CA	55,09348	0
212Lys	CB	30,66554	0	226Asn	CB	37,41468	0
213Asn	H	7,96744	4,14E-04	227Phe	H	7,53551	0,00415
213Asn	N	114,99661	0,0191	227Phe	N	121,0214	0,02393
213Asn	C	173,0707	0	227Phe	C	175,65994	0
213Asn	CA	54,26855	0	227Phe	CA	60,98158	0
213Asn	CB	36,43274	0	227Phe	CB	38,15847	0
214Ala	H	7,90387	0,00156	228Val	H	8,25357	4,66E-04
214Ala	N	119,97168	0,01016	228Val	N	117,73487	0,01904
214Ala	C	174,64558	0	228Val	C	178,14074	0
214Ala	CB	23,30412	0	228Val	CB	31,30056	0
214Ala	CA	49,8036	0	228Val	CA	66,32585	0
215His	H	8,34665	0,00823	229Gln	H	8,02017	0
215His	N	113,02041	0,02228	229Gln	N	113,77023	0
215His	C	175,34427	0	229Gln	C	177,20364	0
215His	CA	53,72404	0	229Gln	CA	57,72768	0
215His	CB	32,68363	0	229Gln	CB	27,4295	0
216Leu	H	9,81237	0,01027	230Lys	H	7,0379	0,00151
216Leu	N	125,14651	0,03307	230Lys	N	117,71942	2,23E-04
216Leu	C	179,73321	0	230Lys	C	176,92835	0
216Leu	CB	43,80595	0	230Lys	CB	31,61394	0
216Leu	CA	56,36335	0	230Lys	CA	58,42067	0

<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>	<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>
231Tyr	H	7,01424	0,00115	248Trp	C	176,19794	0
231Tyr	N	115,842	0,00928	249Lys	H	8,97928	0
231Tyr	C	175,00352	0	249Lys	N	120,74965	0
231Tyr	CB	38,51728	0	249Lys	C	176,98141	0
231Tyr	CA	57,31328	0	249Lys	CA	54,78285	0
232Leu	H	6,65691	0,00416	249Lys	CB	35,96952	0
232Leu	N	118,8798	0,02403	250Trp	H	8,84552	0,01126
232Leu	C	177,9802	0	250Trp	N	125,01791	0,01064
232Leu	CA	55,34734	0	250Trp	C	181,81834	0
232Leu	CB	41,85989	0	250Trp	CA	56,13219	0
233Gly	H	8,42997	0,00303	250Trp	CB	32,59378	0
233Gly	N	109,40505	0,03789	251Val	H	9,15774	0,00143
233Gly	C	173,87081	0	251Val	N	116,17856	0,03369
233Gly	CA	44,5941	0	251Val	C	175,794	0
237Thr	H	7,12069	4,27E-04	251Val	CA	64,1163	0
237Thr	N	116,02132	0,03211	251Val	CB	32,48411	0
237Thr	C	173,40366	0	252Asp	H	7,22812	0,00288
237Thr	CA	61,39009	0	252Asp	N	117,34063	0,00427
237Thr	CB	72,59913	0	252Asp	C	177,42989	0
238Trp	HE1	10,63406	0,00413	252Asp	CB	40,97756	0
238Trp	NE1	129,89664	0,06645	252Asp	CA	52,4863	0
239Met	H	8,49958	3,99E-04	253Gly	H	7,91776	1,12E-04
239Met	N	113,57265	0,01361	253Gly	N	107,24434	0,00649
239Met	C	176,21924	0	253Gly	C	174,77362	0
239Met	CA	52,80253	0	253Gly	CA	44,92643	0,01456
239Met	CB	37,37294	0,05024	254Thr	H	8,16865	0,00214
240Gly	H	10,23679	0,00208	254Thr	N	120,74896	0,00662
240Gly	N	110,76895	0,00156	254Thr	C	174,95855	0,03591
240Gly	C	172,29432	0	254Thr	CA	65,10833	0,03191
240Gly	CA	50,00816	0,05784	254Thr	CB	68,46641	0,03521
241Leu	H	8,23399	8,94E-04	255Asp	H	8,97164	0,00844
241Leu	N	128,49386	0,0041	255Asp	N	128,02388	0,00486
241Leu	C	173,81598	0,00677	255Asp	C	176,54434	0
241Leu	CA	54,4057	0	255Asp	CA	56,07443	0
241Leu	CB	45,94084	0	255Asp	CB	42,0457	0
242Ser	H	8,4485	7,79E-04	256Tyr	H	8,81314	0,00129
242Ser	N	119,69475	0,0161	256Tyr	N	127,34971	0,01541
242Ser	CA	58,09533	0	256Tyr	C	179,38508	0
242Ser	CB	65,27255	0	256Tyr	CA	59,99958	0
242Ser	C	172,28895	0	256Tyr	CB	39,32608	0
245Glu	H	8,30456	0,00339	257Ala	H	9,21714	0,01097
245Glu	N	116,1253	0,007	257Ala	N	123,24202	0,03686
245Glu	C	176,92817	0	257Ala	C	179,61646	0
245Glu	CA	55,43779	0	257Ala	CB	18,59874	0
245Glu	CB	28,68189	0	257Ala	CA	55,45388	0
246Gly	H	7,65765	0,00398	258Thr	H	7,58439	0,00221
246Gly	N	107,42585	0,01601	258Thr	N	103,98526	0,00701
246Gly	C	172,66956	0	258Thr	C	174,94025	0
246Gly	CA	45,01535	0	258Thr	CA	60,84729	0
247Ala	H	6,88291	0,00249	258Thr	CB	69,57961	0
247Ala	N	120,96586	0,00363	259Gly	H	7,25304	5,46E-04
247Ala	C	176,12816	0	259Gly	N	111,12314	0,02191
247Ala	CB	21,10804	0	259Gly	C	173,04433	0
247Ala	CA	50,05222	0	259Gly	CA	44,26595	0
248Trp	H	8,11404	0,00884	260Phe	H	8,30475	0
248Trp	N	124,43037	0,02363	260Phe	N	121,2468	0
248Trp	CA	58,29107	0	260Phe	C	175,05174	0
248Trp	CB	27,75654	0	260Phe	CB	39,8261	0

<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>	<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>
260Phe	CA	58,79749	0	282Cys	H	8,53174	0,00368
261Gln	H	7,03957	0,00903	282Cys	N	116,22154	0,07353
261Gln	N	123,71662	0,02436	282Cys	CB	48,44278	0
261Gln	C	173,42846	0	282Cys	CA	58,99633	0
261Gln	CA	53,71707	0	282Cys	C	172,40292	0
261Gln	CB	31,91441	0	283Ala	H	8,55617	0,00259
263Trp	H	7,24441	0,00739	283Ala	N	125,71484	0,0163
263Trp	N	120,00984	0,03085	283Ala	C	175,60865	0
263Trp	C	178,02576	0	283Ala	CA	50,86677	0
263Trp	CA	57,40847	0	283Ala	CB	21,17337	0,04143
263Trp	CB	29,96643	0	284His	H	9,279	0,00276
264Lys	H	9,71771	0,00501	284His	N	116,14812	1,75E-04
264Lys	N	126,45381	0,0442	284His	C	173,50039	0
264Lys	C	173,43243	0	284His	CA	53,14673	0
264Lys	CB	33,34103	0	284His	CB	33,62649	0,03694
264Lys	CA	54,67386	0	285Phe	H	9,73434	0,00133
267Gln	H	8,19961	8,53E-04	285Phe	N	116,01581	0,00543
267Gln	N	118,24217	0,01582	285Phe	C	181,21533	0
267Gln	C	172,64559	0	285Phe	CA	57,87112	0
267Gln	CA	51,30411	0	285Phe	CB	41,24428	0
267Gln	CB	28,05049	0	286His	H	9,28407	0,0022
269Asp	H	9,58294	0,00386	286His	N	126,74732	0,04459
269Asp	N	125,10722	0,0436	286His	C	175,61304	0
269Asp	C	176,77917	0	286His	CA	56,56915	0
269Asp	CA	52,78155	0	286His	CB	30,48357	0
269Asp	CB	41,07143	0	289Gly	H	8,89175	9,94E-04
271Trp	H	7,80477	0	289Gly	N	108,96062	0,02253
271Trp	N	120,49649	0	289Gly	C	174,05874	0
271Trp	C	176,66187	0	289Gly	CA	44,92982	0
271Trp	CA	59,2902	0	290Arg	H	8,24864	4,22E-04
271Trp	CB	27,84194	0	290Arg	N	119,77651	0,01707
271Trp	NE1	129,95724	0	290Arg	C	177,07561	0
271Trp	HE1	10,36139	0	290Arg	CA	56,44291	0
272Gln	H	7,77089	0,00262	290Arg	CB	30,56197	0
272Gln	N	119,39266	0,0441	291Trp	H	8,30735	0,00102
272Gln	C	178,56683	0	291Trp	N	118,88653	0,00614
272Gln	CB	29,09183	0	291Trp	C	176,42008	0
272Gln	CA	54,38665	0	291Trp	CA	53,40069	0
277Gly	H	8,14876	0,00215	291Trp	CB	30,99901	0
277Gly	N	109,82006	0,03715	292Asn	H	8,66105	9,15E-04
277Gly	C	174,00307	0	292Asn	N	118,93046	0,0417
277Gly	CA	44,55289	0	292Asn	C	172,38127	0
278Gly	H	8,37845	0,02138	292Asn	CA	51,77069	0
278Gly	N	108,5769	0,02643	292Asn	CB	42,42563	0
278Gly	CA	44,59556	0	294Asp	H	10,09865	0,00197
279Gly	H	7,97517	1,30E-04	294Asp	N	123,8242	0,03315
279Gly	N	107,53179	0,01551	294Asp	C	174,42749	0
279Gly	C	174,91083	0	294Asp	CA	51,3179	0
279Gly	CA	44,41543	0	294Asp	CB	46,58721	0
280Glu	H	8,46145	0,00285	296Cys	H	8,66883	0,00696
280Glu	N	124,69273	0,04493	296Cys	N	123,52254	0,041
280Glu	C	177,49606	0	296Cys	C	174,72707	0
280Glu	CA	56,59369	0	296Cys	CA	58,63572	0
280Glu	CB	31,58655	0,04316	296Cys	CB	45,38847	0
281Asp	H	7,79749	0,00283	297Gln	H	7,61198	0,00589
281Asp	N	124,76121	0,00888	297Gln	N	115,51574	0,01175
281Asp	C	179,20802	0	297Gln	C	175,64755	0
281Asp	CB	41,65207	0	297Gln	CA	56,62892	0



<i>Residue</i>	<i>Nuclei</i>	<i>Shift (ppm)</i>	<i>SD</i>
297Gln	CB	28,85595	0
298Arg	H	7,83552	0
298Arg	N	119,98083	0
298Arg	C	174,80001	0
298Arg	CB	27,73027	0
298Arg	CA	54,58142	0
300Tyr	H	7,57686	0,00353
300Tyr	N	126,71753	0,02908
300Tyr	C	176,95876	0
300Tyr	CA	52,78489	0
300Tyr	CB	35,29682	0
301His	H	7,71819	7,95E-04
301His	N	117,96324	0,02616
301His	C	172,90884	0
301His	CA	56,71901	0
301His	CB	29,62553	0
302Trp	H	9,11856	0
302Trp	N	116,85123	0
302Trp	C	172,40989	0
302Trp	CA	58,48122	0
302Trp	CB	32,63768	0
303Val	H	6,67499	0,00381
303Val	N	117,84164	0,12189
303Val	CA	59,2336	0
303Val	C	177,471	0
303Val	CB	35,58008	0
304Cys	H	9,45293	0,00173
304Cys	N	122,38493	0,00745
304Cys	C	172,37791	0
304Cys	CA	50,84196	0,39307
304Cys	CB	39,96687	0
305Glu	H	9,87796	0,00227
305Glu	N	124,49249	0,027
305Glu	C	173,15392	0
305Glu	CA	54,35361	0
305Glu	CB	33,88255	0
306Ala	H	9,28009	5,65E-04
306Ala	N	128,95276	0,03049
306Ala	C	176,63436	0
306Ala	CA	49,64911	0
306Ala	CB	23,7179	0,03805
307Gly	H	8,45327	0,00253
307Gly	N	107,19147	0,03401
307Gly	C	173,86118	0
307Gly	CA	44,95436	0
308Leu	H	8,00699	0,00156
308Leu	N	122,63561	0,03885
308Leu	C	178,50288	0
308Leu	CB	42,45474	0
308Leu	CA	55,93625	0
309Gly	H	8,57046	0,00268
309Gly	N	110,8753	0,023
309Gly	C	174,22429	0
309Gly	CA	45,41496	0
316His	H	7,88928	0,00329
316His	N	126,75109	0,03394
316His	C	179,921	0
316His	CB	31,25069	0
316His	CA	57,58636	0

Figure 8.  $^1\text{H}$ -1D-NMR spectra focused on the anomeric region of CRD MGL-OS R1 (A) MGL-OS R3 (B) and MGL-LPS O157:H7 (C) mixtures in buffer B at 303K.

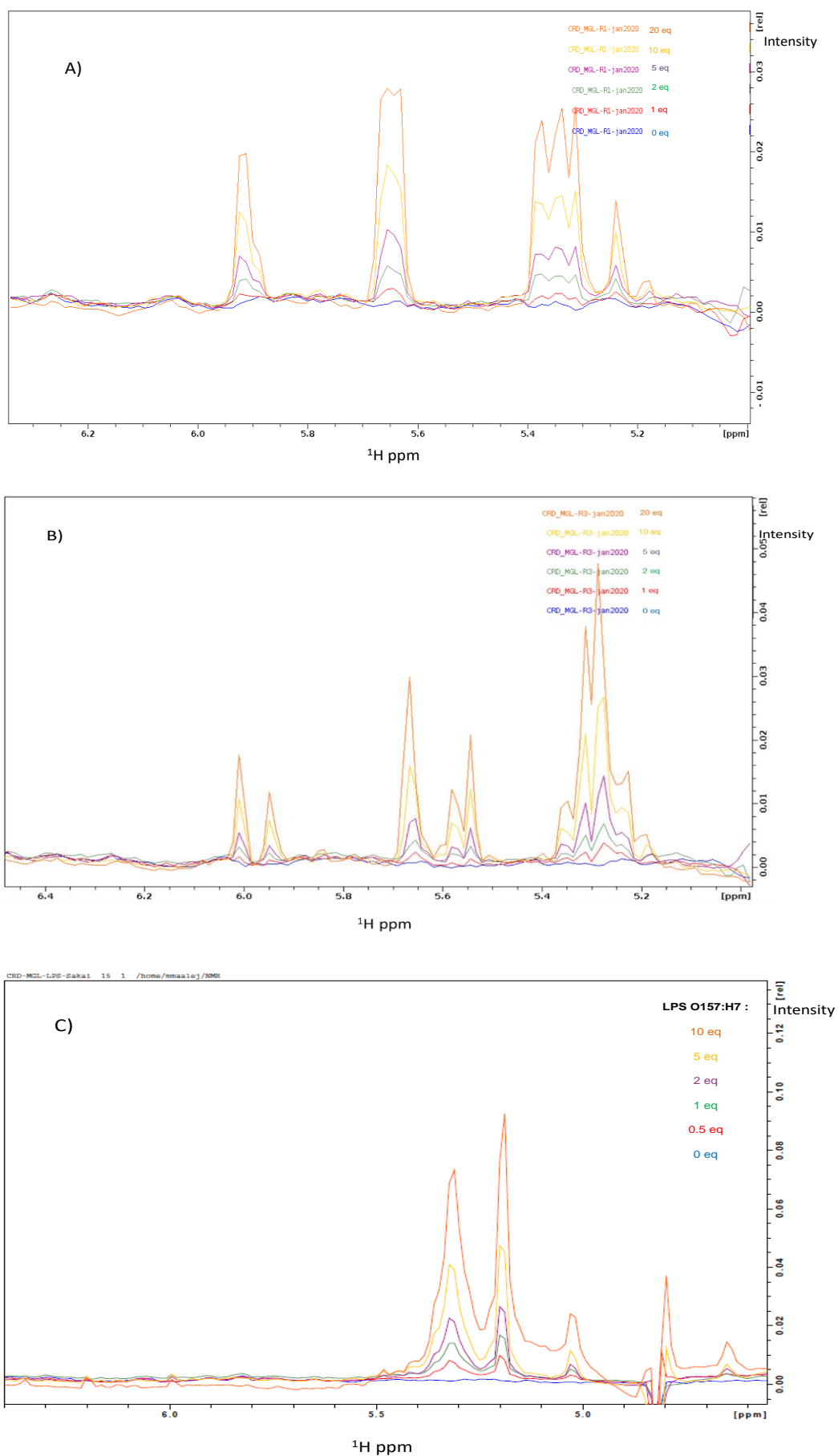


Figure 9. Global fitting of CSPs (above threshold) from residues experiencing perturbations following LPS O157:H7 addition during NMR titrations.

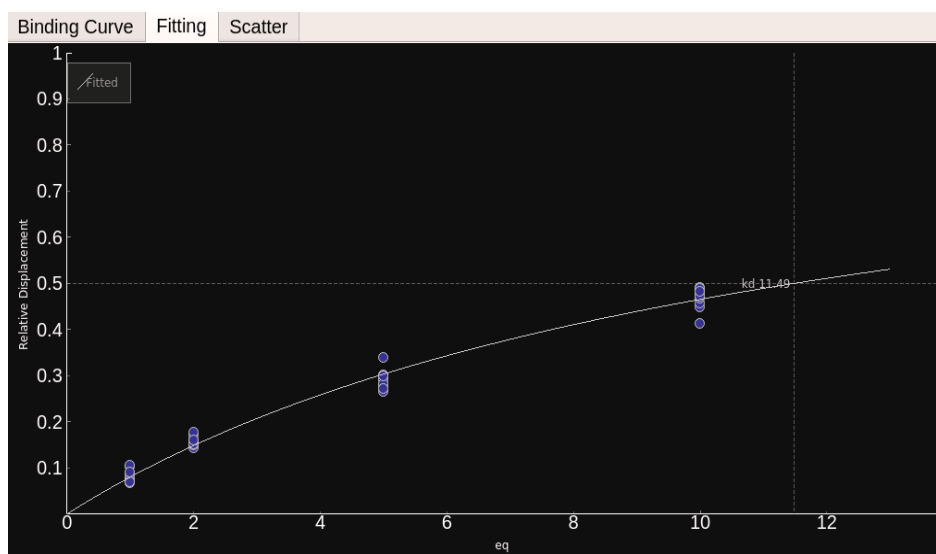
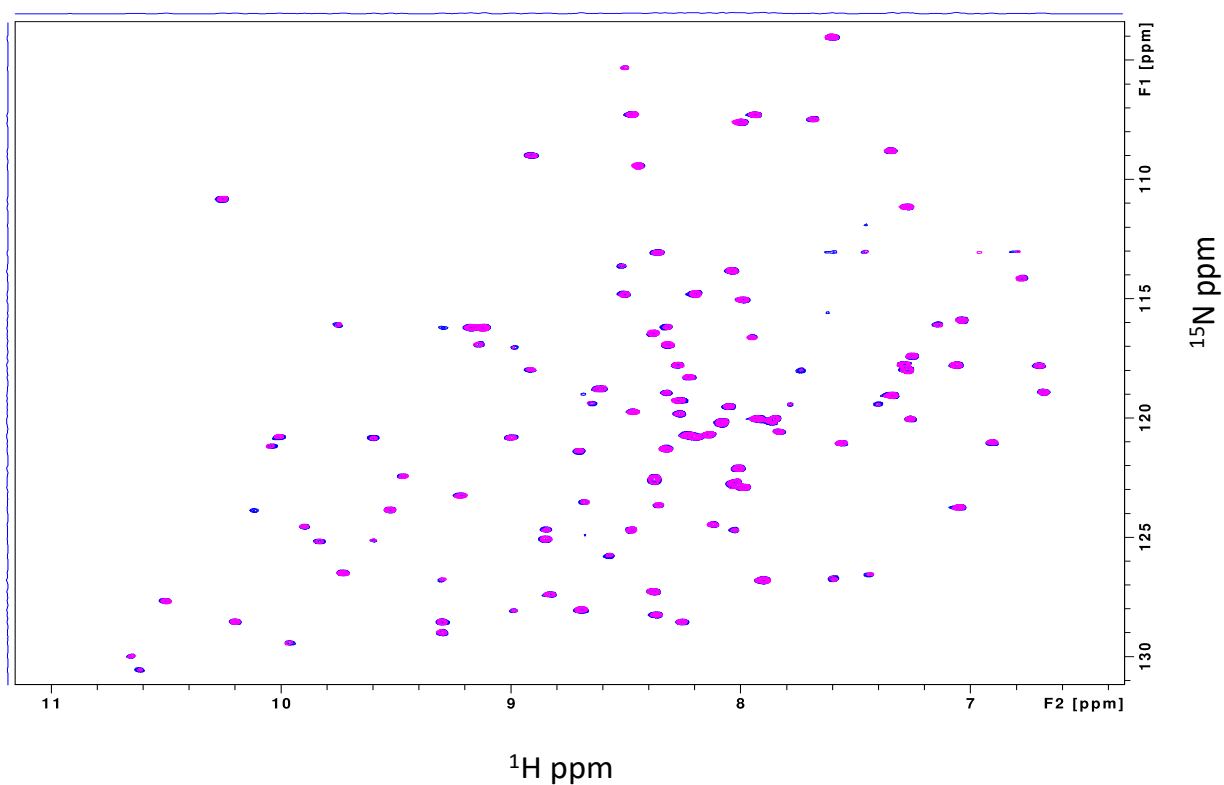
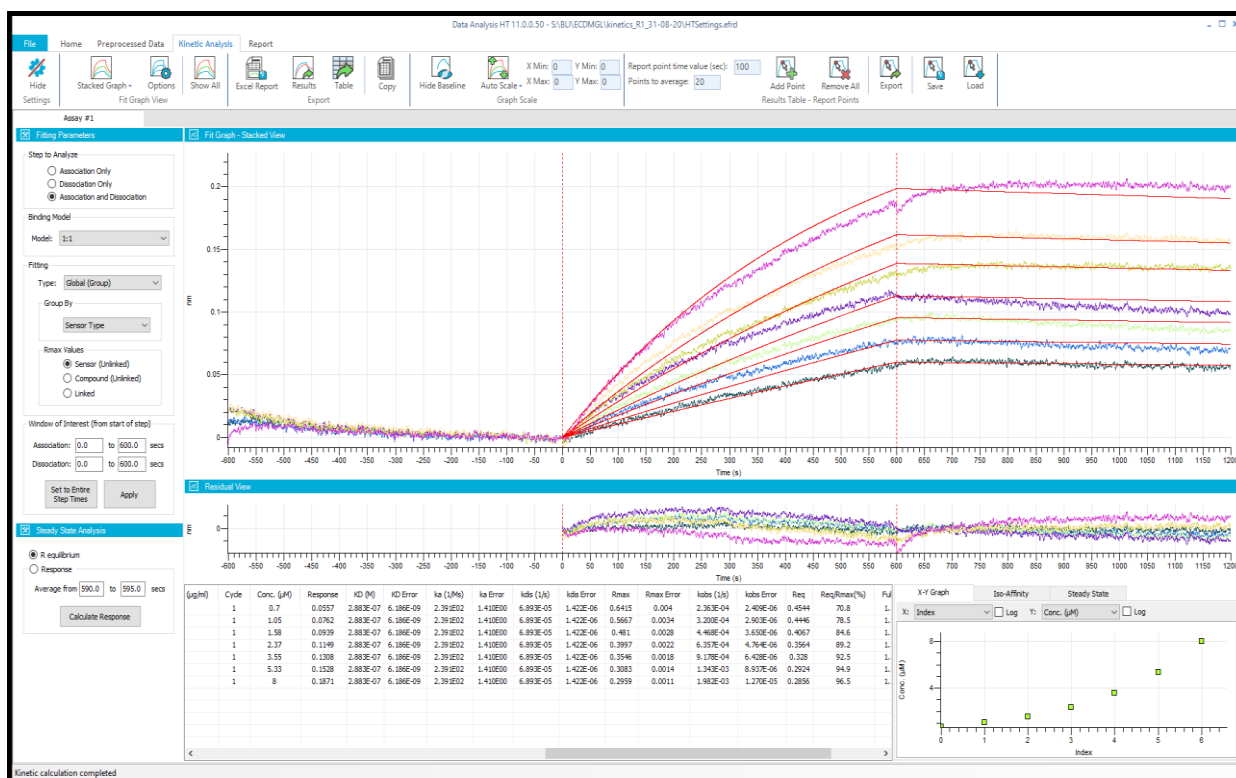


Figure 10.  $[^1\text{H},^{15}\text{N}]$ -2D-Best-TROSY-HSQC superimposed spectra of  $50\ \mu\text{M}$   $^{15}\text{N}$ -CRD MGL in apo (blue) and at 20 eq. molar ratio of DDM (Pink). For simplicity, backbone resonances are not labelled.





**Figure 12. BLI analyses of the interaction between human biotinylated ECD MGL (53 nM) and LOS R1 at different concentrations (0.7-8  $\mu$ M).** Kinetic parameters as well as LOS R1 tested concentrations (0.7- 8  $\mu$ M) are indicated in the table generated from Data analysis HT 11.1.0.25<sup>®</sup> software.



**Figure 13. BLI analyses of the interaction between human biotinylated ECD MGL (53 nM) and LOS R3 and LPS O157:H7 at different concentrations (0.7-8  $\mu$ M).**

