



HAL
open science

Investigation of LPS recognition by immunity C-type lectin receptors in a cell-surface mimicking environment

Massilia Abbas

► **To cite this version:**

Massilia Abbas. Investigation of LPS recognition by immunity C-type lectin receptors in a cell-surface mimicking environment. Structural Biology [q-bio.BM]. Université Grenoble Alpes [2020-..], 2024. English. NNT : 2024GRALV032 . tel-04878955

HAL Id: tel-04878955

<https://theses.hal.science/tel-04878955v1>

Submitted on 15 Jan 2025

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'UNIVERSITÉ GRENOBLE ALPES

École doctorale : CSV- Chimie et Sciences du Vivant

Spécialité : Biologie Structurale et Nanobiologie

Unité de recherche : Institut de Biologie Structurale

Etude de la reconnaissance du LPS par les récepteurs immunitaires lectine de type C dans un environnement imitant la surface cellulaire

Investigation of LPS recognition by immunity C-type lectin receptors in a cell-surface mimicking environment

Présentée par :

Massilia ABBAS

Direction de thèse :

Cedric LAGURI
CHARGE DE RECHERCHE, CNRS
Michel THEPAUT
CNRS, IBS

Directeur de thèse

Co-encadrant de thèse

Rapporteurs :

Alba SILIPO
FULL PROFESSOR, Università degli Studi Napoli Federico 2
Latifa ELANTAK
CHARGÉE DE RECHERCHE HDR, CNRS délégation Provence et Corse

Thèse soutenue publiquement le **9 septembre 2024**, devant le jury composé de :

Olivier RENAUDET , PROFESSEUR DES UNIVERSITES, Université Grenoble Alpes	Président
Cedric LAGURI , CHARGE DE RECHERCHE HDR, CNRS délégation Alpes	Directeur de thèse
Alba SILIPO , FULL PROFESSOR, Università degli Studi Napoli Federico 2	Rapporteure
Latifa ELANTAK , CHARGÉE DE RECHERCHE HDR, CNRS délégation Provence et Corse	Rapporteure
Julie BOUCKAERT , DIRECTRICE DE RECHERCHE, CNRS délégation Hauts-de-France	Examinatrice

Invités :

Michel Thépaut
INGENIEUR DE RECHERCHE, CNRS.



Acknowledgements

Over the course of my three years at the IBS, I have received help, support and guidance from countless individuals. First and foremost, I would like to thank my supervisor, Michel Thepaut. Your support, thoughtful advice and keen ability to recognize when I was overwhelmed have been invaluable. It has been a pleasure to work with you.

I am grateful to the members of my PhD thesis defense jury: Prof. Olivier Renaudet, Prof. Alba Silipo, Dr. Latifa Elantak and Dr. Julie Bouckaert. Thank you for your valuable insights and discussions during the defense. Likewise, I would like to thank the members of my PhD thesis committee Dr. Anne Imberty, Dr. Lauriane Lecoq, Dr. Annabelle Varrot and Dr. André Zapun, for their thoughtful feedback throughout my PhD journey. Thank you to Dr. Samantha Micciulla, Dr. Jean-Luc Pellequer and Dr. Jean-Marie Teulon for your help and contribution to this work.

I would like to express my deepest gratitude to everyone who supported me along this path, especially to the two research groups that I had the privilege of being part of.

I am sincerely thankful to all the biomolecular NMR Spectroscopy group members I worked alongside during my PhD. Your guidance, and encouragement were so much appreciated, and I learned a lot from each one of you. Bernhard and Adrien thank you for your invaluable support and kindness. Alicia, thank you for your assistance with the NMR spectrometer, your contribution to the project and beyond. Lionel, thank you for making the lab such an enjoyable place to work.

A special thank you to Catherine and Beate. Your endless support, guidance and mentorship truly meant the world to me. To Rida, Rime, and Isabel, I am deeply grateful for your continuous assistance and support throughout this journey. Thank you all for being there to help with every question and challenge I faced. Working alongside you and getting to know each of you made this experience so much more enjoyable, and I will always cherish those moments we shared.

Faustine and Béatrice, thank you for all the fun moments, the mutual support, and great conversations we shared in the office. To Arthur, Annelise, Arijit, and Astrid, thank you for creating such a positive atmosphere in the lab and for all the shared time, thoughts, and laughter. You all made my time in Grenoble truly unforgettable!

Acknowledgements

I am equally grateful to all the Membrane & Pathogen group members who warmly welcomed me into their lab during the final year of my PhD. Franck, thank you for your insightful feedback and meaningful discussions. Isabelle, Aline, Estelle, and Claudine, thank you for your assistance. A special thanks to Perrine, Maria, Serena, and Lama. Your friendship and support made this last year much more pleasant and enjoyable.

On a personal note, I am profoundly thankful to my family for your unconditional love, support, and encouragement. Your belief in me has been my greatest source of strength. The biggest thanks of all goes to Nassim. Thank you for your incredible patience and support throughout this journey. You have always believed in me and lifted me up during my low moments and consistently given me confidence in myself. You have been truly awesome.

Abstract

The envelope of Gram-negative bacteria comprises Lipopolysaccharides (LPSs) representing the main glycolipid component of its surface. LPSs are constituted of three main parts: lipid A linked to an Oligosaccharide chain (OS), which in turn is linked to an O-antigen polysaccharide portion. Given their dense packing and structural variability, LPSs are key elements in antimicrobial resistance and virulence. As surface exposed components, they are potent activators of the immune system of plants, animals, and humans. Their lipid A moiety is detected by the immune system either extracellularly by the TLR4 cascade or intracellularly by the caspase system. Whereas, their glycan part is found to be recognized by C-type Lectin Receptors (CLRs) present on Antigen Presenting Cells (APC). A protein family to which key roles have been attributed in host defence and homeostasis.

In this study, we investigated the interaction involving human Macrophage Galactose-type Lectin MGL and *E. coli* surface glycans. We demonstrated the ability of MGL to bind *E. coli* R1 core OS by integrative approaches spanning from the cellular to the atomic level. Fluorescence microscopy and flow cytometry primarily revealed the strong ability of MGL to bind *E. coli* R1 type core surfaces, while SPR provided an estimation of the interaction affinity. Nevertheless, this interaction was found to occur regardless of the canonical calcium-dependent glycan binding site. NMR spectroscopy was used to identify a novel carbohydrate binding site on the opposite surface of the canonical interaction site within MGL Carbohydrate Recognition Domain (CRD). A model of the trimeric MGL was built using a combination of Small-Angle X-ray Scattering (SAXS) and AlphaFold modelling. This model showed the convenient 3D arrangement of MGL CRDs presenting up to six accessible glycan binding sites (2 per CRD) favourable to bind LPSs at the bacterial surface with enhanced affinity.

MGL interaction with bacterial glycans was further monitored in a cell surface mimicking model using Styrene-Maleic Acid (SMA) copolymer LPS nanodiscs. To that end, a protocol for the preparation of LPS nanodiscs originating from various strains ranging from laboratory to pathogenic *E. coli* strains was established. This protocol was successfully applied on purified LPS and LPS extracted from outer membranes. The resulting membrane-mimetic models were studied and proved suitable for several biophysical methods; their size distribution and thickness were assessed by Atomic Force Microscopy. The distinct components of bacterial outer membranes could be observed at atomic scale by solid-state NMR. LPS nanodiscs have been effectively employed to monitor interactions with MGL immunity C-type lectin by Quartz Crystal Microbalance with Dissipation monitoring (QCM-D) and BioLayer Interferometry

(BLI), and now will be used to build a model depicting the arrangement of MGL ECD on LPS membrane nanodiscs.

Résumé

L'enveloppe des bactéries Gram-négatives est constituée de LipoPolySaccharides (LPS) qui représentent le principal composant glycolipidique de sa surface. Les LPSs sont constitués de trois parties principales : le lipide A, lié à une chaîne Oligo-Saccharidique (OS), qui elle-même est liée à une partie polysaccharidique O-antigène. En raison de leur densité et de leur variabilité structurelle, les LPS sont des éléments clés de la résistance aux antimicrobiens et de la virulence. En tant que composants exposés en surface, ils sont considérés comme activateurs du système immunitaire des plantes, des animaux et des humains. La partie lipide A est détectée par le système immunitaire soit de manière extracellulaire par la cascade TLR4, soit de manière intracellulaire par le système des caspases. En revanche, leur partie glycane est reconnue par les récepteurs de lectine de type C (CLR) présents sur les cellules présentatrices d'antigènes (APC). Il s'agit d'une famille de protéines à laquelle des rôles clés ont été attribués dans la défense et l'homéostasie de l'hôte.

Dans cette étude, nous avons étudié les interactions entre la lectine MGL de type galactose des macrophages humains et les glycanes de surface d'*E. coli*. Nous avons démontré la capacité de la MGL à se lier à l'OS R1 d'*E. coli* par des approches intégratives allant du niveau cellulaire au niveau atomique. La microscopie à fluorescence et la cytométrie en flux ont premièrement révélé la forte capacité du MGL à lier les surfaces d'*E. coli* de type R1, tandis que SPR a fourni une estimation de l'affinité de cette interaction. Néanmoins, il a été constaté que cette interaction se produisait indépendamment du site canonique de liaison des glycanes dépendant du calcium. La spectroscopie RMN a été utilisée pour identifier un nouveau site de liaison des glucides sur la surface opposée au site d'interaction canonique dans le domaine de reconnaissance des glucides (CRD) de MGL. Un modèle de la MGL trimérique a été construit en combinant la diffusion des rayons X aux petits angles et la modélisation AlphaFold. Ce modèle a montré l'arrangement 3D particulier des CRD de la MGL présentant jusqu'à six sites de liaison de glycanes (2 par CRD) favorisant ainsi la liaison des LPS à la surface de la bactérie avec une affinité accrue.

L'interaction de la MGL avec les glycanes bactériens a été étudié dans un modèle imitant la surface cellulaire en utilisant des LPS reconstitués en nanodisque avec du copolymère d'acide

styrène-maléique. À cette fin, un protocole de préparation de nanodisques de LPS provenant de diverses souches, allant de souches de laboratoire à des souches pathogènes d'*E. coli*, a été établi. Ce protocole a été appliqué avec succès à des LPS purifiés et à des LPS extraits de membranes externes. Les modèles membranaires mimétiques obtenus ont été étudiés et se sont révélés compatibles avec plusieurs méthodes biophysiques ; leur distribution de taille et leur épaisseur ont été évaluées par microscopie à force atomique. Les différents composants des membranes externes bactériennes ont pu être observés à l'échelle atomique par RMN du solide. Les nanodisques LPS ont également été utilisés avec succès pour étudier les interactions avec la lectine de type C de l'immunité MGL par QCM-D et BLI, et seront maintenant utilisés pour construire un modèle détaillé décrivant l'arrangement de l'ECD de la MGL sur les nanodisques de LPS de membrane.

Table of contents

INTRODUCTION	1
I. Bacterial cell envelope: Composition and architecture	1
I.1. Bacterial membranes: structure and properties	1
I.2. The outer membrane: a distinguishing feature of Gram-negative bacteria.....	3
I.3. Lipopolysaccharides: structure and biogenesis	5
I.3.1. LPS chemical structure and nomenclature	5
I.3.2. LPS Biogenesis and transport	7
I.3.2.1. Lipid A biogenesis	7
I.3.2.2. Core oligosaccharide and O-antigen biogenesis	8
I.3.2.3. LPS transport to the outer membrane	10
I.4. LPS Leaflet: an effective permeability barrier	15
I.4.1. Strong lateral interactions between LPS molecules	15
I.4.2. Modification of LPS Structure	16
II. Immunogenicity of LPS	18
II.1. LPS detection in mammals	18
II.1.1. TLR4-dependent detection of LPS.....	18
II.1.2. TLR4-independent detection of LPS.....	21
II.1.2.1. Extracellular detection of LPS.....	21
II.1.2.2. Intracellular detection of LPS	23
II.2. LPS detection in plants	25
II.3. LPS detection by lectins.....	26
II.3.1. C-type lectin receptors: overview	26
II.3.2. CLRs in bacterial glycans recognition	30
II.3.3. Focus on Macrophage Galactose-type Lectin MGL	32
II.3.3.1. Structural features.....	32
II.3.3.2. Functional features.....	34
II.4. Bacterial escaping strategies: pathogens and commensals	35
III. LPS: an antibiotic target.....	36
III.1. Antibiotic discovery and mode of action	36
III.2. Mechanism of antibiotic resistance used by bacteria	38
III.2.1. Resistance through target modification and protection	39
III.2.2. Resistance through antibiotic inactivation and degradation	39

III.2.3. Resistance through active efflux.....	40
III.2.4. Resistance through influx reduction and permeability control	41
III.3. LPS: a distinguishing target against bacterial resistance	42
III.3.1. Targeting LPS biosynthesis	42
III.3.2. Targeting LPS transport	44
III.3.3. LPS direct targeting with polymyxin antibiotics	45
IV. State-of-the-art approaches for LPS studies	48
IV.1. LPS structural characterization.....	48
IV.1.1. LPS extraction and fractionation methodologies.....	48
IV.1.1.1. LPS isolation and detection	48
IV.1.1.2. LPS fractionation	49
IV.1.2. LPS compositional analysis.....	50
IV.1.3. LPS structure elucidation.....	53
IV.1.3.1. Mass spectrometry	53
IV.1.3.2. Nuclear magnetic resonance spectroscopy	54
IV.2. LPS-lectin receptor interaction studies.....	58
IV.2.1. Glycan arrays for LPS-lectin interaction screening	58
IV.2.2. Lectin arrays for host–microorganism interaction screening.....	60
IV.2.3. Imaging approaches for LPS-lectin interaction studies.....	62
IV.2.3.1. Atomic Force Microscopy	62
IV.2.3.1. Flow cytometry and fluorescence microscopy	63
IV.2.4. NMR spectroscopy for LPS-lectin interaction studies.....	65
IV.2.4.1. Saturation Transfer Difference NMR spectroscopy	65
IV.2.4.2. Chemical shift perturbation	67
IV.2.5. Other approaches for carbohydrate-lectin interaction studies.....	69
IV.3. Position of the PhD project as it relates with the state-of-the-art	71
IV.3.1. What about Styrene-Maleic Acid Lipid Particles: SMALPs.....	73
Results	76
V. MGL binds <i>E. coli</i> surface with high avidity.....	76
V.1. Production of recombinant MGL ECD lectin	78
V.1.1. MGL ECD constructions	78
V.1.2. MGL ECD expression in minimum media.....	80
V.1.1. Refolding and purification of MGL ECD	82
V.2. MGL ECD binds <i>E. coli</i> OSs <i>via</i> a new binding interface	86

V.3. SPR interaction analysis of LOS micelles with CLRs	89
V.4. Flow cytometry analysis of MGL binding onto R1 and R3 cells	92
V.5. What about MGL binding to pathogenic <i>E. coli</i> strains?	93
V.6. Published article	96
V.7. Conclusion and discussion	121
VI. Biomimetic LPS nanodiscs for interaction studies at the surface of Gram-negative bacteria	124
VI.1. Isolation and characterization of SMA LPS nanodiscs from <i>E. coli</i> strains	126
VI.2. Structural analysis of LPS nanodiscs by solid-state NMR	131
VI.3. SMA LPS Nanodiscs are a good model to explore interactions with antibiotics and immune receptor proteins	144
VI.4. Assessment of LOS nanodiscs interactions using QCM-D	147
VI.5. Conclusion and discussion	151
VII. Structural studies of LPS-MGL interactions in membrane mimetics	154
VII.1. Electron microscopy for MGL-LPS nanodiscs structural studies	155
VII.1.1. MGL ECD chemo-selective labelling approach for negative-staining EM ...	156
VII.2. Perspectives for the MGL ECD project	163
VII.2.1. Mutants design for further MGL glycan secondary binding site investigation	163
VII.2.2. Design of shorter MGL ECD versions for structural studies	165
VII.2.3. Expression trials of designed MGL ECD constructs	167
VII.3. Conclusion	169
General conclusion and future perspectives	170
VIII. Conclusion	170
IX. Future perspectives	173
IX.1. SAXS/SANS for MGL-LPS nanodisc model construction	173
IX.2. Extension of nanodiscs technology to other bacterial strains with potential structural characterization approaches	175
IX.3. Super-resolution microscopy and AFM imaging	176
IX.4. Link the <i>in-vitro</i> interaction to the immune system modulation	178
Material and methods	180
References	187
Annexes	205

List of figures

Figure 1.1: Overall schematical organization of Gram-positive and Gram-negative bacteria cell envelopes.....	1
Figure 1.2: Schematic representation of the cellular pathways of outer membrane components biogenesis and transport.....	4
Figure 1.3: Schematic representation of LPS structure and core OS variability	5
Figure 1.4: The Raetz pathway for lipid A-Kdo2 biogenesis in <i>E. coli</i> K-12.....	8
Figure 1.5: Simplified representation of core oligosaccharide biogenesis in <i>E. coli</i>	9
Figure 1.6: The three O-antigen synthesis pathways	10
Figure 1.7: Transport pathway of LPS	11
Figure 1.8: Crystal structure of LPS transporter LptB ₂ FG from <i>P. aeruginosa</i> and its proposed model for LPS extraction	13
Figure 1.9: LptADE transport machinery	15
Figure 1.10: Covalent lipid A modifications in <i>E. coli</i> and <i>Salmonella</i>	17
Figure 2.1: Simplified TLR4–MD2 signalling pathways	19
Figure 2.2: Side view of the X-ray structure of TLR4-MD2-LPS complex (PDB: 3FXI).....	20
Figure 2.3: LPS extracellular recognition by TRP channels.....	22
Figure 2.4: LPS-mediated caspase non-canonical inflammasome activation.....	24
Figure 2.5: Domain organization of Macrophage Galactose-type Lectin (MGL) at the surface of antigen presenting cells	27
Figure 2.6: CRD domain overall organization.....	28
Figure 2.7: Illustration of the major signalling pathways used by CLRs upon pathogen recognition	30
Figure 2.8: Overview of some CLRs expressed on different antigen-presenting cells with an indication of some of their respective recognized bacterial species	31
Figure 2.9: Crystal structure of hMGL CRD domain	33
Figure 3.1: Classification of antibiotics on the basis of their targets.....	37
Figure 3.2: Overview of antibiotic resistance strategies in bacteria	38
Figure 3.3: Chemical structures of polymyxin B and colistin	46
Figure 4.1: Illustration of the cleavage sites of LPS deacylation reactions	50
Figure 4.2: Derivatization protocol used for LPS monosaccharide branching determination	51

Figure 4.3: Strategy used for the determination of absolute configuration of LPS monosaccharides	52
Figure 4.4: MALDI-MS spectra of intact and O-deacylated LOS of <i>Shewanella pacifica</i>	54
Figure 4.5: Summary of the described NMR approaches employed for core OS assignment	56
Figure 4.6: <i>P. aeruginosa</i> LPS O-antigen interaction with gentamicin	57
Figure 4.7: Interaction of the extracellular domain of DC-SIGN to mycobacterial glycans	59
Figure 4.8: Binding of enteropathogenic and enterohemorrhagic <i>E. coli</i> strains to the lectin array	61
Figure 4.9: Investigation of DC-SIGN binding to <i>E. coli</i> R1 core OS by flow cytometry and fluorescence microscopy	64
Figure 4.10: Schematic representation of the principle of STD NMR	66
Figure 4.11: Interaction of <i>E. coli</i> R1 LOS reconstituted in DDM micelles with surface-oriented DC-SIGN lectin	70
Figure 4.12: Chemical structure of the core OSs of the different <i>E. coli</i> strains used in this study	71
Figure 4.13: Styrene-maleic acid in membrane research	74
Figure 5.1: Amino acid sequences of MGL variants and the construct used in this study	79
Figure 5.2: Alignment of MGL ECD ^{wt} and MGL ECD ^{D245H} mutant sequences	80
Figure 5.3: Expression test of MGL ECD in M9 minimal media	82
Figure 5.4: Inclusion bodies refolding steps of [U- ² H, ¹⁵ N] MGL-ECD followed on a 12% SDS-PAGE	83
Figure 5.5: Purification of MGL ECD	84
Figure 5.6: Mass spectrometry analysis of MGL ECD ^{H262A}	85
Figure 5.7: Superposition of ¹⁵ N- ¹ H correlation spectra of MGL-ECD and MGL-CRD	87
Figure 5.8: Chemical structure of <i>E. coli</i> R1 and R3 core oligosaccharides used in this study	88
Figure 5.9: NMR spectroscopy interaction studies of MGL ECD with OS R1 and R3	89
Figure 5.10: SPR direct interaction of R1 and R3 LOS reconstituted in DDM micelle with MGL ECD oriented surface	90
Figure 5.11: SPR interactions of R1 and R3 LOS DDM micelle with DC SIGN ECD oriented surface	91
Figure 5.12: Flow cytometry quantification of MGL ECD binding onto <i>E. coli</i> R1 and <i>E. coli</i> R3 cells	92

Figure 5.13: Flow cytometry quantification of MGL ECD labelled with AF647 bound to R1 and R3 cells.....	93
Figure 5.14: LPSs chemical structures of <i>E. coli</i> strains O113:H21 and O157:H7 LPSs used in this study	94
Figure 5.15: Flow cytometry monitoring of the influence of O-antigen portion on MGL ECD binding onto <i>E. coli</i> cells	95
Figure 5.16: Combined SAXS-AlphaFold model of MGL-ECD	121
Figure 6.1: Overview of Gram-negative bacteria cell envelope	124
Figure 6.2: Preparation and characterization of SMA nanodiscs.....	126
Figure 6.3: Representative images of different Gram-negative membranes separation sucrose density gradients	128
Figure 6.4: outer membrane nanodiscs characterization using DLS and negative staining EM	129
Figure 6.5: AFM characterization of SMA nanodiscs	131
Figure 6.6: NMR detection of different LPS regions	132
Figure 6.7: Solution NMR 1D ¹ H spectrum of SMA O157 _{pur}	133
Figure 6.8: Solution NMR ¹ H- ¹³ C correlation spectra of the O-antigen carrying nanodiscs	134
Figure 6.9: Solution NMR ¹ H- ¹³ C correlation spectra of SMA R1 _{pur} nanodiscs.....	135
Figure 6.10: Spectra quality monitoring following temperature variation	136
Figure 6.11: Monitoring LPS nanodiscs phase transition temperature using solid state NMR spectroscopy.....	138
Figure 6.12: Solid state NMR ¹ H- ¹³ C hCH CP correlation spectrum of SMA R1 _{pur} nanodiscs	139
Figure 6.13: Solid state NMR characterization of LPS nanodiscs from <i>E. coli</i> O157:H7	140
Figure 6.14: Solid state NMR characterization of SMA O157:H7 nanodiscs	141
Figure 6.15: ssNMR spectroscopy dynamic analysis of SMA R3 _{OM}	143
Figure 6.16: Interactions studies of immobilized SMA R1 _{pur} with PmB by BLI	145
Figure 6.17: BLI interaction studies of immobilized SMA R1 _{pur} with MGL ECD	146
Figure 6.18: SMA R1 _{pur} nanodiscs and MGL ECD interactions monitoring by QCM-D and AFM.....	148
Figure 6.19: SMA R1 _{pur} nanodiscs and PmB interactions monitoring by QCM-D.....	150
Figure 7.1: Schematic representation of MGL ECD on bacterial membranes	154
Figure 7.2: Negative staining Electron Microscopy characterization.....	155

Figure 7.3: Summary of the different reactions carried out for MGL ECD chemo-selective labelling.....	158
Figure 7.4: Mass spectrometry analysis of MGL ECD stock solution before Strep tag cleavage	159
Figure 7.5: Mass spectrometry analysis of MGL ECD after factor Xa cleavage	160
Figure 7.6: Purification of MGL ECD following SrtA reaction	161
Figure 7.7: Mass spectrometry analysis of MGL ECD Following SrtA ligation reaction.....	163
Figure 7.8: Docking and molecular dynamics of R1 OS outer core binding onto MGL new OS binding site.....	164
Figure 7.9: MGL ECD and DC-SIGN ECD models.....	165
Figure 7.10: Design of shorter MGL ECD constructions of different coiled-coil lengths	166
Figure 7.11: Expression test of MGL ECD constructs.....	167
Figure 7.12: Expression test of MGL ECD constructs	168
Figure 8.1: MGL-LPS nanodiscs model construction	174
Figure 8.2: Super-resolution microscopy and AFM imaging for MGL-LPS nanoscale characterization	177
Figure 8.3: Immunological assessment of MGL-LPS interactions.....	179

List of tables

Table 1: Summary of LPS biogenesis inhibitors that have entered pre-clinical development	43
Table 2: Summary of LPS transport inhibitors in pre-clinical development	45
Table 3: minimal media composition.....	81
Table 4: List of the expected on observed masses following Factor Xa cleavage reaction...	160
Table 5: Expected and observed masses following SrtA labelling reaction of 3G-MGL ECD.	162
Table 6: Summary of the different strains, culture conditions used for MGL constructs expression tests.	168

List of Abbreviations

AF647	Alexa Fluor 647
AFM	Atomic Force Microscopy
APC	Antigen Presenting Cells
BLI	BioLayer Interferometry
BSA	Bovine Serum Albumin
CLR	C-type Lectin Receptor
CP	Cross-Polarization
CRD	Carbohydrate Recognition Domain
CSP	Chemical Shift Perturbation
DBCO	DiBenzoCycloOctyne
DC-SIGN	Dendritic Cell-Specific Intracellular adhesion molecule-Grabbing Non-integrin
DDM	n-Dodecyl β -D-maltoside
DLS	Dynamic Light Scattering
ECD	ExtraCellular Domain
EM	Electron Microscopy
FSC	Forward Scatter
GalNAc	N-acetylgalactosamine
GlcNAc	N-acetyl glucosamine
Heptose	L-glycero-D-manno-heptose
hMGL	human MGL
HSQC	Heteronuclear Single-Quantum Coherence
IM	Inner Membrane
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
IPTG	IsoPropyl- β -D-ThioGalactopyranoside
Kdo	3-deoxy-D-manno-oct-2-ulosonic acid
LOS	LipoOligoSaccharides
LPS	LipoPolySaccharide
MAS	Magic Angle Spinning
MGL	Macrophage Galactose-type Lectin
MIC	Minimum Inhibitory Concentration

NMR	Nuclear Magnetic Resonance
OM	Outer Membrane
OS	OligoSaccharide
PCP	Phenol/Chloroform/light Petroleum
PmB	Polymyxin B
QCM-D	Quartz Crystal Microbalance with Dissipation monitoring
REDOR	Rotational Echo Double Resonance
R-LPS	Rough LPS
SAXS	Small-Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulfate – PolyAcrylamide Gel Electrophoresis
S-LPS	Smooth LPS
SMA	Styrene-Maleic Acid
SMALPs	SMA Lipid Particles
SPR	Surface Plasmon Resonance
SrtA	Sortase A
SSC	Side Scatter
ssNMR	Solid-State NMR
STD	Saturation Transfer Difference
TOCSY	Total COrrelation SpectroscopY
TROSY	Transverse Relaxation Optimized SpectroscopY

INTRODUCTION

I. Bacterial cell envelope: Composition and architecture

I.1. Bacterial membranes: structure and properties

Biological membranes stand out as the principal interface between the cell and its surrounding, with key structural and functional roles. These membranes exhibit varied structures and architecture across organisms and cell types. Bacterial cell envelopes present a sophisticated multilayered complex acting as a protection barrier for bacteria from external threats. Back in 1884, Hans Christian Gram, a Danish bacteriologist, developed a staining technique to classify bacteria into two groups, based on differences in their cell envelope structure. The Gram-positive bacteria group that retained Gram's stain, and the Gram-negative bacteria group that didn't ¹.

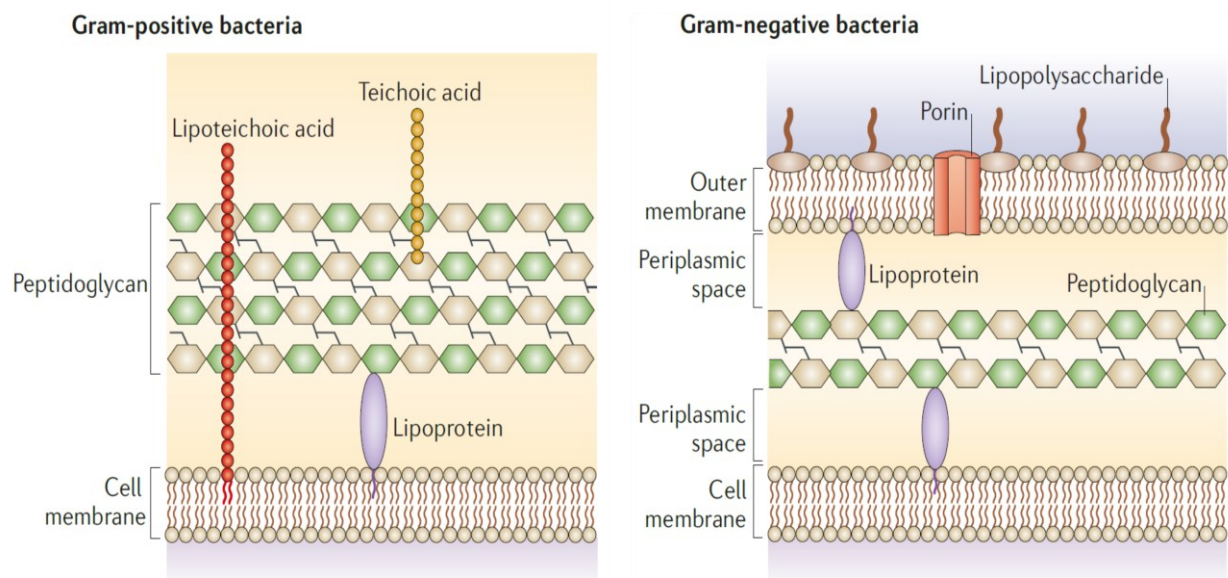


Figure 1.1: Overall schematical organization of Gram-positive and Gram-negative bacteria cell envelopes. Gram-positive bacteria have a single lipid membrane surrounded by a thick peptidoglycan cell wall. In contrast, Gram-negative bacteria cell wall comprises two distinct membranes separated by a thin peptidoglycan layer in the periplasmic space. Adapted from ²⁹².

The Gram-positive cell envelope consists of a cytoplasmic lipid bilayer surrounded by a thick peptidoglycan (PG) cell wall (Figure 1.1). The cell membrane is composed of phosphatidylethanolamine (PE) and phosphatidylinositol (PI) in its inner leaflet, phosphatidylglycerol (PGly) in the outer leaflet, and cardiolipin (CL) distributed in both leaflets². The PG layer with a width of up to 40-80 nm is significantly thicker in Gram-positive bacteria compared to the PG layer in Gram-negative bacteria (~7-8 nm thick)³. These PG layers usually contain long anionic polymers known as teichoic acids, mainly composed of glycerol phosphate, ribitol phosphate, or glucosyl phosphate, and constitute over 60% of Gram-positive cell wall mass¹. Teichoic acids are either covalently attached to PG, and form wall teichoic acids (WTA), or anchored to membrane lipids head groups, forming lipoteichoic acids (LTA)⁴. In contrast, Gram-negative bacteria contain two distinct membranes, the inner and outer membranes (IM and OM, respectively), separated by the periplasm containing a PG layer (Figure 1.1). IM is mostly composed of PE, PGly, and CL in both leaflets. Unlike IM, the OM is asymmetric in its lipid distribution, where the inner leaflet consists exclusively of phospholipids, whereas the outer leaflet is additionally made up of lipopolysaccharides (LPS)^{1,5}.

The PG layer in Gram-negative bacteria comprises repeating units of $\beta(1\rightarrow4)$ linked N-acetyl glucosamine (GlcNAc) and N-acetyl muramic acid (MurNAc) disaccharide. The lactoyl group of MurNAc is attached to oligopeptide stems, made of two to five amino acids, mostly five in Gram-negative bacteria. The peptide stem is often constituted of L-alanine (L-Ala), γ -D-glutamate (D-iGlu), meso-diaminopimelic acid (mDAP), and D-Ala–D-Ala. In Gram-positive bacteria, L-lysine (L-Lys) is often at the third position instead of mDAP amino acid⁶. Stem peptides may be cross-linked, either directly or *via* peptide bridges, by connecting the carboxyl group of D-Ala at position 4 and the amino group of mDAP (or L-Lys) at position 3⁷. However, some varieties either in the glycan strand or the peptide stem (length and/or cross-linkage percentage) exist for different bacterial strains, growth conditions, and external factors. Glycan strands can undergo some chemical modifications, including N-deacetylation and O-acetylation of both sugars and N-glycolylation of MurNAc. These variations were identified in both Gram-negative and Gram-positive bacteria, notably O-acetylation in *Staphylococcus aureus*, *Enterococcus hirae*, *Streptococcus pneumoniae*, *Neisseria gonorrhoeae*, and *Neisseria meningitidis*⁸. Similarly, variations at the peptide stem level have been reported where non-canonical amino acid modifications can occur during synthesis, more frequently at position 3 where in certain species other mono (L-homoserine, L-Ala, L-Glu) or diamino (meso-lanthionine, L-2,4-diaminobutyric acid, D-Lys) acids are added⁷. Finally, the most common

cross-link involves a D-Ala⁴ and mDAP³ residues of the donor and receptor stem, respectively. Other cross-linkages have been identified, typically mDAP³-mDAP³ in *E. coli*⁶, together with the D-iGlu²-mDAP³ cross-link in coryneform bacteria, especially the phytopathogenic corynebacteria⁷. It is believed that all of these peptidoglycan structural modifications contribute to bacterial virulence and alteration of host defence. That was reported for peptide stem amino acids modifications, in L-Ala, D-iGlu, or mDAP/L-Lys, which reduced their recognition by host antimicrobial receptors⁹. Similarly, glycan backbone deacetylation and O-acetylation in some pathogenic strains such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, respectively, are crucial for lysis evasion by host-lysozyme resistance¹⁰.

1.2. The outer membrane: a distinguishing feature of Gram-negative bacteria

Like the rest of biological membranes, the outer membrane (OM) is a lipid bilayer, in fact, an asymmetric one with phospholipids confined in the inner leaflet and a mixture of lipopolysaccharide (LPS) and phospholipids in the outer leaflet¹¹. Supporting evidence of the asymmetric arrangement of OM was presented where phospholipids of *Salmonella typhimurium* intact cells showed no susceptibility to hydrolysis by phospholipases, indicating their inaccessibility on the OM surface¹². In addition to lipids, numerous proteins reside in the OM: namely β -barrel proteins called outer membrane proteins (OMPs), and lipoproteins (Lpps). The latter are attached to the OM's inner leaflet through an N-terminal lipidic tail and are implicated in a variety of functionalities, including iron uptake, cellular adhesion, OM biogenesis and PG synthesis^{13,14}. A variety of topologies has been introduced for Lpps: they can adopt a self-exposed topology (multiple have been identified in *Borrelia burgdorferi* implicated in host evasion and cell adhesion including VlsE)¹⁵, form transmembrane channels in OM (CsgG and Wza lipoproteins of secretion channels in *Escherichia coli*)¹⁴, or exist in complexes with β -barrel proteins (LptE lipoprotein, required for LPS insertion in OM, forms a complex with its β -barrel partner LptD)¹⁴.

OMPs on the other hand, are transmembrane proteins with an antiparallel β -barrel conformation. Some OMPs function as passive diffusion porins of small molecules (e.g. OmpF, and OmpC), while others ensure specific nutrients import (e.g. LamB and PhoE)¹. β -barrels present a large variety in size, architecture, and oligomeric states. OmpX is the smallest OMP known to date while LptD is the largest one, with 8 and 26 β -strands, respectively^{16,17}. They can contain large extracellular loops with periplasmic domains (e.g. OmpA)¹⁸, or interact with

lipoproteins (e.g. OmpA which interacts with RcsF) ¹⁹. Finally, they exist as single monomer barrels such as for OmpA ¹⁸, or form trimers like OmpF ²⁰.

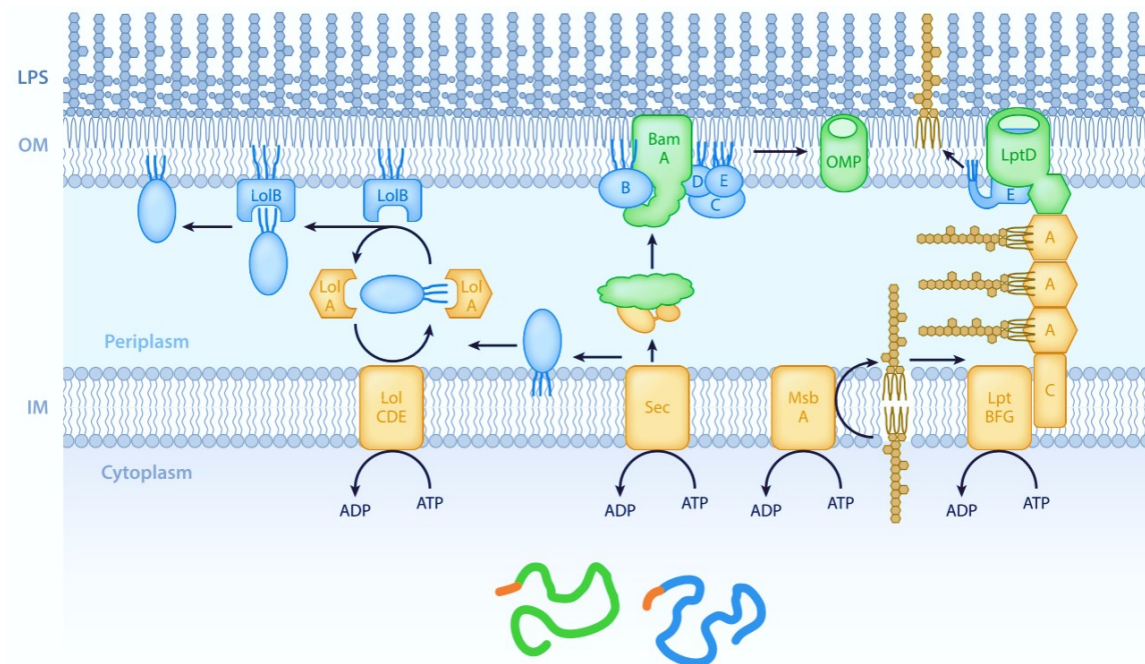


Figure 1.2: Schematic representation of the cellular pathways of outer membrane components biogenesis and transport. Outer membrane proteins, β -barrels (green) and lipoproteins (blue), are initially synthesized with a signal sequence (orange). The Sec complex translocates them across the IM, where lipoproteins are extracted from the IM by the LolCDE machinery. Both are then escorted by protein chaperones to OM receptors, LolB and Bam complex, where OMPs are assembled into the OM and lipoproteins inserted in the inner leaflet of OM. LPSs on the other hand are translocated across the IM by MsbA, and transported to the cell surface by the Lpt machinery. LPS: Lipopolysaccharide, OM: outer membrane, IM: inner membrane, OMPs: outer membrane proteins. Adapted from ¹³.

Biogenesis, transport, and insertion of each OM component are achieved by different types of machinery. All of them working in a very coordinated fashion with cell growth. Briefly, LPS synthesis is carried out at the cytoplasmic face of the inner membrane. Subsequently, the synthesized molecule is removed from the inner membrane, crosses the periplasm, and finally assembles at the cell surface. A process mediated by MsbA and the Lpt complex machinery in an ATP-dependent manner (Figure 1.2) ^{13,21}. Similarly, proteins destined for OM are synthesized in the cytoplasm with a signal sequence at their N-terminus allowing their translocation from the cytoplasm by the Sec complex. They then cross the periplasm bound to chaperones which deliver them to outer membrane receptor proteins, LolB and the β -barrel

assembly machine (BAM) for lipoproteins and OMPs, respectively, which ensure their insertion in OM and folding, in the case of OMPs (Figure 1.2) ¹³.

1.3. Lipopolysaccharides: structure and biogenesis

1.3.1. LPS chemical structure and nomenclature

LPSs, at the outer leaflet of Gram-negative bacterial membranes, are glycoconjugates constituted of three distinct parts: an acylated lipid A linked to an oligosaccharide chain (Core OS), which in turn is linked to a surface-exposed O-antigen polysaccharide region (O-PS) (Figure 1.3) ²¹. Incomplete synthesis of LPS can take place in Gram-negative bacteria and form rough phenotype bacteria where LPS, referred to as Rough LPS (R-LPS, or LOS for LipoOligoSaccharides), is lacking the O-PS region. More mutation can occur and form other truncated versions of LOS, ReLPS (also known as deep rough LPS constituted only of lipid A and Kdo), RdLPS (formed by lipid A and core OS inner core), and RcLPS (Lipid A-inner core OS and incomplete outer core OS) ²². On the other hand, in bacteria with a smooth phenotype, the O-PS is added to LOS and LPSs are then referred to as S-LPS (Figure 1.3).

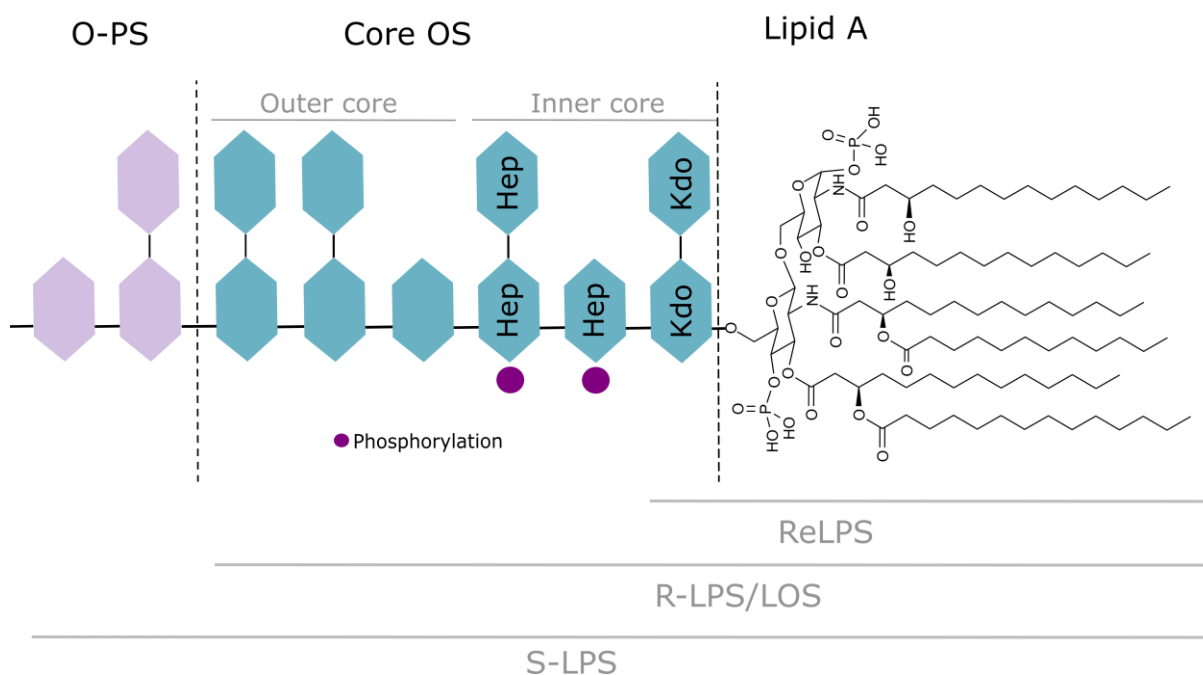


Figure 1.3: Schematic representation of LPS structure and core OS variability. Lipid A (black) anchors LPS in the membrane. Core OS (blue) subdivided in an inner and an outer core. O-PS repeating units (light purple). Hep: heptose, Kdo: 3-deoxy-D-manno-oct-2-ulosonic acid, blue and purple hexagons are hexoses of various nature.

The lipid A moiety is the most conserved part of LPS within species, and consists of a β -(1 \rightarrow 6) di-glucosamines (GlcN), phosphorylated on positions 1 and 4', and acylated with four

acyl chains through an amide bond at positions 2 and 2', and ester bonds at positions 3 and 3' (Figure 1.3)²³. Given their conserved structure, lipid A molecules act as potent activators of the innate immune system receptors, including toll-like receptor 4 (TLR4)²⁴. Chemical variations in lipid A primary structure have been reported to play crucial roles in LPS toxicity and antimicrobial resistance. The addition of polar groups to lipid A GlcN, for instance, a phosphoethanolamine (PEtN) or a 4-amino-4-deoxy-L-arabinose (L-Ara4N), neutralizes its negative charge and reduces bacterial susceptibility to cationic antimicrobial peptides²⁵. Furthermore, modifications in the acylation patterns dictate lipid A agonistic and antagonistic effects. LPSs with hexa-acylated lipid A (the case for most enteric bacteria like *E. coli*) have been found to be the most agonist one as the acylation degree correlates with cytokines induction. Whereas tetra-acylated lipid A (from *Yersinia Pestis*) and penta-acylated ones (from *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*) are found to be antagonistic in human cells²³.

The core OS is covalently linked to lipid A *via* a 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) linked to lipid A GlcN hydroxyl group at position 6'. OS region consists of a group of 10-15 sugars, and comprises an inner and an outer core. The former is composed of L-glycero-D-manno-heptose (heptose) and Kdo residues. Whereas the outer core is typically made up of non-charged hexoses, mainly three backbone residues and two side chain ones (Figure 1.3). More variability in composition, position, and linkage is displayed in the outer core region, distinguishing the five core OS types found in *E. coli*: R1, R2, R3, R4, and K-12²⁶. Core OS is responsible for pathologies associated with LOS-producing strains, notably *Haemophilus influenzae*. Furthermore, pathogens express LOS molecules presenting antigenic similarities with the human host (known as "antigenic mimicry"), to facilitate their evasion of the immune response. this was reported in *Neisseriae* and *H. influenzae* which add a sialic acid moiety, present on mammalian cell surfaces, to their sugar domain and escape the immune surveillance^{21,27}.

Attached to the outer core is the terminal part of LPS, the O-PS, a polymer of repeating sugar units, ranging from one to five thousand, of two to eight different monosaccharides (heteroglycans), or identical sugars in some bacteria (homoglycans)^{21,28}. Besides the frequent structures presented above, in S-LPS, some core OS can be substituted by other polysaccharide polymers, including capsular polysaccharides²⁹ and the enterobacterial common antigen (ECA)^{28,30}. Given the heterogeneities that can be found in the composition, position, length, linkages, and the presence of noncarbohydrate moieties in the oligosaccharide units within species²⁸, O-PS is the LPS's portion with the most variability, with more than 180 O-PS and

80 K-antigen structures in *E. coli* alone^{29,31}. O-PSs are implicated in bacterial interactions with host immune cells. It was even found that O-PS length dictates *Salmonella* virulence, where O-PS of 4 to 15 units were found to confer complement resistance, whereas longer O-PSs (16 to 35 repeating units) efficiently activated the complement³². Other than host cells, O-PSs are also known for their interaction with bacteriophages. These latter interact with O-PS through their tail spike proteins (TSPs) resulting in O-PS depolymerization and degradation as a first infection step before their addressing to a secondary receptor³³.

1.3.2. LPS Biogenesis and transport

1.3.2.1. Lipid A biogenesis

LPS biogenesis begins in the cytoplasm and at the cytoplasmic side of the inner membrane with serial enzymatic reactions to produce Kdo-lipid A moiety, the so-called “Raetz pathway” (Figure 1.4). The first reaction in lipid A synthesis consists of the acylation of the sugar nucleotide UDP-GlcNAc by the UDP-GlcNAc acyltransferase (LpxA). The acyl chain length may vary between species, but in *E. coli*, β -hydroxymyristate (14:0(3-OH)) is the preferred substrate of LpxA²³. Next, a deacylation step of the product, UDP-3-O-(acyl)-GlcNAc, by the zinc metalloenzyme LpxC takes place, followed by a second acylation, addition of a 14:0(3-OH) molecule to the free amino function, by LpxD to form UDP-2,3-diacylglucosamine. This product is then cleaved by the pyrophosphatase LpxH at its pyrophosphate bond to yield 2,3-diacylglucosamine-1-phosphate (lipid X). This latter is then condensed with another UDP-2,3-diacylglucosamine molecule, by LpxB, and forms a β -(1'→6) linked disaccharide-1-phosphate constituted of four 14:0(3-OH) acyl chains and a linked phosphate^{23,24}. LpxK, a specific kinase, phosphorylates the disaccharide at position 4' and forms lipid IV_A. Before completion of lipid A acylation, two Kdo residues are added by a bifunctional Kdo-transferase, WaaA also known as KdtA, which synthesizes the α -Kdo-(2→4)-Kdo disaccharide and transfers it to O-6' of lipid IV_A. Finally, the last step in lipid A synthesis consists of the introduction of two acyls, a lauroyl (12:0) by LpxL, and a myristoyl (14:0) by LpxM, to the linked C14:0(3-OH) residues at positions 2' and 3' of the distal glucosamine²³. All nine enzymes mentioned above are either soluble (LpxA, LpxC, and LpxD), peripheral (LpxB and LpxH), or integral membrane proteins (LpxK, LpxL, and WaaA)²³.

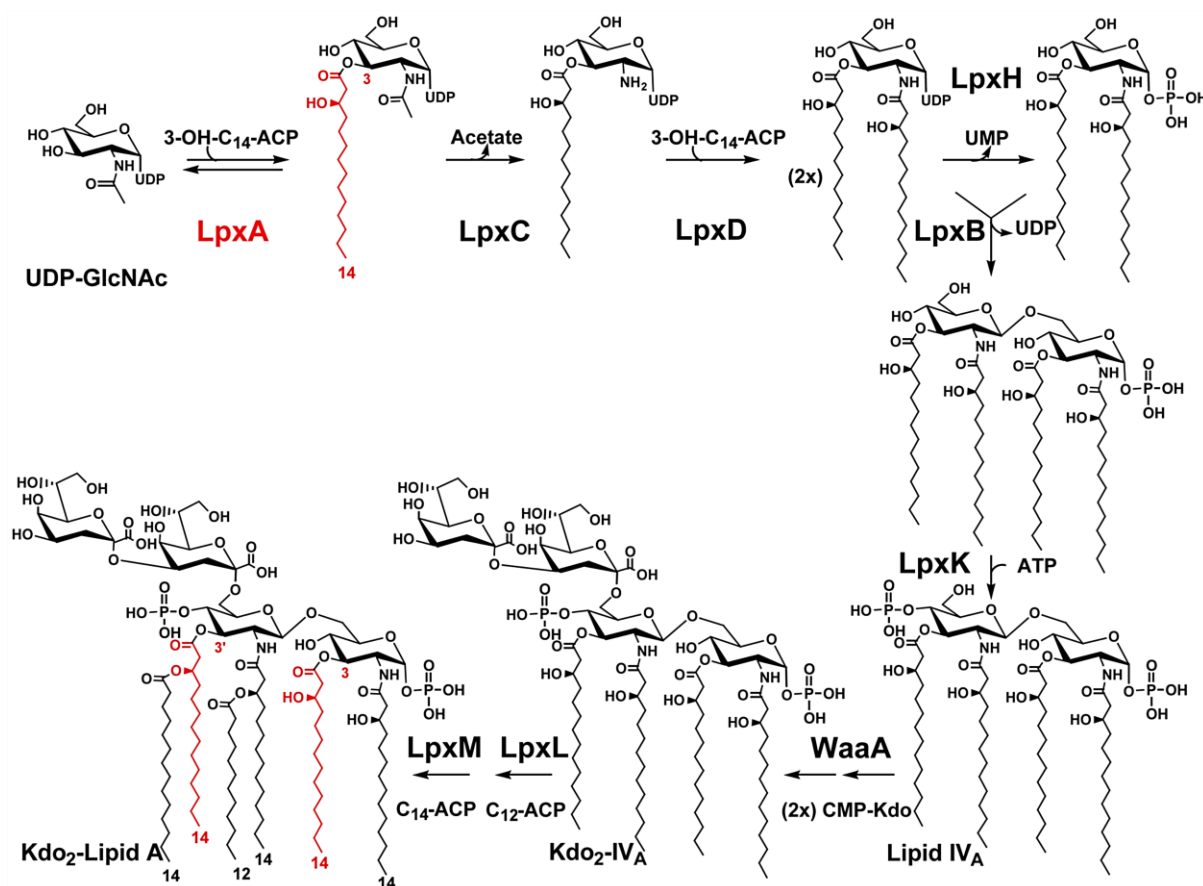


Figure 1.4: The Raetz pathway for lipid A-Kdo2 biogenesis in *E. coli* K-12. Raetz pathway with a highlight of the acyl chains added by LpxA in red as well as the nine relevant enzymes involved in each step of the synthesis. Adapted from ²⁴.

1.3.2.2. Core oligosaccharide and O-antigen biogenesis

At this stage, the core OS synthesis, also completed in the cytoplasmic side of the inner membrane, can begin. Both inner and outer cores are assembled in a stepwise fashion by a sequential addition of activated sugar nucleotides. That is carried out by a series of glycosyltransferases also known as Waa proteins that ensure both the glycosyl moieties transfer and their phosphorylation (Figure 1.5). The *waa* locus consists of three operons: *gmhD*, *waaQ*, and *waaA*. The *gmhD* operon codes for two heptosyltransferases WaaF and WaaC, whereas, *waaA* encodes for the Kdo-transferase responsible of Kdo2 synthesis and transfer to lipid IV_A. Finally, the *waaQ* operon codes for proteins relevant in the biosynthesis and modification of the outer core ²⁴.

Two heptose residues are added by two heptosyltransferases, WaaC and WaaF, followed by reactions catalysed by WaaP, WaaQ, and WaaY in order. WaaP catalyses the phosphorylation of the heptose added by WaaC, while WaaQ transfers the third heptose of the core which is then

phosphorylated by WaaY. On the other hand, the outer core synthesis starts with a transfer of a glucose to the heptose by WaaG using UDP-glucose as a donor substrate, onto which other hexoses are transferred by WaaO, WaaR, and finally the final hexose group that serves as a O-antigen receptor by WaaU, a heptose group in K12 core type from *E. coli*³⁴. Upon completion, rough-type LPS is translocated from the inner to the outer leaflet of the inner membrane by an ABC transporter, MsbA, in an ATP-dependent manner (Figure 1.7)³⁵. It has been reported that MsbA acts as a checkpoint in LPS biogenesis since it recognizes both the phosphoglucosamine headgroup and the correct acylation number on lipid A, preventing the transport of any LPS synthesis intermediates³⁶.

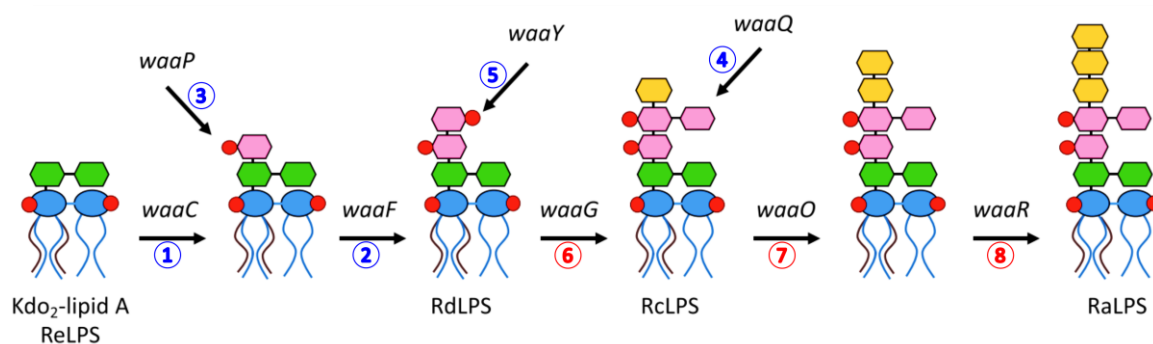


Figure 1.5: Simplified representation of core oligosaccharide biogenesis in *E. coli*. Core OS synthesis steps with a highlight of the genes involved in each step. The order of the enzymatic reactions is indicated with those corresponding to the inner and outer core synthesis in blue and red, respectively. Adapted from²².

In the periplasm, for smooth-type phenotype bacteria, an O-antigen moiety is added onto the core–lipid A by the O-antigen ligase WaaL to form S-LPS¹¹. The assembly of the repeating units begins separately from core-lipid A and takes place in the cytoplasm on a membrane-embedded lipid carrier known as undecaprenyl phosphate (Und-P). Unlike other glycosyltransferases, initiating enzymes recognize the hydrophobic Und-P instead of receptor sugars. Three different mechanisms for O-antigen synthesis and export across the inner membrane have been described: Wzy-dependent pathway, ABC transporter-dependent pathway, and synthase-dependent pathway (Figure 1.6)³⁷.

In the case of the Wzy-dependent pathway, the O-units are sequentially incorporated into the Und-P by WecA, WbbL, WbbJ, WbbK and WbbI. The repeating units are transferred across the inner membrane by the Wzx flippase. Wzy, an O-antigen polymerase combines the repeating units and yields long polysaccharide chains, which are transferred to core-lipid A by the WaaL ligase. For the ABC transporter-dependent pathway, the entire repeating unit is

assembled in the cytoplasm and translocated by an ATP-dependent transporter system (ABC) across the inner membrane where it is ligated by WaaL. Finally, in the synthase-dependent pathway, a synthase simultaneously polymerizes and translocates the O-antigen to the periplasmic side of the inner membrane where it is also ligated by WaaL (Figure 1.6) ^{21,37}.

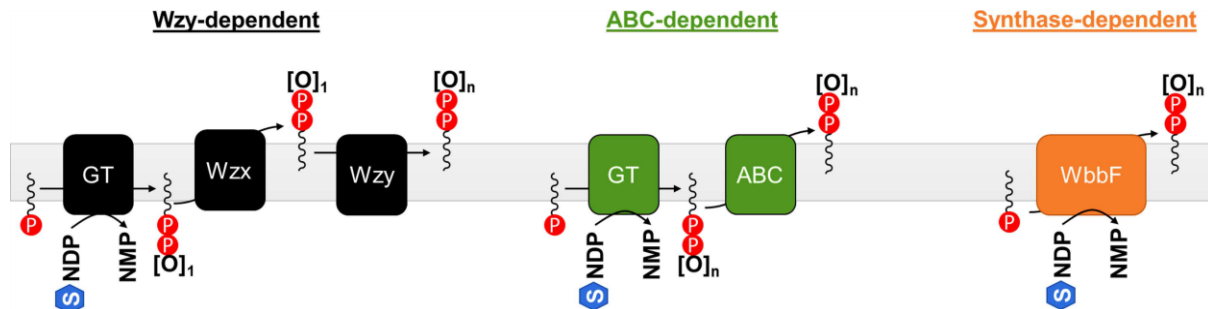


Figure 1.6: The three O-antigen synthesis pathways. Wzy-dependent, ABC-dependent, and synthase-dependent pathways. GT: glycosyltransferase, NDP: sugar nucleotide diphosphate, NMP: nucleotide monophosphate, [O]: O-antigen repeating units, n: number of units repetitions. The lipid carrier Und-P is shown as a black oscillating curve with a terminal P in red. Adapted from ³⁴.

1.3.2.3. LPS transport to the outer membrane

Fully synthesized LPSs are transported onto the surface of the outer membrane by a seven-protein complex machinery (LptA to LptG) (Figure 1.7). In *E. coli*, the Lpt complex is subdivided into an inner membrane complex, including LptB₂CFG, and an outer membrane complex of LptDE, both bridged by one or several copies of the periplasmic protein LptA ³⁵. The transport begins with LptB and LptFG, ATPase and transmembrane components, respectively, which intervene and extract LPS from the outer leaflet of the inner membrane. The LptB₂CFG complex forms a stable domain with LptC, a transmembrane helix and a jellyroll-like periplasmic domain that delivers LPS to a periplasmic transporter, LptA ³⁸. This latter facilitates LPS transport by forming an extended bridge spanning the periplasm. LptA transfers LPS to the final LPS-binding protein LptDE, forming a 1:1 stoichiometry plug and barrel structures in the outer membrane with LptE lipoprotein inside LptD lumen, for LPS insertion into the membrane. LPS transport is facilitated thanks to the adapted β-jellyroll fold adopted

by the different Lpt proteins acting as protecting shields to LPS acyl chains across the aqueous periplasm^{35,39}.

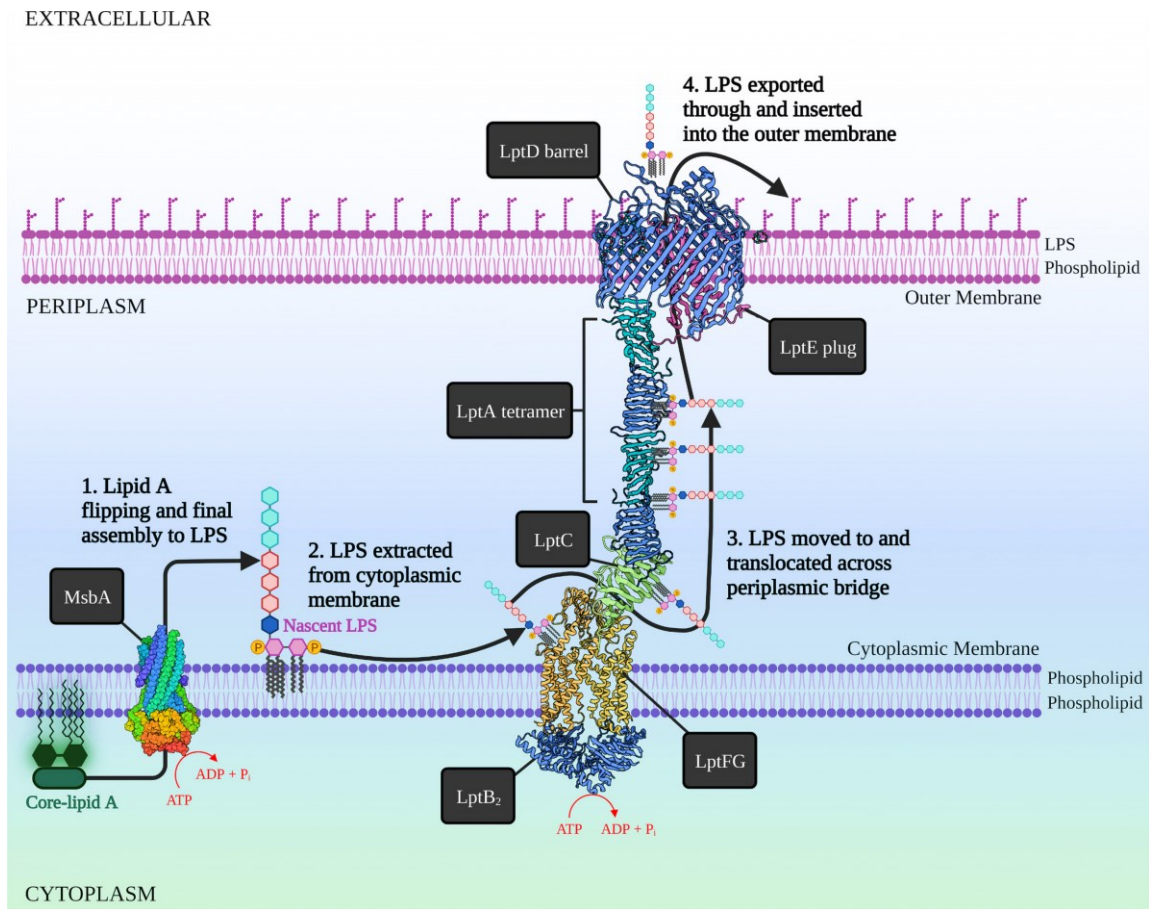


Figure 1.7: Transport pathway of LPS. Following the core oligosaccharide-lipid A synthesis, R-LPS is flipped across the inner membrane by MsbA. LPS biosynthesis is completed by ligation of O-antigen units by WaaL ligase in the periplasmic side of the inner membrane. Finally, LPS is transported to the outer membrane by the Lpt machinery. Adapted from¹⁴⁴.

The breakthroughs in Lpt structures contributed much to our understanding of LPS extraction and journey across the periplasm and support the described “PEZ” model in reference to the spring-filled candy dispenser. Following this comparison, the LptB₂FG complex represents the spring at the bottom of the dispenser which propels LPS, the candy, through the tube (in this case representing LptC and LptA) to the cap (LptD and LptE)⁴⁰. The LptB₂FG complex resembles other ABC transporter’s configurations. However, unlike ABC transporters, which translocate their substrates from one side of the membrane to another, LptB₂FG extracts it from the same leaflet. In LptB₂FG, the LptB₂ homodimer, with its nucleotide-binding domains (NBD), is coupled to the LptF/G transmembrane complex. LptF together with LptG consist of a transmembrane domain (TMD) composed of six transmembrane (TM) helices, a β -jellyroll periplasmic domain, and coupling helices that interact with LptB₂ on the cytoplasmic side (Figure 1.8). Within their TMDs, limited interactions between their TMs occur and form a V-shaped cavity in the inner membrane, which can be opened further upon conformational changes triggered by ATP binding and hydrolysis by LptB₂⁴¹. Based on the observed structural features, a model for LPS extraction has been proposed, where the complex cycles three main conformational states (Figure 1.8)⁴¹:

- 1- A resting state where no nucleotide is bound
- 2- LPS-loaded state: an ATP-bound state where LptB units get closer, triggering conformational changes in LptF/G TMDs inducing the cavity opening and LPSs’ lipid A entry.
- 3- LPS-extracted state: an ATP-hydrolysed state where lipid A is extracted out of the inner membrane and expelled into LptF/G periplasmic domains. Upon ADP release, LptB₂FG returns to the resting mode.

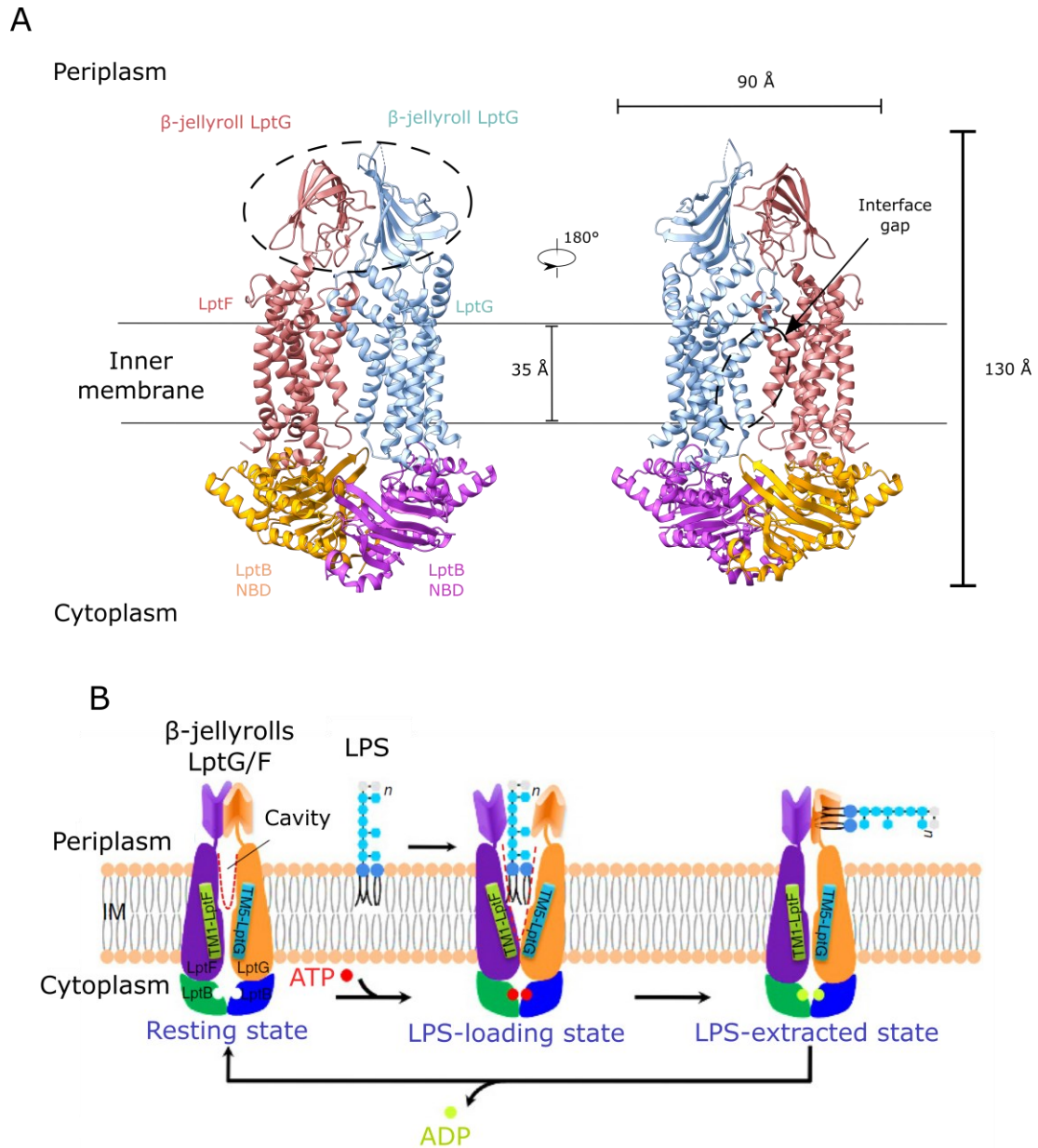


Figure 1.8: Crystal structure of LPS transporter $LptB_2FG$ from *P. aeruginosa* and its proposed model for LPS extraction. (A) Structure of the $LptB_2FG$ transporter (PDB: 5X5Y) shown in cartoon (left) and its corresponding 180° rotation view (right). The β -jellyroll domains and the interface gaps on the surface of $LptB_2FG$ are circled. (B) The $LptB_2FG$ complex undergoes different conformational states for LPS extraction: a resting state (a nucleotide-free state), an LPS-loaded state (ATP-bound state), and an LPS-extracted state (ATP hydrolysed state). Adapted from ⁴¹.

LptB₂FG forms a stable complex with LptC. The latter protein comprises an N-terminal TM helix (TM_C) and a periplasmic β-jellyroll domain. Single-particle cryo-EM was used to further characterize this transport mechanism by investigating LptB₂FG and LptB₂FGC structures and proposed another model for LPS extraction by LptB₂FGC⁴². LptC's TM helix was found to insert within the LptF/G TM helices, assumed to mediate electrostatic interactions with LPS in LptB₂FG cavity, and widely opens the complex's cavity. It is thus suggested that this insertion induces conformational changes that weaken LPS binding in LptB₂FG and increases its flexibility in LptB₂FGC⁴². Upon TM_C dissociation and LptC/F β-jellyroll domains association, a highly positively charged surface around the bound LPS is created. Following the ATP hydrolysis by LptB₂, LPS is then expelled from the TMDs towards the periplasmic β-jellyroll domain of LptF then LptC⁴².

LptA, with its 16 antiparallel β-strands adopts a β-jellyroll configuration, a fold that is shared with the other Lpt periplasmic domains described above. LptC C-terminus interacts with the N-terminus of LptA, and the C terminus of LptA interacts with the N-terminal periplasmic domain of LptD with no interaction with LptE^{41,43,44}. Crystal structures have shown that LptA forms oligomers with each of the LptA molecules interacting with an adjacent LptA molecule in a head-to-tail manner⁴⁵. This stacking creates a continuous hydrophobic β-jellyroll bridge for LPS transport across the periplasm with lipid A protection towards the β-jellyroll domain of LptD (Figure 1.9). LptD harbours a massive C-terminal transmembrane portion of 26 antiparallel β-strands spanning 50 Å with the LptE lipoprotein inside the lumen, and a β-jellyroll N-terminal extending from the periplasm to the outer membrane, a 2 β-sheets arranged in a V-shape, both joined by short turns and loops on the periplasmic and extracellular sides, respectively (Figure 1.9)⁴⁶. This configuration suggests that lipid A inserts into the membrane directly by the β-jellyroll fold, whereas the hydrophilic polysaccharide first gets in the β-barrel's lumen⁴⁰. A lateral gate in LptD β-barrel is formed between β1 and β26 strands which is facilitated by reduced hydrogen bonding between them in LptD. This allows lipid A diffusion and direct entry into the membrane, while sugar moieties transit to the β-barrel lumen before their translocation to the outer leaflet of the outer membrane following the extracellular loop 4 opening (Figure 1.9)^{46,47}. Unlike other Lpt proteins, LptE doesn't act as a direct transporter but is found to be involved in functional LptD assembly, and membrane permeability maintenance. LptE also facilitates LPS transfer and prevents its aggregation at the inner leaflet of the outer membrane⁴⁰.

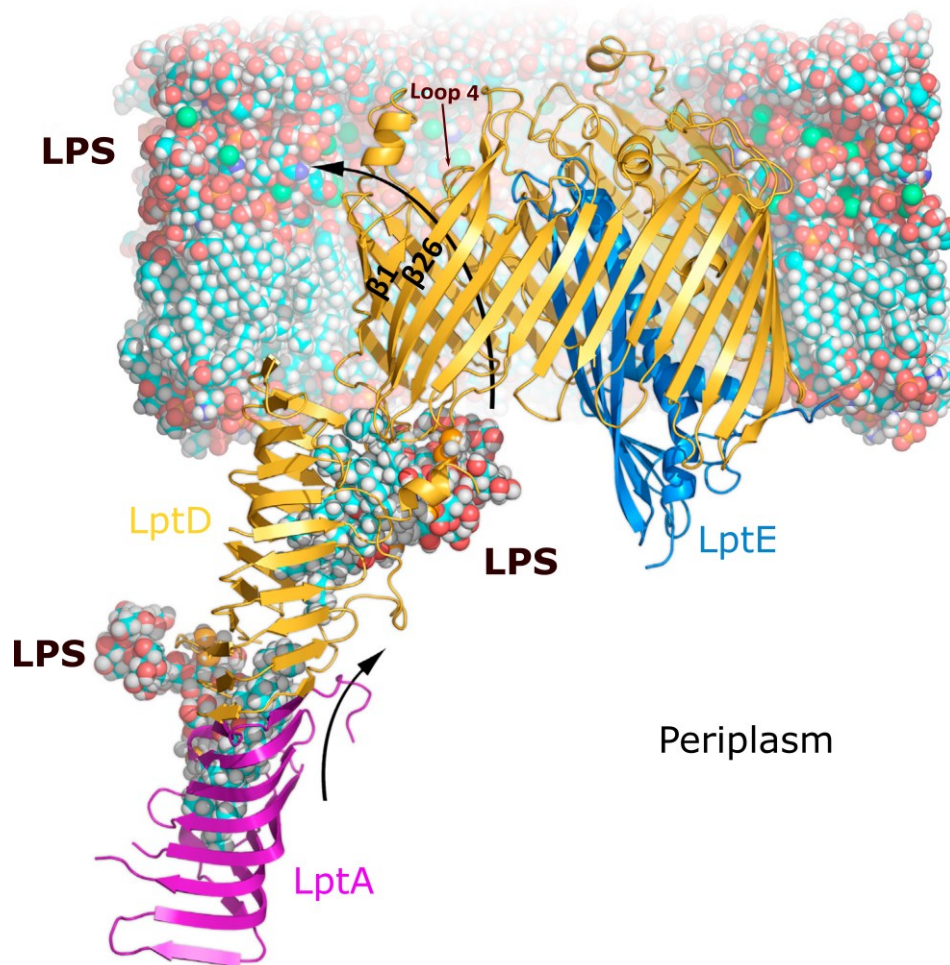


Figure 1.9: LptADE transport machinery. LptA (purple) delivers LPS to the β -jellyroll N-terminal domain of LptD (orange). A lateral gate opens between $\beta 1$ and $\beta 26$ strands allowing Lipid A to insert directly into the membrane, and the sugar moiety to get to the lumen of LptD. Following extracellular loop 4 movement and opening, the polysaccharide is translocated to the outer leaflet of the outer membrane. Calcium ions are represented with green sphere. Proteins and LPS are represented as cartoons and spheres, respectively. Adapted from ⁴⁶.

1.4. LPS Leaflet: an effective permeability barrier

1.4.1. Strong lateral interactions between LPS molecules

LPS leaflet establishes a permeability barrier that prevents the entry of toxic compounds into the cell. This characteristic is attributed to LPSs strong amphiphilic nature. Lipid A with its acyl moieties confers LPS a hydrophobic property that prevents the entry of hydrophilic compounds. In contrast, the polysaccharide regions (core OS and O-antigen) provide a hydrophilic character and prevent entry of hydrophobic molecules ³⁴. Production of such a barrier requires interactions between LPS molecules with a dense packing within the bilayer.

That is mediated by the presence of several saturated acyl chains per LPS molecule (six to seven in *E. coli* and *Salmonella*) which can lead to strong lateral interactions creating a gel-like lipid interior, and result in a decrease in the membrane fluidity⁴⁸. In addition to hydrophobicity-driven interactions, the sugar head groups and lipid A can contribute to LPS interactions through hydrogen bonding. The 3-OH-myristic acids at positions 2 and 3 harbour two hydroxyl groups within the membrane. Together with the 4-OH group of the reducing glucosamine residue, they can act as H-bond donors. Additionally, the polysaccharide region carries different groups that could act as H-bond donors and acceptors. The interaction between the different LPS molecules is further structured by coordination to divalent cations (Mg^{2+} and Ca^{2+}). These metal-ligand bonds neutralize the high negative charges residing within LPS, including two monophosphates in lipid A at positions 1 and 4', carboxyl groups of KDO residues, and phosphate substituents on the heptoses. These bonds also greatly facilitate neighbouring of LPS molecules to be cross-linked, and consequently contribute to membrane integrity and permeability^{39,48}. It was demonstrated that treatment of cells with a divalent cations chelator, EDTA, results in outer membrane destabilization and LPS release from the outer membrane⁴⁹.

1.4.2. Modification of LPS Structure

Bacteria adapt their LPS structure in response to environmental stress or external stimuli. *E. coli* adjusts its membrane fluidity at low temperatures through the expression of LpxP, an acyltransferase that incorporates a C16:1 palmitoleic acid instead of a lauroyl (12:0) acyl chain normally added by LpxL (Figure 1.10)²⁵. Lipid A modifications are known to take place in response to the level of divalent cations and the presence of cationic antimicrobial peptides (CAMPs). A sensing system of these changes has been described in *Salmonella* and *E. coli* known as the PhoQP regulatory system⁴⁸. In response to the sensing of these factors, the PhoQ kinase is activated and catalyses the phosphorylation and activation of the PhoP regulator. Once phosphorylated, PhoP induces the activation of the expression of genes encoding LPS-modifying enzymes^{34,48}. Some of these enzymes result in lipid A acyl chain modification, while others control changes in the glucosamine disaccharide moiety of LPS. The latter often occurs following the PmrAB system activation by PhoQP in response to CAMP sensing. Upon phosphorylation, PmrAB codes for enzymes needed for glucosamine modification, namely ArnT and EptA. ArnT mediates the transfer of 4-amino-4-deoxy-L-arabinose (L-Ara4N) to lipid A at position 4'-phosphate, whereas EptA transfers a phosphoethanolamine (PEtN) to lipid A at position 1'-phosphate or 1'-phosphate and 4'-phosphate, in the absence of L-Ara4N or under certain growth conditions (Figure 1.10)²⁵. The

addition of these positively charged moieties reduces the negative charge of lipid A recognized by CAMPs.

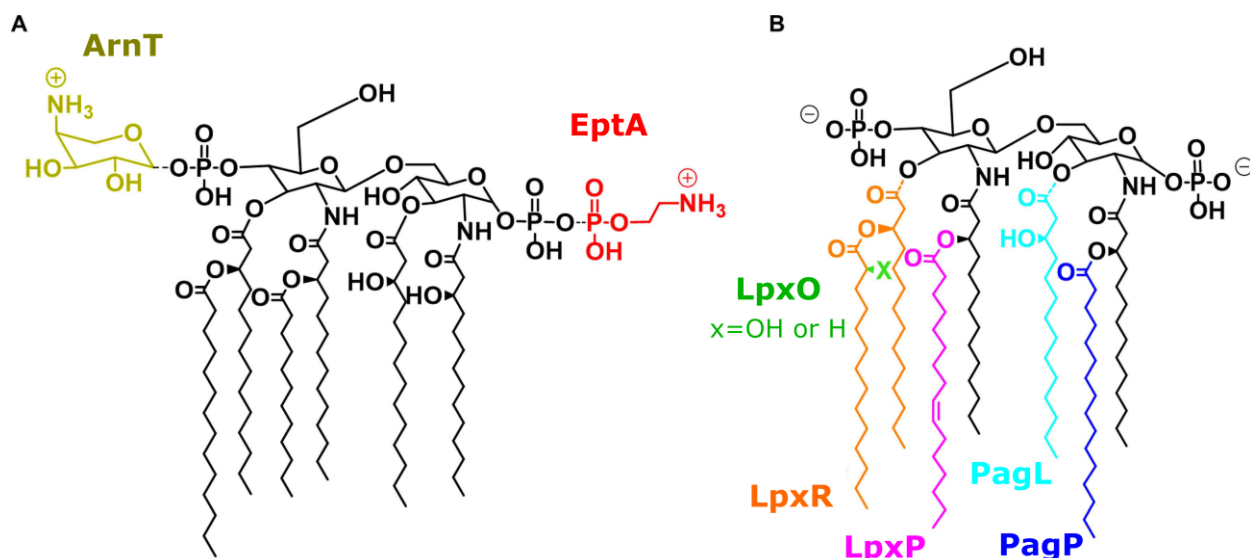


Figure 1.10: Covalent lipid A modifications in *E. coli* and *Salmonella*. The known lipid A modifications occurring at the glucosamine phosphates (A) or at the acyl chains (B) are indicated. Involved enzymes are indicated alongside their corresponding substitution. *ArnT* and *EptA* are under the control of *PmrAB*. *PagP* and *PagL* are regulated by *PhoQP*. *LpxO* and *LpxR* are not regulated by either *PhoP/Q* or *PmrAB*. *LpxP*, which replaces *LpxL* at low temperatures, is part of the conserved Raetz pathway. Adapted from ^{25,34}.

In addition to glucosamine modification, acyl chains of lipid A are not spared. *PagP*, an acyl transferase outer membrane, can also be activated by the sensing system. This acyltransferase transfers a palmitate (C16:0) to the hydroxyl group of 3-hydroxy-myristic acid residue linked to position 2 of lipid A (Figure 1.10), forming heptacylated LPS. The increase in lipid A acylation is thought to increase the LPS leaflet stability by increasing the hydrophobic interactions between LPS molecules, resulting in reducing the outer membrane's fluidity thus, reducing CAMP susceptibility ⁵⁰. Other bacterial strains than *E. coli*, namely *S. typhimurium*, carry other lipid A-modifying enzymes. *PagL*, an outer membrane lipase, mediates the removal of the 3-hydroxymyristoyl chain at position 3 of the lipid A moiety, while *LpxR*, also an outer membrane lipase, cleaves the 3'-acyloxyacyl groups at position 3' (Figure 1.10). These modifications are likely implicated in pathogenesis through cytokine response modulation ^{51,52}. Finally, an S-2-hydroxymyristate-modified lipid A at position 3' can be produced in a *PhoQP*-dependent manner. This is carried out by an oxygenase *LpxO* which converts the myristic acid into a 2-hydroxymyristate (Figure 1.10). This modification contributes to LPS leaflet stability by making up for the OH groups removal by *PagP* and maintaining the number of H-bond donors ^{48,53}.

II. Immunogenicity of LPS

II.1. LPS detection in mammals

II.1.1. TLR4-dependent detection of LPS

The host's immune system senses a plethora of pathogens through pattern recognition receptors (PRRs). PRRs detect microbe-specific molecular signatures known as pathogen-associated molecular patterns (PAMPs)⁵⁴ and activate downstream signalling pathways leading to the modulation of innate immune responses and the production of inflammatory cytokines, type I interferon (IFN), as well as other mediators⁵⁴. Toll-Like Receptors (TLRs), as one of many PRRs of the mammalian innate immune system, are documented as PRRs for bacterial, viral and parasitic molecular patterns⁵⁵. The TLR family counts 10 members in humans (TLR1 to TLR10) and 12 in mice (TLR1-TLR9 and TLR11-TLR13), distributed either on the cell surface or in intracellular compartments of innate immune cells, including dendritic cells and macrophages, or non-immune cells. Intracellular TLRs recognize bacteria and viruses' nucleic acids released after degradation (as is the case for TLR3, TLR7, TLR8 and TLR9), whereas cell surface TLRs recognize microbial membrane agents (as is the case for TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10)^{54,56}. All TLRs share similar structural features: an N-terminal extracellular binding domain consisting of leucine-rich repeats (LRR/EBD), a transmembrane domain, and a C-terminal intracellular Toll/Interleukin-1 receptor (TIR) domain that induces downstream signalling cascades⁵⁷.

TLR4 is found to recognize bacterial LPS through its lipid A moiety. This detection starts with a soluble LPS-binding protein (LBP) that senses released LPS in the serum and avidly binds to them through its N-terminal domain ($K_D = 10^{-9}$ M)⁵⁸. LPS aggregates are then transferred to the Cluster of Differentiation 14 (CD14) protein, a glycosylphosphatidylinositol (GPI)-anchored membrane protein, following the interaction of LBP C-terminal end with CD14⁵⁸. LPS in complex with CD14 is delivered to the TLR4-MD2 complex resulting in TLR4-MD2-LPS homodimerization with another TLR4-MD2-LPS complex, forming the M-shaped TLR4-MD2-LPS homodimer, which in turn leads to dimerization of the intracellular TIR domains. Subsequently, signal transduction takes place following the recognition of the TIR domains by cytoplasmic adaptor proteins. Two major pathways have been described, according to the involved adaptor protein: (i) the MYD88-dependent pathway involving a myeloid differentiation primary response protein 88 (MYD88), and (ii) a MYD88-independent pathway consisting of a TIR-domain-containing adapter inducing interferon- β (TRIF). The MYD88 pathway occurs at the cell surface and induces the production of pro-inflammatory

cytokines. This includes interleukin-6 (IL-6), IL-12, and tumour necrosis factor (TNF) ⁵⁹. In contrast, the MYD88-independent pathway takes place within the endosomes after endocytosis of CD14 and dimerized TLR4, and is characterized by the production of interferon- β (IFN β) and IFN-inducible proteins (Figure 2.1) ^{21,59}.

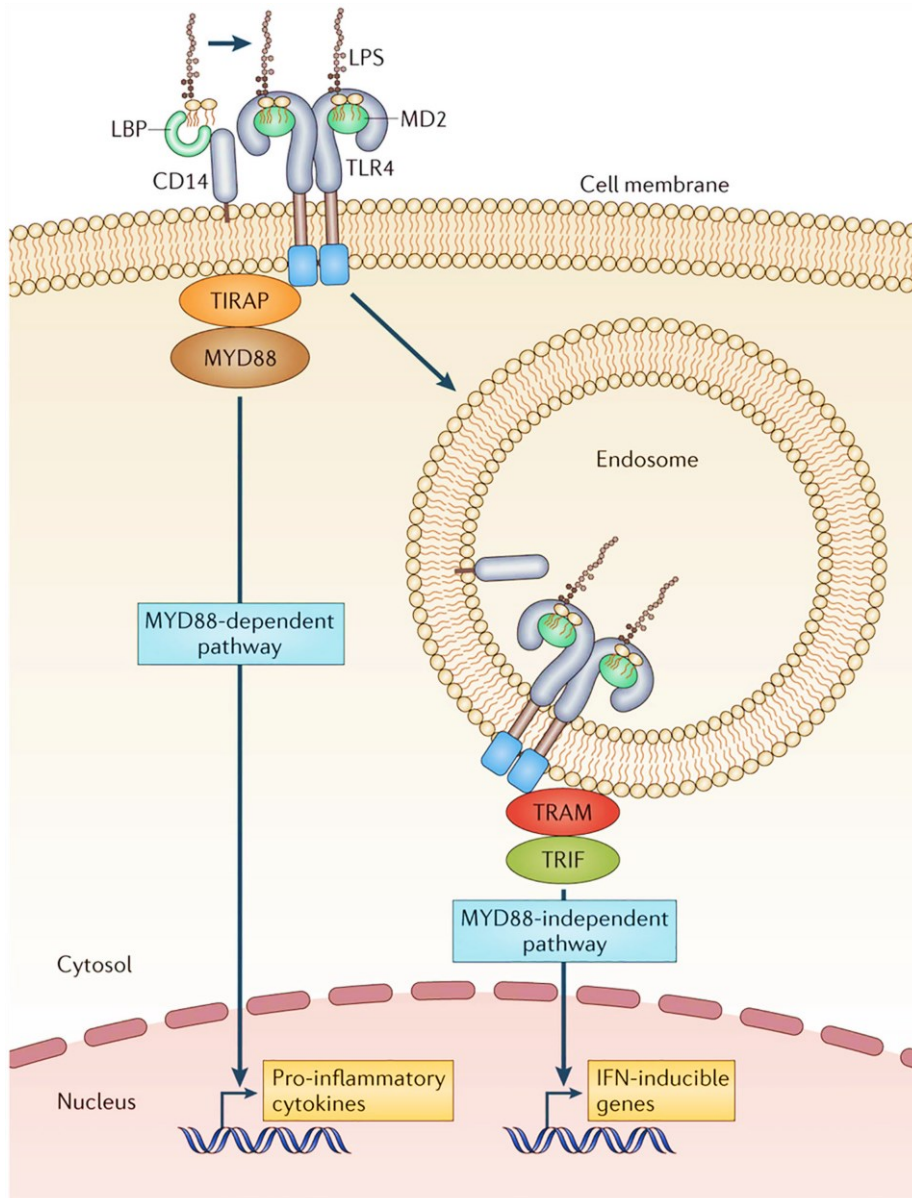


Figure 2.1: Simplified TLR4–MD2 signalling pathways. LPS is recognized by LBP and CD14, which delivers it to the TLR4-MD2 complex. Upon Lipid A recognition, two responses can take place: the myeloid differentiation primary response protein 88 (MYD88)-dependent response, which results in the production of pro-inflammatory cytokines, and the MYD88-independent pathway which occurs in endosomes and induces the expression of interferon induced genes. Adapted from ⁵⁹.

Crystal structures of TLR4/MD2-LPS from both humans and mice revealed a conserved recognition mechanism for *E. coli* hexa-acylated lipid A⁶⁰. TLR4 and MD2 complex is formed prior to LPS recognition where MD2 binds to the N-terminal and central regions of TLR4 extracellular domain. Upon LPS binding, five of the six acyl chains of lipid A are completely embedded in the hydrophobic pocket of MD2. In contrast, the sixth chain is exposed on the MD2 surface and mediates hydrophobic interactions with conserved phenylalanines of the other TLR4 (TLR4'). Furthermore, the 1-phosphate group is located near the dimerization interface of TLR4/MD2 and forms ionic interactions with positively charged residues from TLR4, TLR4', and MD2. Finally, MD2 loops F126 and L87 interact with TLR4' molecule. All of these interactions promote the homodimerization of two TLR4-MD2-LPS complexes and the formation of the active M-shaped complex (Figure 2.2)^{57,61}.

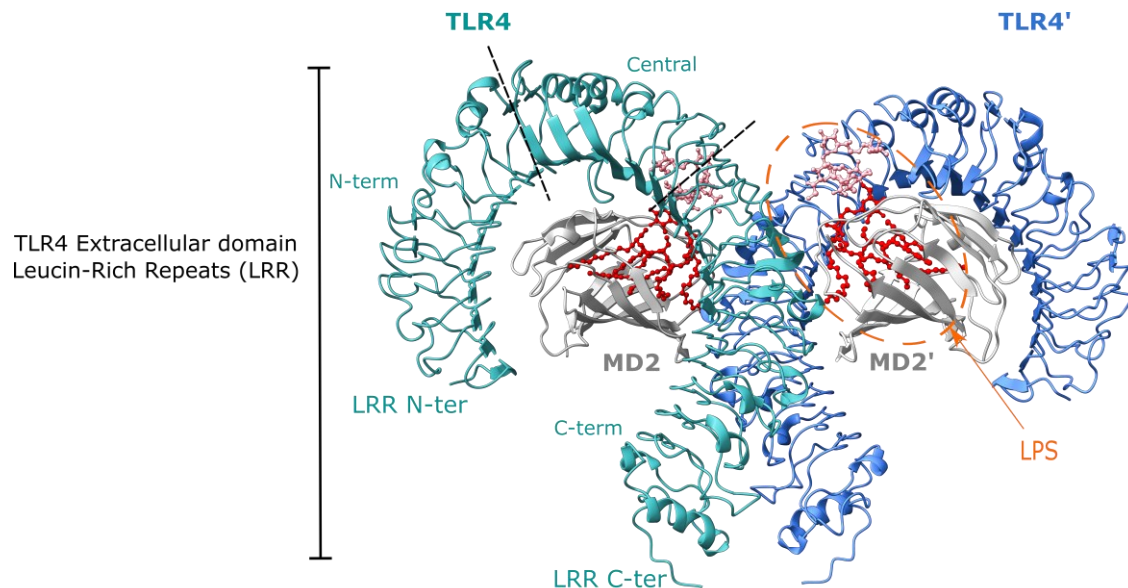


Figure 2.2: Side view of the X-ray structure of TLR4-MD2-LPS complex (PDB: 3FXI). TLR4 extracellular domain (in green and blue for TLR4 and TLR4', respectively) is divided into N-terminal, central, and C-terminal regions. MD2 is represented in grey, lipid A of LPS (ball and stick) is coloured in red, and the inner core carbohydrates of LPS are coloured in pink. Adapted from⁶¹.

The number of lipid A acyl chains was found to govern the immunological activity of LPS. In contrast to hexa-acylated lipid A, which acts as an agonist for mammalian cells, the tetra-acylated lipid IVa presented species-specific agonistic or antagonistic activities^{60,61}. In human MD2 complex with lipid IVa, the ligand glucosamine backbone adopts an inverted orientation with the glucosamine 4'-phosphate facing the dimerization interface. This results in a deeper arrangement of acyl chains within the binding pocket and prevents dimerization with another TLR4-MD2-LPS complex. In contrast, lipid IVa occupied a similar

conformational space as the hexa-acylated lipid A on mouse MD2. These observations thus provide the structural basis behind both the agonistic effect of lipid IVa with mouse TLR4/MD2 and its antagonist effect with human TLR4/MD2^{23,57}. Additionally, the phosphate groups are also important for LPS endotoxic activity. Deletion of one of them results in a 100-fold reduced endotoxic activity and weak activation of the innate immune response⁶².

II.1.2. TLR4-independent detection of LPS

TLR4-MD2-mediated sensing has long been considered the sole pathway for LPS recognition. However, recent studies reported that circulating LPS can also induce a number of receptors in a TLR4-independent manner. These pathways will be described below on the basis of LPS localisation.

II.1.2.1. Extracellular detection of LPS

LPS was found to activate a number of Transient Receptor Potential (TRP) cation channels which are present in airway epithelial cells and sensory neurons. Mammals count of 28 TRP members subdivided into six groups based on the sequence homology: TRPP (polycystin), TRPV (vanilloid), TRPM (melastatin), TRPC (canonical), TRPA (ankyrin), and TRPML (mucolipin). TRP channels share similar structural features: identical or homologous tetramers with each monomer consisting of a six transmembrane (TM) domain, with long intracellular carboxy and amino termini domains harbouring regulatory modules that can vary within groups^{63,64}. TRPs are non-selective cation channels whose activation leads to Ca²⁺ influx (except for TRPM4 and 5) and represent key gateways in downstream signalling pathways regulation.

In sensory neurons, different TRP channels have been identified as LPS sensors, including TRPA1, TRPV1, TRPM3, and TRPM8. Their activation causes Ca²⁺ influx and triggers acute responses such as pain and the local release of neuropeptides, including calcitonin gene-related peptide (CGRP), leading to neurogenic inflammation (Figure 2.3)⁶³. The mechanism behind TRP activation by LPS is still unrevealed, however, it was suggested that TRPA1 activation is directly modulated by the LPS-induced mechanical perturbation of the cytoplasmic membrane with no intracellular mediator needed^{21,65}. Furthermore, Lipid A was found to be necessary for TRPA1 activation, after it was completely inhibited following treatment of neuronal cells with polymyxin B. In addition, lipid A structure modifications resulted in a variation in TRPA1 activation efficiency with the hexa-acyl lipid A being the strongest activator of TRPA1, as observed for TLR4-MD2⁶⁵. On the other hand, in epithelial

cells, TRPV4 was identified as an LPS sensor. This recognition induces local secretion of nitric oxide (NO) production subsequently to Ca^{2+} influx, as a bactericidal defence mechanism (Figure 2.3) ^{63,66}.

In summary, these studies showed that TRP ion channels, notably TRPA1 and TRPV4, endow sensory neurons and airway epithelial cells, respectively, with the ability to promptly detect LPS presence, and initiate the defence against pathogens in a TLR4-independent manner. Moreover, given the overlapping expression of TRP channels and TLRs in various tissues ⁶³, investigating the functionality of dual detection mechanisms and the crosstalk between intracellular signalling pathways activated by both TLR4 and TRP channel activation holds compelling avenues for further investigations.

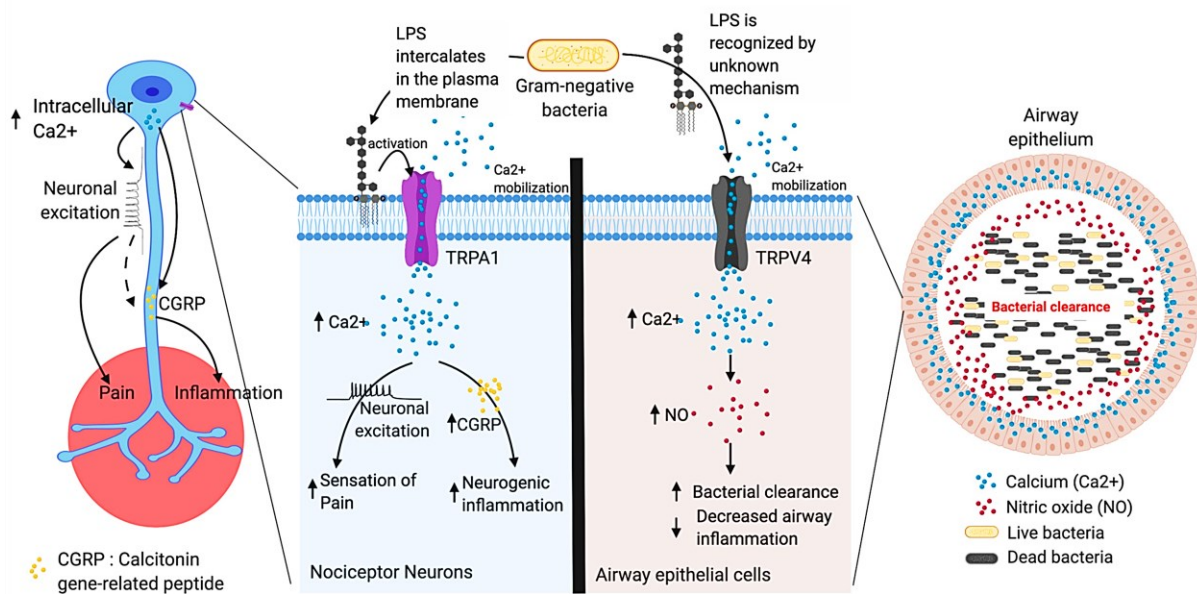


Figure 2.3: LPS extracellular recognition by TRP channels. LPS recognition by TRPA1 in nociceptive neuron results in Ca^{2+} influx and neuropeptides (CGRP) release leading to neurogenic inflammation. Whereas, LPS sensing in the airway epithelium by TRPV4 and the subsequent Ca^{2+} influx results in nitric oxide (NO) production facilitating pathogen clearance. TRP: transient receptor potential, CGRP: calcitonin gene-related peptide, NO: nitric oxide. Adapted from ⁶⁵.

II.1.2.2. Intracellular detection of LPS

While membrane-bound TRP channel and TLR4 recognize LPS extracellularly or within endosomes, specific receptors step up and sense LPS presence at the host cytoplasm. These intracellular sensors consist of caspase-11 (in mice) and caspase-4 and -5 (in humans). Caspases are conserved endoproteases implicated in cell apoptosis and inflammation with a cysteine protease activity that cleaves target proteins only after an aspartate residue ⁶⁵. Different pathways have been presented for LPS entry into the cytosol. Outer membrane vesicles (OMVs), released by bacteria, are endocytosed and result in LPS delivery into the cytosol from endosomes ⁶⁷. Moreover, other studies revealed the implication of a ubiquitous and conserved high mobility group box 1 (HMGB1) protein in LPS access to the cytosol. HMGB1 binds LPS and induces its internalization by macrophages *via* the receptor for advanced glycation end-products (RAGE). Subsequently, HMGB1 disrupts the endosomal membrane leading to LPS leakage into the cytosol ⁶⁸.

Intracellular LPS, released from Gram-negative bacteria-containing vacuoles by guanylate-binding protein GBP (or HMGB1), is sensed by caspase-11 (in mice) and caspase-4 and -5 (in humans), and activates the non-canonical inflammasome. The interaction between caspase and LPS molecules is established by the binding of the caspase domain termed the caspase activation and recruitment domain (CARD) to the lipid A moiety with high affinity. This assembly, accomplished similarly in human and mice, triggers the oligomerization of caspase monomers and their catalytic activity resulting in pyroptosis and secretion of pro-inflammatory cytokines, IL-1 β and IL-18 (Figure 2.4) ⁶⁹.

The molecular mechanism behind pyroptosis involves an effector protein gasdermin D (GSDMD). The latter comprises conserved amino- and carboxy-terminal domains, NTD and CTD. The membrane pore-forming activity of NTD is inhibited in the resting state through its binding with CTD. Following activation, caspase cleaves GSDMD after an aspartate residue in the link between CTD and NTD which is consequently translocated to the plasma membrane where it oligomerizes and induces pore formation ⁵⁵. These pores enable Ca²⁺ influx and cell swelling leading to cell lysis. Furthermore, pro-inflammatory cytokines, IL-1 β and IL-18, are also released. In contrast, the processing and secretion of pro-inflammatory cytokines require the activation of NLRP3 involved in pro-caspase 1 maturation which in turn induces the proteolysis and maturation of pro-inflammatory cytokines (Figure 2.4) ⁶⁹.

Once again, lipid A chemistry governs another inflammatory response. Similarly to what was reported for both TLR4-MD2 and TRP channels, specific lipid A structures are required for caspase oligomerization and subsequent activation. Penta- and hexa-acylated lipid A are

identified as potent activators of caspase-11 activity. In contrast, tetra-acylated lipid A from *Francisella novicida* and *Yersinia pestis* did not result in non-canonical inflammasome activation^{65,69}.

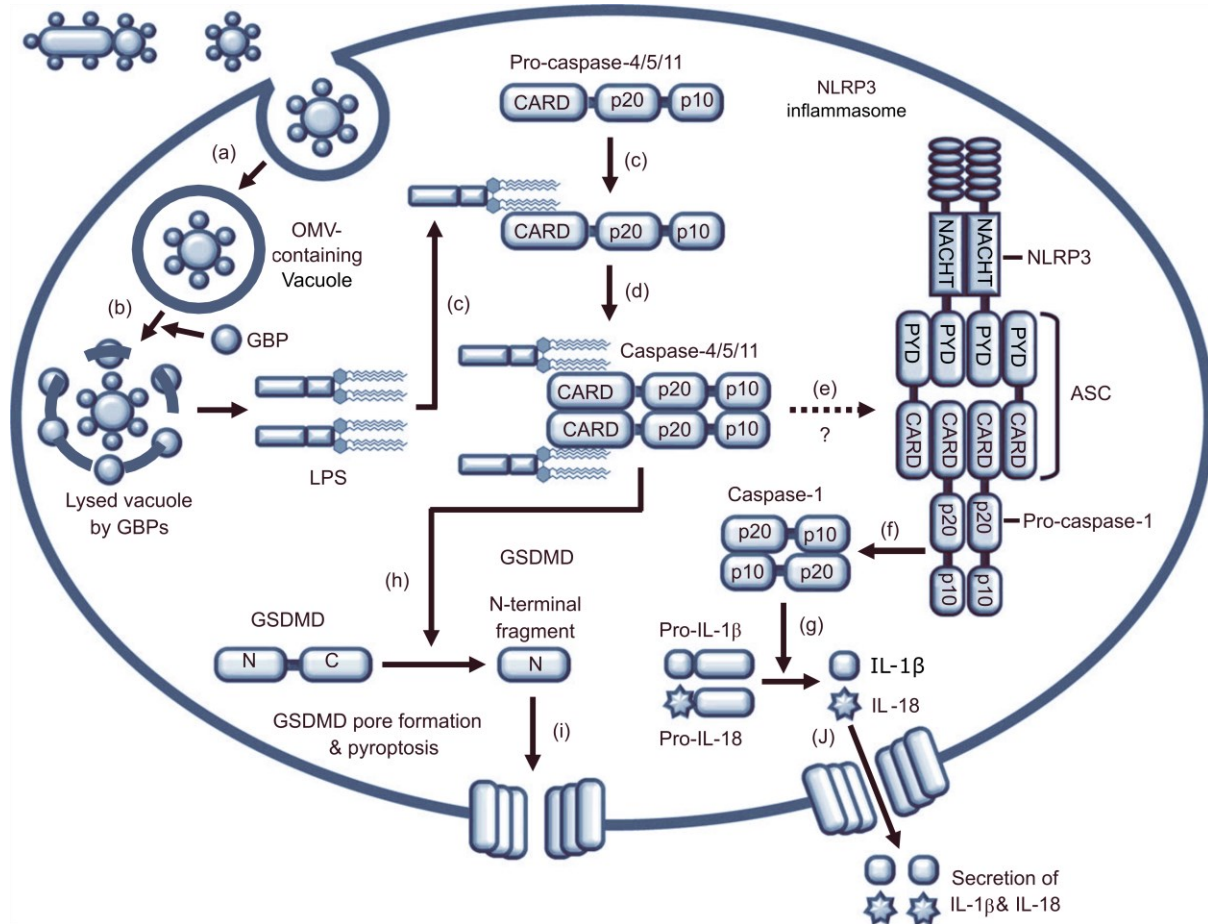


Figure 2.4: LPS-mediated caspase non-canonical inflammasome activation. LPS containing structures are internalized into macrophage by endocytosis (a), then released by GBP (b). The released LPS binds to pro-caspase 4/5/11 CARD motif through its lipid A (c). Activated caspase oligomerizes (d) and activates the NLRP3 canonical inflammasome (e) leading to maturation of pro-caspase-1 (f) which in turn induces maturation of pro-inflammatory cytokines IL-1 β and IL-18 (g). Active caspase-4/5/11 also cleaves GSDMD into N- and C-terminal fragments (h). The cleaved N-terminal fragment binds the cell membrane, oligomerizes, and forms membrane pores, thus resulting in pyroptosis (i). Mature IL-1 β and IL-18 are secreted through GSDMD pores (j). OMV: outer membrane vesicles, Caspase: cysteine-aspartic protease, GBP: guanylate-binding protein, LPS: lipopolysaccharide; GSDMD: gasdermin D, IL: interleukin, CARD: caspase recruit domain, NACHT: nucleotide-binding and oligomerization domain; PYD, pyrin domain. Adapted from⁶⁹.

II.2. LPS detection in plants

Plants are continuously exposed to a variety of potentially pathogenic microorganisms. The plant's innate immune system recognizes the PAMPs by their respective PRRs resulting in an induced immune response referred to as microbe- or pathogen-triggered immunity (MTI or PTI) ⁷⁰. Up to date, identified plant PRRs are exclusively located on the cell surface. They are either receptor-like kinases (RLKs) consisting of a ligand-binding extracellular domain, a transmembrane domain, and an intracellular kinase domain, or receptor-like proteins (RLPs). Unlike RLKs, RLPs lack the intracellular signalling domain suggesting that they consistently operate alongside RLKs to induce downstream regulation signals upon ligand binding ⁷¹. LPS has been reported to trigger several immune responses in plants, namely oxidative burst, nitric oxide production, Ca²⁺ influx, cell wall alterations, and pathogenesis-related (PR) gene expression ⁷⁰.

The model plant *Arabidopsis thaliana* recognizes LPS from *Pseudomonas spp.* and *Xanthomonas campestris* through the lectin S-domain receptor-like kinase LipoOligosaccharide-specific Reduced Elicitation (*AtLORE*). Chemical isolation and degradation of this LPS identified lipid A moiety being the main detected portion by LORE ⁷². Nevertheless, LORE-mediated immune response is only limited to *Pseudomonas* and *Xanthomonas* but does not apply to typical LPS from *E. coli* or *Salmonella enterica* ⁷². *A. thaliana* was also found to respond by nitric oxide production and activation of defence genes expression such as glutathione S-transferases and cytochrome P450 to LPS from different bacteria, including *E. coli*, *Burkholderia cepacia*, and *Pseudomonas spp.* ⁷³, and by a late burst of reactive oxygen species (ROS) to *Pseudomonas* lipid A ⁷⁴. These findings point out the presence of different mechanisms for *Pseudomonas* lipid A perception in *A. thaliana*, which would presumably be translated into diverse responses.

Other studies were carried out to understand the contribution of LPS moieties, lipid A and core oligosaccharide OS, in the plant immune response. These studies were conducted on lipooligosaccharides (LOS) from the plant pathogen *X. campestris* pv. *campestris* (Xcc) strain 8004 whose effect was tested on the immune response in *A. thaliana* leaves ⁷⁵. The structure of LOS Xcc was first determined and was found to comprise dense negatively charged groups in the lipid A-inner core region with exclusive features such as a phosphoramidate group. Intact LOS induced the expression of defence genes PR1 and PR2, with dual temporal phases in the transcript levels: an early accumulation at 12 h and a later, more significant accumulation after 20 h. On the other hand, LOS' chemically obtained derivatives (lipid A and core OS) induced the expression of PR1 and PR2 with different response kinetics. Core OS activated only the

early phase, whereas lipid A induced the later phase, suggesting that different LPS moieties are recognized independently with two distinct plant receptors ⁷⁵.

Alteration in structures within LPS can happen and may lead to bacterial resistance and attenuation of the induced host's immune response. In correlation with what was reported on the effect of LPS structure modifications in mammal sensing, a similar picture can occur for plant immune systems. The structure of LOS from non-pathogenic Xcc mutant strain 8530 presented a truncated core region with alterations in its lipid A acylation and phosphorylation patterns. With respect to wild-type Xcc lipid A, which was mainly hexa-acylated, Xcc 8530 lipid A was mainly penta-acylated with both phosphate groups substituted with phosphoethanolamine groups (PEtN) ⁷⁶. These modifications greatly influenced its ability to trigger plant immune response in *A. thaliana* through defence genes' expression as described in the wild-type strain LOS ⁷⁵ and highlight important features in LPS's lipid A chemistry (net negative charge and acylation degree) that govern its activity and perception.

Unlike LPS immune sensing in mammals, the concentration of LPS required for the plant immune responses described above is within the range of 5–100 µg/mL, in contrast to the pg/mL to ng/mL range concentration in mammals. These comparisons suggest that LPS binding with plant receptors is of low affinity ⁷³ compared with other bacterial PAMP factors detected at subnanomolar levels ⁷⁷. It was suggested that this difference might be attributed to PAMPs physicochemical properties within the purified preparation. LPS/LOS for instance are amphiphilic molecules with low solubility that tend to form aggregates above critical aggregation concentrations which provide limited access of LPS to their corresponding receptors on the plant cell wall ⁷⁸.

II.3. LPS detection by lectins

II.3.1. C-type lectin receptors: overview

The human immune system is equipped with a variety of PRRs which detect specific molecular features of foreign organisms. Among these PRRs, the C-type lectin receptors (CLRs) superfamily expressed by antigen-presenting cells (APCs), including dendritic cells (DC) and macrophages, are found as transmembrane proteins or are secreted as soluble proteins ⁷⁹. CLRs are known for their ability to recognise exposed carbohydrate structures present on self- and non-self-molecular motifs. This recognition is mediated by one or multiple carbohydrate recognition domains (CRD) in a Ca²⁺-dependent manner ⁸⁰. Some transmembrane CLRs share the same domain organization consisting of an extracellular domain (ECD),

comprising a coiled-coil domain and the CRD, a transmembrane domain, and an intracellular domain with or without a signalling motif (Figure 2.5) ⁸¹.

Based on their topology, transmembrane CLR receptors are divided into two groups: type I receptors with their N-terminus pointing outwards the cytoplasm of the cell, and type II receptors with an N-terminus pointing inwards the cytoplasm ⁸³. Within both groups, CLR receptors can be categorized depending on the amino acid motif determining their glycan specificity within CRDs. CLR receptors with an EPN (Glu-Pro-Asn) motif are associated to glycans with disposed equatorial 3- and 4-OH groups like fucose, mannose, glucose, and N-acetylglucosamine (GlcNAc) glycans

⁸¹. In contrast, CLR receptors with a QPD motif (Gln-Pro-Asp) recognize glycans with axial 4-OH groups like galactose and N-acetylgalactosamine (GalNAc) terminated glycans ^{81,84}.

CLR receptors exhibit varying numbers of CRDs spanning from a single domain, including dendritic cell-specific intracellular adhesion molecule (ICAM)-grabbing non-integrin receptor (DC-SIGN) and Macrophage Galactose-type Lectin (MGL), to eight or ten different CRDs as in macrophage mannose receptor (MMR) ⁸⁵. All CRDs share a series of structural features: they consist of six to seven β -strands organized into two antiparallel β -sheets with two flanking α -helices ($\alpha 1$ and $\alpha 2$). The overall domain has a double-loop structure, including a huge loop with N- and C-terminus β -strands ($\beta 1$ and $\beta 5$) joined together to form an antiparallel β -sheet, and stabilized with a conserved disulfide bridge (linking $\beta 5$ and $\alpha 1$). The second loop, called the long loop region, is stabilized by another conserved disulfide bridge. Whereas, the second antiparallel β -sheet is formed by strands $\beta 2$, $\beta 3$ and $\beta 4$ (Figure 2.6) ⁸⁶. In the case of long-form CRDs, a short N-terminal extension forming an additional β -hairpin is also identified, which is stabilized by a disulfide bridge ⁸⁶. Another common feature is the presence of calcium ions in CRDs, with up to four Ca^{2+} binding sites described so far (Figure 2.6).

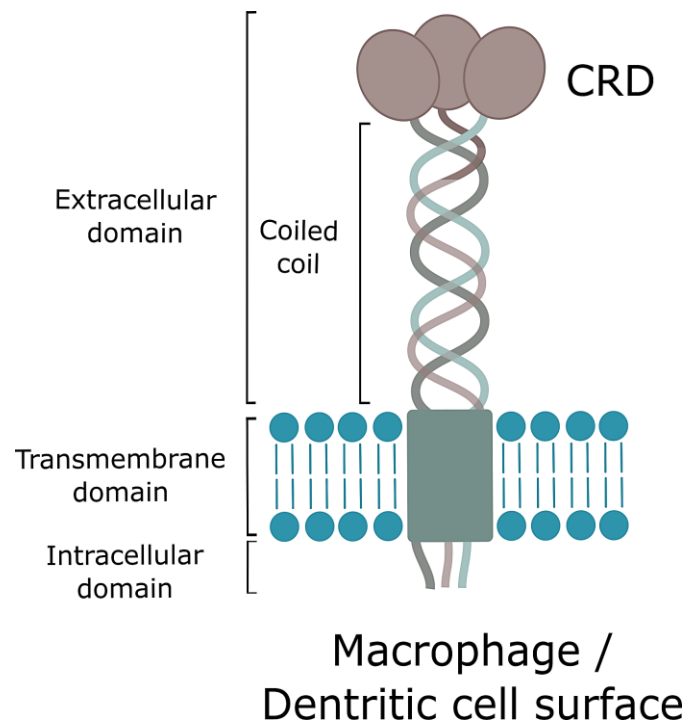


Figure 2.5: Domain organization of Macrophage Galactose-type Lectin (MGL) at the surface of antigen presenting cells. Adapted from ⁸².

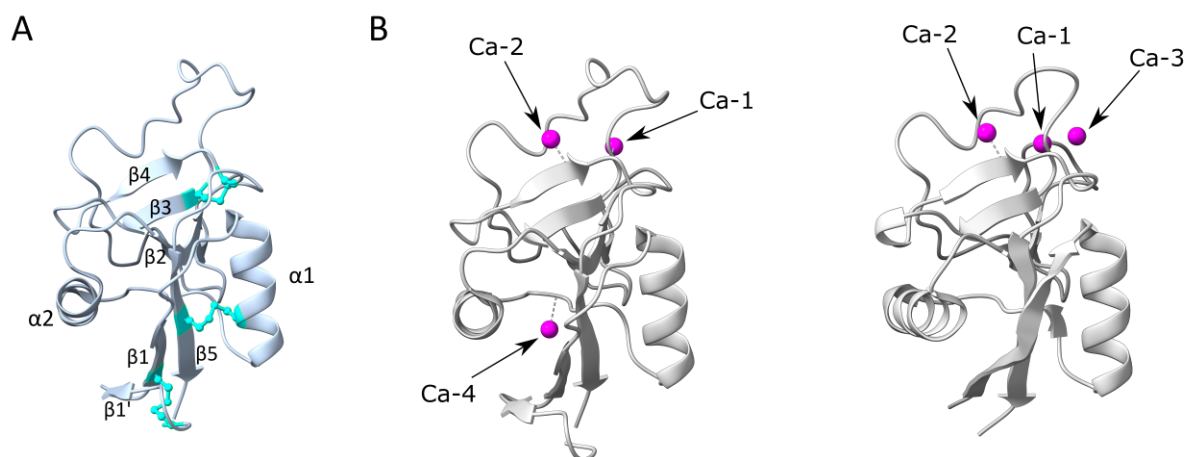


Figure 2.6: CRD domain overall organization. (A) CRD secondary structure elements numbering (PDB: 1K9I). (B) Ca^{2+} -binding sites in human ASGPR-1 in left (PDB: 1DV8, left) and rat MBP-A (2MSB, right). Ca^{2+} ions are shown as magenta spheres. Adapted from ⁸⁶.

CLR CRDs' binding to mono-carbohydrates is usually of low affinity (in the millimolar range), although higher affinities have been reported for CLRs. That could be attributed to both the density of the presented glycans on the cell surface together with the degree of multimerization of CLR. Thereby, lectins offset for their CRDs low affinities towards mono-sugars by the fact that the recognized targets are highly glycosylated, and that they themselves are multimeric proteins. These properties combined together lead to strong interaction affinities through the different avidity-contributing modes that could occur at the cell surface, namely multivalence, clustering, and rebinding effects. In this way, the arrangement of several CRDs in multimers enables simultaneous binding of multivalent ligands resulting in an enhanced affinity (avidity effect) and kinetics of binding ⁸⁷. This concept known as the “face-to-face” mechanism is proposed for the binding of tetrameric DC-SIGN which displayed an increased affinity for its multivalent mannan ligands, with possible additional secondary interactions on the CRDs contributing to the binding ⁸⁸. While a $\sim 10^6$ -fold increase in affinity of the asialoglycoprotein receptor was reported for a trivalent ligand with N-acetyllactosamine (LacNAc) branches relative to monovalent LacNAc ^{87,89}.

Different functions have been attributed to CLRs in host defence and homeostasis including cell-cell adhesion, host-pathogen interaction, antigen uptake, and modulation of innate immunity^{90,91}. The ability of CLRs to activate or inhibit the immune response is guided by the intracellular motif. Some CLRs present an immunoreceptor tyrosine-based inhibitory motif (ITIM) which usually mediates immune-suppressive functions. Whereas CLRs with an immunoreceptor tyrosine-based activation motif (ITAM) result in activation functions⁹¹. Some CLRs, including Macrophage inducible C-type lectin (Mincle) can induce indirect signalling through association with the ITAM motif containing adaptor molecules, Fc receptor γ -chain (FcR γ) or DAP12. Whereas others, such as Dectin-1, involve direct signalling through their cytoplasmic ITAM motif. These two mechanisms lead to the activation of spleen tyrosine kinase (Syk) which catalyses the phosphorylation of ITAM tyrosine residues. Downstream events involving a complex comprising CARD9, mucosa-associated lymphoid tissue lymphoma translocation protein 1 (Malt1), and B cell lymphoma 10 (Bcl10) take place⁸⁰, resulting in cellular responses induction, including phagocytosis, DC maturation, inflammasome activation, and cytokine and other mediator production (Figure 2.7)⁸⁰. On the other hand, ITIM-bearing receptors like dendritic cell immunoreceptor (DCIR) mediate inhibitory functions through the recruitment of tyrosine phosphatases (SHP-1 and SHP-2) to its cytoplasmic domain. This results in a modulation of signalling pathways induced by other PRRs through inhibition of TLR-induced cytokine production (IL-12 and TNF production by TLR8 or IFN α and TNF production by TLR9)^{80,92}. Finally, another category of CLRs without defined ITAM or ITIM domains exists, as in the case of MMR, DC-SIGN, and MGL. These last two lectins mediate endocytosis, and possess cytoplasmic domains with YXX Φ motifs (Y represents a tyrosine, x represents any amino acid, and Φ denotes a hydrophobic amino acid)⁹³ facilitating clathrin-mediated endocytosis through direct interaction with the adaptor protein AP-2 (Figure 2.7)⁹⁴⁻⁹⁶.

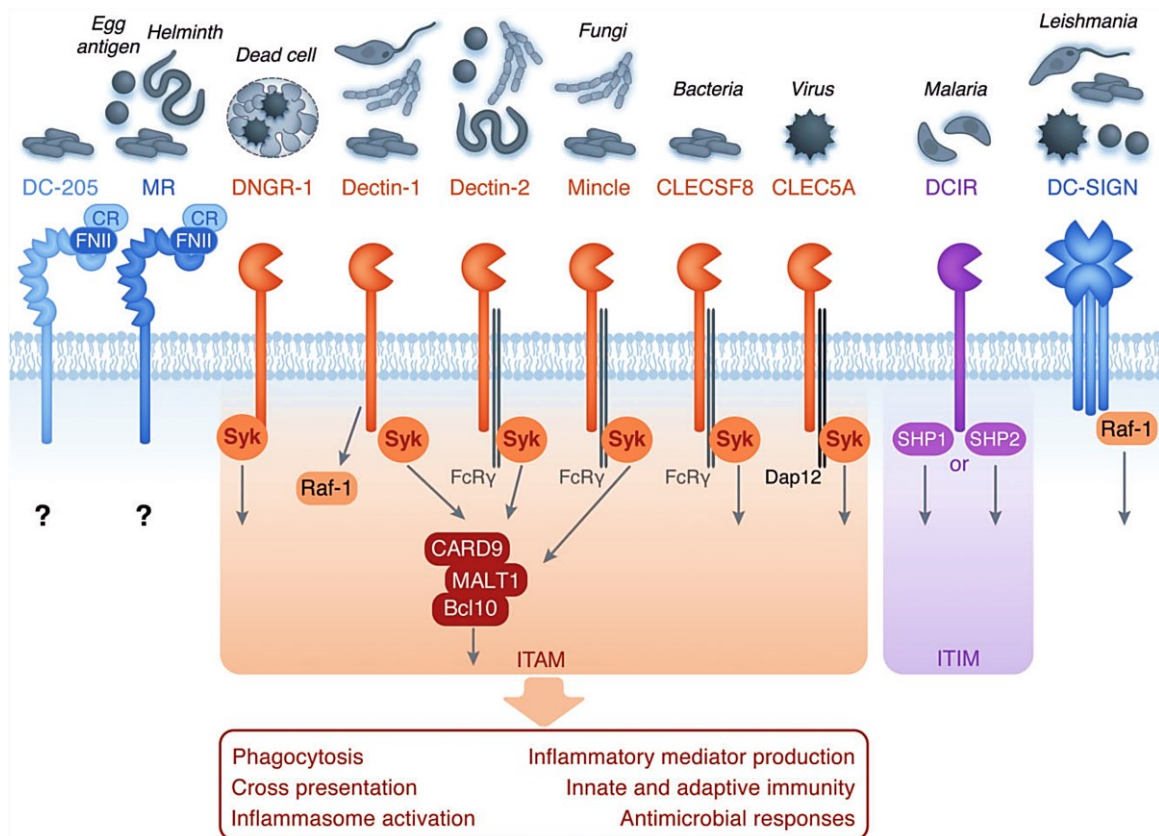


Figure 2.7: Illustration of the major signalling pathways used by CLR upon pathogen recognition. ITAM: immunoreceptor tyrosine-based activation motifs, ITIM: immunoreceptor tyrosine-based inhibitory motifs. CR: cysteine-rich domain. FNII: Fibronectin domain. Adapted from ⁸⁰.

II.3.2. CLR in bacterial glycans recognition

Bacterial cell walls are decorated with a variety of glycan structures, including LPSs in gram-negative, and PG in gram-positive bacteria. These glycoconjugates are potent ligands for CLR. *Mycobacterium tuberculosis* (Mtb) recognition by CLR is very well characterized and can involve multiple CLR including dendritic cell-associated C-type lectin-1 (Dectin-1), Mincle, DC-SIGN, and MMR. Each of these CLR recognises different glycan features on Mtb surface: DC-SIGN recognizes mannose-capped lipoarabinomannan (ManLAM) and α -glucan in Mtb, and prevents DC maturation and induces IL-10 production. MMR also binds to ManLAM and mediates bacterial phagocytosis by limiting phagosome–lysosome fusion within macrophages. Whereas Mincle recognizes the mycobacterial cord factor, trehalose-6,6'-dimycolate (TDM), and modulates the immune response by granuloma formation ^{80,97}. However, these CLR appear to be redundant in mycobacterial recognition *in vivo*, since only

a subset of CLR-deficient mice models showed increased susceptibility to mycobacterial infection^{81,97}.

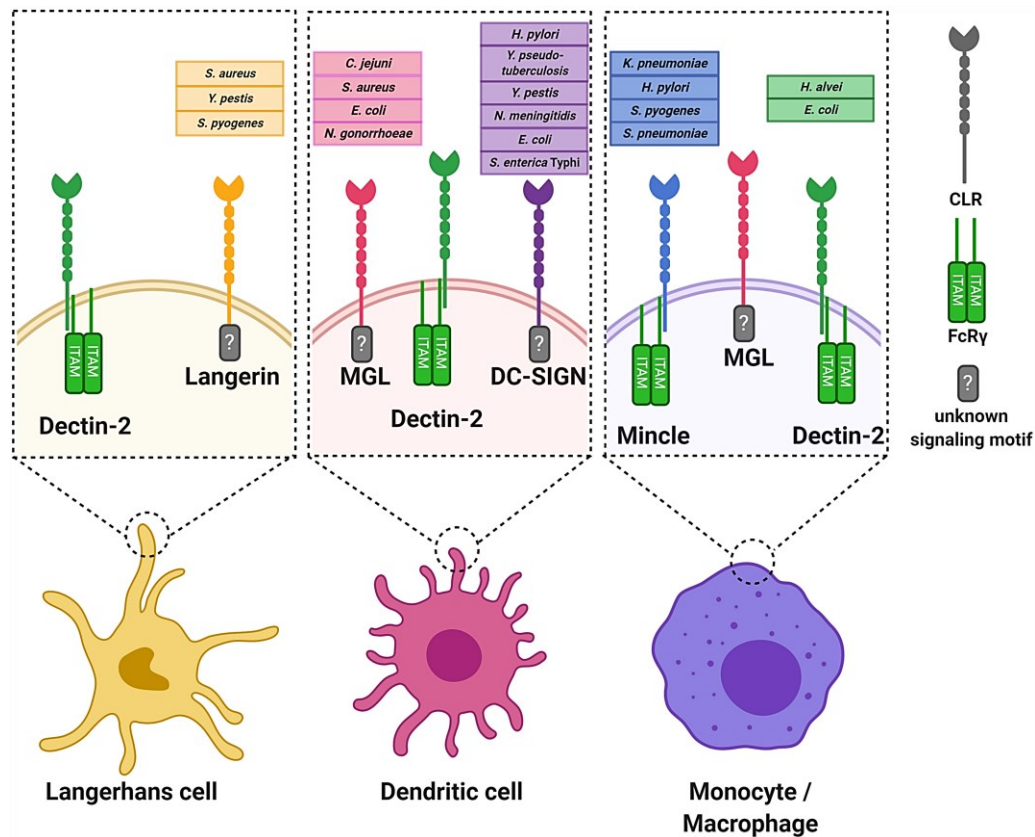


Figure 2.8: Overview of some CLR receptors expressed on different antigen-presenting cells with an indication of some of their respective recognized bacterial species. For each CLR, the corresponding signalling motif, if known, is stated. Adapted from⁸¹.

In addition to *M. tuberculosis*, other bacteria are also recognized by CLR receptors (Figure 2.8). DC-SIGN interacts with a wide range of bacterial pathogens including *Helicobacter pylori*, *Mycobacterium leprae*, *Lactobacillus acidophilus* commensal bacteria, and *E. coli* F470 (R1) strain^{81,92,98}. The MMR also recognizes other bacterial species, including *Mycobacterium kansasii*, *Klebsellia pneumoniae*, and *Streptococcus pneumoniae*⁹². Mincle on the other hand, recognizes *K. pneumoniae*, *H. pylori*, and also Group A Streptococcus (GAS or *Streptococcus pyogenes*) which results in the production of pro-inflammatory cytokines, and reactive oxygen species upon recognition of monoglucosyldiacylglycerol (MGDG), a component of lipoteichoic acid anchor^{81,99}. In contrast, MGL recognizes *Staphylococcus aureus*, *Campylobacter jejuni*, *Nesseria gonorrhoeae*, and *E. coli* glycan structures^{81,100,101}.

Pathogen recognition by CLR receptors results in diverse immune responses depending on the induced signalling pathway. Importantly, previous studies revealed that these interactions are

not always supportive of the host defence. Instead, bacteria take advantage of these interactions to evade immunity and ensure their survival. Once again, lectin-mediated immune escaping of Mtb is one of the most studied evasions. Mtb entry into the cell can be mediated by several CLRs. Mtb can enter macrophages by MMR upon binding to its ManLAM. Subsequent to phagocytosis, Mtb prevents phagosomes maturation by limiting phagosome–lysosome fusion thus allowing intracellular survival¹⁰². Additionally, Mtb through interaction with DC-SIGN enters DCs and results in the inhibition of DC maturation and production of anti-inflammatory cytokine IL-10¹⁰³. Other bacteria have evolved and developed camouflage strategies of their surface glycan structures to avoid recognition by the host immune system. That was described for some *Klebsiella* serotypes lacking the manno(rhamno)biose repeats in their capsular polysaccharides, which could prevent recognition by host lectins^{97,104}. In contrast, *Neisseria sp.* produces LOS structures subject to natural variations in their terminal carbohydrate residues. These variations are recognized by different CLRs, including DC-SIGN and MGL. More importantly, both CLRs induce similar DC maturation, but differ in the produced cytokine profile and T cell polarization, thus shifting subsequent immune responses in favour of bacterial survival¹⁰⁰. Finally, *H. pylori* takes advantage of the Mincle receptor to evade the innate immune system recognition by upregulating Mincle expression in macrophages. While Mincle-knockdown human macrophages presented an elevated pro-inflammatory cytokines production with low anti-inflammatory cytokines expression during infection, Mincle expression and interaction with *H. pylori* act as a maintaining balance between pro- and anti-inflammatory cytokine production during infection¹⁰⁵.

II.3.3. Focus on Macrophage Galactose-type Lectin MGL

II.3.3.1. Structural features

The human MGL (hMGL), also known as CD301 or CLEC10A, is a 30 kDa type II transmembrane C-type lectin receptor. In contrast to most other human lectins, MGL's structure displays a characteristic QPD motif at the long loop region associated with galactose (Gal) and N-acetylgalactosamine (GalNAc) terminated glycans binding. Structurally, MGL comprises an N-terminal cytoplasmic domain with an YxxΦ endocytosis motif, a transmembrane domain, and an extracellular domain (ECD). This latter comprises a coiled-coil trimerization neck and a C-terminal carbohydrate recognition domain (CRD) (Figure 2.5)^{93,106}.

The crystal structure of the CRD of hMGL revealed a similar organization as described for other C-type lectins⁸⁶. It consists of five β-strands arranged in two β-sheets flanked on

opposite sides by two α -helices (Figure 2.9). Three intrachain disulfide bridges are found involving cysteine residues 181-192, 209-304, and 282-296. hMGL-CRD, like other long-form CRDs, displays a short N-terminal extension forming an extra β -hairpin which is stabilized by a disulfide bridge (between C181 and C192) connecting the N-terminus to β 1⁹³.

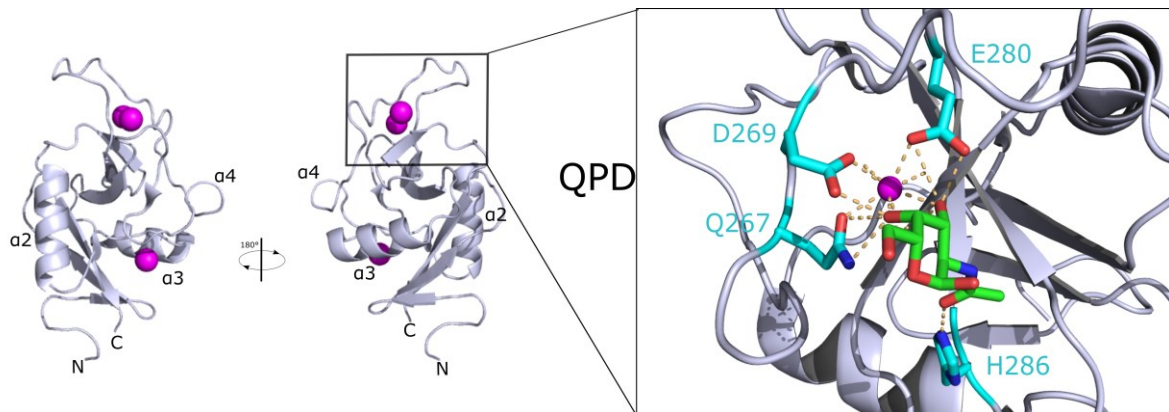


Figure 2.9: Crystal structure of hMGL CRD domain. The structure of the CRD domain of hMGL (PDB: 6PY1, grey) is shown on the left. The observed three calcium ions are shown as pink sphere. A close-up view on GalNAc sugar binding to the calcium-binding site is shown on the right. The CRD amino acids involved in GalNAc binding are shown as sticks. GalNAc is represented as green sticks. Adapted from⁸².

MGL-CRD crystal structure revealed that in the absence of a ligand, no electron density was obtained for a portion of the long loop presumably due to its flexibility and thus could not be modelled. In contrast, in a complex with galactose-containing ligands, the long loop is ordered and binds the two additional calcium atoms (Figure 2.9). The equatorial 3-OH and axial 4-OH groups of Gal/GalNAc residues are bound to the calcium ion. Additional contacts established by the 2-acetamido group (NHAc) with H286 through a water-mediated hydrogen bond justify the MGL binding preference for GalNAc with an over 70-fold greater affinity over the simple galactopyranoside unit of Gal (Figure 2.9)^{93,107}.

Human MGL (hMGL) is exclusively expressed by macrophages and dendritic cells within the immune system. Humans possess a single MGL gene, which undergoes extensive splicing, resulting in different MGL isoforms with different lengths. Alternative splicing mainly targets exon 6, encoding the terminal neck region. Three isoforms of hMGL have been described¹⁰⁶: MGL isoform 2 originally cloned from macrophages (National Center for Biotechnology Information accession number NP_006335), isoform 3 cloned from dendritic cells (accession number NP_001316999.1), and isoform 1 (accession number NP_878910)

cloned from immature dendritic cells that has an additional insertion of 27 amino acids in the neck domain^{106,108}. In contrast, mice have two related lectins, termed MGL1 (CD301a) and MGL2 (CD301b), with the latter being homologous to hMGL^{107,109}.

II.3.3.2. Functional features

MGL stands out as the sole C-type lectin receptor within the human immune system that specifically recognizes terminal N-acetylgalactosamine (GalNAc) glycans. Numerous tumour-associated carbohydrate antigens (TACAs) have been proposed as hMGL ligands, such as (i) the Thomsen nouveau (Tn)-antigen (GalNAc- α -1-O-Ser/Thr) frequently detected in cancers (90% of carcinomas) but is masked on healthy cells¹¹⁰, (ii) GalNAc moieties present in pathogens such as helminth parasites, and (iii) self-antigens such as glycosphingolipids (GM2 and GD2)¹¹¹. In mice, MGL2 shares the specificity of hMGL with an additional capacity to bind terminal galactose glycans, while MGL1 presents a high specificity for Lewis X (Gal- β (1 \rightarrow 4)Fuc- α (1 \rightarrow 3)-GlcNAc) structures¹⁰⁹. Following recognition of Tn-structures, MGL can act as an endocytic receptor expressed on antigen-presenting cells, and participate in antigen uptake and presentation, resulting in T-cell activation and induction of the adequate immune response. On the other hand, MGL interactions can mediate immunosuppressive responses resulting in efficient cancer cell survival, proliferation, and metastasis¹⁰⁹.

Antigen-presenting cells expressing hMGL are distributed throughout various regions of the human body. This facilitates the recognition of pathogens that display the GalNAc epitope within tissues, such as glycan structures in bacteria. Different bacterial strains were described for MGL recognition. *Neisseria gonorrhoeae* LOS phenotype C, a LOS bearing a terminal GalNAc, was described as the first bacterial ligand for hMGL. This binding was found to shift dendritic cells' cytokine secretion and T helper cell differentiation toward Th2 profiles. These changes led to bacterial survival, as the immune polarization toward Th2 profiles is considered as a less effective response against *N. gonorrhoeae* infection^{81,109}.

MGL is also described to recognize two gastrointestinal bacteria, *E. coli* R1 and *Campylobacter jejuni*, either by their LOS terminal GalNAc residues or through the expression of N-glycosylated proteins, respectively^{81,101,109}. *C. jejuni* was reported to bind MGL to suppress IL-6 production by dendritic cells. In contrast, *C. jejuni* mutant strains lacking N-glycan protein modifications, which do not bind to hMGL, induced pro-inflammatory cytokines production¹¹².

Additionally, in mice, MGL1 was found to play an important antibacterial and regulatory role during *Mycobacterium tuberculosis* and *Klebsiella pneumoniae* lung infections. These infections resulted in MGL1 upregulation and accumulation of its expressing cells at the

inflammation sites. More importantly, MGL1 silencing resulted in increased mortality compared with MGL1-sufficient wild-type mice. Moreover, MGL1 deficiency triggered the production of pro-inflammatory cytokines (IL-1 β , IL-6) and showed a hyperinflammatory response. Altogether, this evidence indicates that MGL1-binding Lewis X epitopes play an immunomodulatory role in the lung microenvironment by reducing excessive inflammatory responses ¹⁰⁹.

II.4. Bacterial escaping strategies: pathogens and commensals

As mentioned throughout the previous sections, bacterial pathogens adapt their LPS structures in order to escape the immune response or acquire antibiotic resistance. Briefly, the main modifications occurring in lipid A moiety consist in changing the overall negative charge of lipid A through the removal of phosphate(s) or the addition of chemical groups to the lipid A disaccharide backbone. These alterations reduce the net negative charge of LPS and protect bacteria against CAMPs ³⁴. Several pathogens including lung pathogens, such as *P. aeruginosa* and *Burkholderia sp.*, remodel their lipid A structures and add Arap4N. Similarly, some Gram-negative species including *Proteus mirabilis*, the causative agent of urinary tract infections, and *K. pneumoniae*, another human lung pathogen employ Arap4N modification of the LPS to combat CAMPs action ²¹.

Secondly, changes in the acylation pattern in lipid A structure result in the attenuation of endotoxic properties of LPS and immune surveillance evasion ^{23,25}. That is the case for *H. pylori*, *Yersinia pestis*, and *Francisella tularensis* which synthesize tetra-acylated lipid A species with a very weak or no endotoxic effect on the human TLR4-mediated signalling activation ²¹. Both, net charge changes and acylation pattern modifications can occur on lipid A. *H. pylori* is found to synthesise a tetra-acylated lipid A with a PEtN decoration on the glucosamine unit conferring both the resistance to CAMPs and the evasion of TLR4 detection ²¹.

It is noteworthy that modifications occurring in the LPS structure are perceived not only as a way for pathogenic bacteria to survive and colonize the host environment, but also as a strategy employed by harmless bacteria in the commensal and beneficial microbiota to thrive and survive within the host ²¹. The presence of a tremendous number and varieties of Gram-negative bacteria inhabiting our bodies suggests that the host immune system has evolved to meticulously distinguish microbial communities, and tolerate their commensal LPS. In that perspective, immune system modulation and immunological differences are related to the

chemical structure of LPS. Modifications of the acylation and phosphorylation pattern, already described for pathogenic strains, appear to be fundamental for the persistence and survival of commensals in the intestine²¹. In the human gut, Proteobacteria, expressing bis-phosphorylated hexa- or hepta-acylated lipid A forms, contribute to pro-inflammatory LPSs, whereas Bacteroidetes, producing penta- or tetra-acylated, usually mono-phosphorylated forms, contribute to anti-inflammatory LPSs. This suggests that the gastrointestinal tract might be characterized by a continuous immunologic tolerance governed by the balance between the induced inflammatory and the weak inflammatory lipid A¹¹³.

Structural elucidation of LPS from *Bacteroides vulgatus*, a commensal of the human and murine intestine, revealed a particular chemical structure¹¹⁴. The lipid A comprised a mixture of tetra- and penta-acylated species. These ones were exclusively phosphorylated at the reducing glucosamine unit. This chemistry results in an attenuated LPS endotoxic activity and a weak TLR4-mediated activation potency. Furthermore, the inner part of the core oligosaccharide contained a galactofuranose unit, a sugar moiety rarely found in bacterial LPS, which can be recognized by specific lectins favouring bacterial resilience in the intestine¹¹⁴. On the other hand, its O-antigen moiety consisting of a repeating unit of [(→4)-α-L-Rhap-(1→3)-β-D-Manp-(1→)] showed selective binding to DC-SIGN associated with gut lymphoid tissues¹¹⁵.

Unusual chemical LPS structures govern the coexistence of bacteria within the human body. Nevertheless, only little is known about this coexistence, hence the need for thorough characterization of LPS structures derived from commensal bacteria to enhance our comprehension of the interplay between microbes and hosts.

III. LPS: an antibiotic target

III.1. Antibiotic discovery and mode of action

Before antibiotics availability, pneumonia, tuberculosis, and gastrointestinal infections were the major three death causes ¹¹⁶. A revolutionary era of antibiotics began back in 1928 when Alexander Fleming found on his workbench a contamination of a *Staphylococcus aureus* petri dish with *Penicillium notatum* fungi. Subsequently, in 1940, scientists Ernst Chain and Howard Florey successfully isolated Penicillin, which was later approved for clinical use in 1941. In the late 1930s, Selman Waksman initiated the Golden Age of Antibiotic discoveries, between the 1940s and the 1960s. He launched novel antimicrobial screening studies, using microbes as antimicrobial producers, leading to the discovery of numerous antimicrobials made by soil-dwelling actinomycetes, namely neomycin and streptomycin ¹¹⁷.

All clinically used antibiotics since the early 1900s can be divided into three main categories: natural products if isolated directly by large-scale fermentation of bacteria or fungi, semi-synthetic antibacterials for compounds produced by chemical synthesis using natural products as a base, and fully synthetic antibacterials for compounds that are fully produced by chemical synthetic routes ¹¹⁸. Antibiotics can be classified based on their target, in addition to their mode of action, i.e., whether they induce cell death (bactericidal drugs) or inhibit cell growth (bacteriostatic drugs). Most bactericidal antimicrobials target key events in bacteria's cellular homeostasis, namely protein synthesis (targeting ribosomal subunits), nucleic acid synthesis (maintenance of chromosomal topology), and cell wall synthesis (targeting peptidoglycan synthesis) (Figure 3.1) ^{119,120}.

Beta-lactam antibiotics have been a primary choice for the treatment of bacterial infections due to their high specificity. β -lactams and glycopeptides family antibiotics are known to interfere with homeostatic cell wall biosynthesis by causing changes to cell shape and size, inducing cellular stress responses, and cell lysis ¹²¹. β -lactams share a common chemical moiety commonly called “ β -lactam ring” or “azetidinone” ¹²². This beta-lactam ring displays remarkable structural mimicry with the backbone of the D-alanyl-D-alanine, the substrate of transpeptidases responsible for the cross-linking of peptidoglycan units ¹²³. Therefore, penicillin was proposed to covalently bind to transpeptidases, thus called penicillin-binding proteins (PBPs), and block the bacterial cell wall synthesis ¹²⁴. In contrast, glycopeptides, a group of natural and semisynthetic glycosylated peptides, bind to the D-Ala-D-Ala terminus of the lipid II bacterial cell-wall precursor ¹²⁵. This binding results in the prevention of both transpeptidation and transglycosylation through substrate sequestration and

associated steric hindrance¹²⁵, respectively. The oldest glycopeptide described was vancomycin (from the root word “vanquish”) and is still in clinical use since 1955¹²⁵. It should be recalled that β -lactams are effective against both Gram-positive and Gram-negative bacteria. In contrast, glycopeptides exclusively target Gram-positive bacteria due to low permeability of the outer membrane of Gram-negative bacteria¹¹⁹.

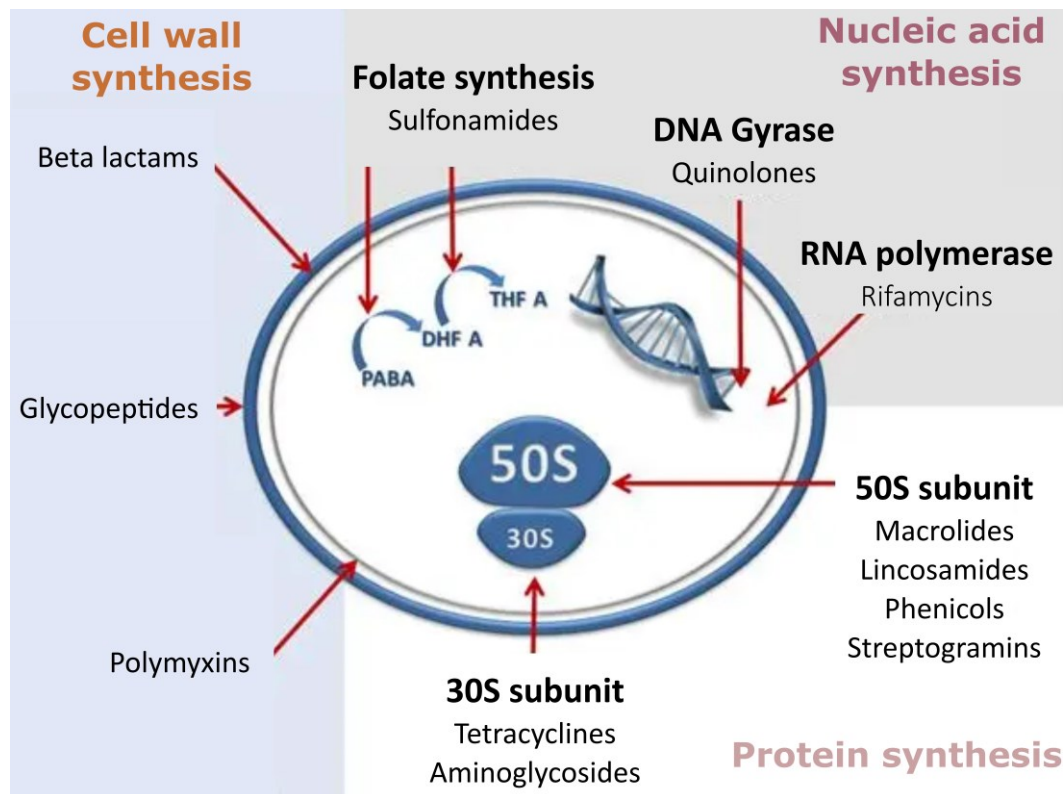


Figure 3.1: Classification of antibiotics on the basis of their targets. Adapted from 120,131.

Antibiotics, since their discovery and utilization in the first half of the 20th century, played a pivotal role in saving millions of lives from severe diseases and bacterial infections. Apart from infectious diseases, antibiotics have also been crucial in facilitating numerous advanced medical procedures, including cancer treatment, organ transplants, and open-heart surgery¹²⁶. Additionally, these compounds made remarkable contributions to animal husbandry and aquaculture¹²⁷.

While various antibiotic families have been introduced over time, not all remain effective today, as the emergence of resistant bacterial strains has hindered their efficacy¹²⁸. Antibiotic resistance is the ability of bacteria to withstand the detrimental effects of an antibiotic that was previously effective against them. Antimicrobial resistance is not new. In

fact, back in 1929, Alexander Fleming had already isolated penicillin-resistant strains, including *E. coli*, *Salmonella enterica* serovar Typhi, and *Haemophilus influenzae* strains¹²⁹. Drug-resistant strains initially appeared in hospital facilities, where most antibiotics were utilized. In the 1930s, sulfonamide-resistant *Streptococcus pyogenes* appeared in military hospitals. Shortly after penicillin introduction, London civilian hospitals encountered penicillin-resistant *Staphylococcus aureus*. Likewise, *Mycobacterium tuberculosis* with streptomycin resistance emerged soon after the antibiotic's discovery¹³⁰. However, the rate of antibiotic-resistance development together with the extreme decline in antibiotic research is the most concerning.

III.2. Mechanism of antibiotic resistance used by bacteria

Regardless of the resistance origin, a variety of mechanisms for bacterial resistance have been proposed (Figure 3.2)¹³¹. Some are directed at the antibiotic itself and ensure its alteration/modification, as in the case for penicillin destroyed by enzymes such as β -lactamases. Others alter the antibiotic's intracellular target, e.g. modification of proteins involved in cell wall synthesis such as PBPs. Finally, other mechanisms target the drugs' transport, notably through decreased influx and active efflux (e.g. resistance to the tetracyclines)¹³⁰.

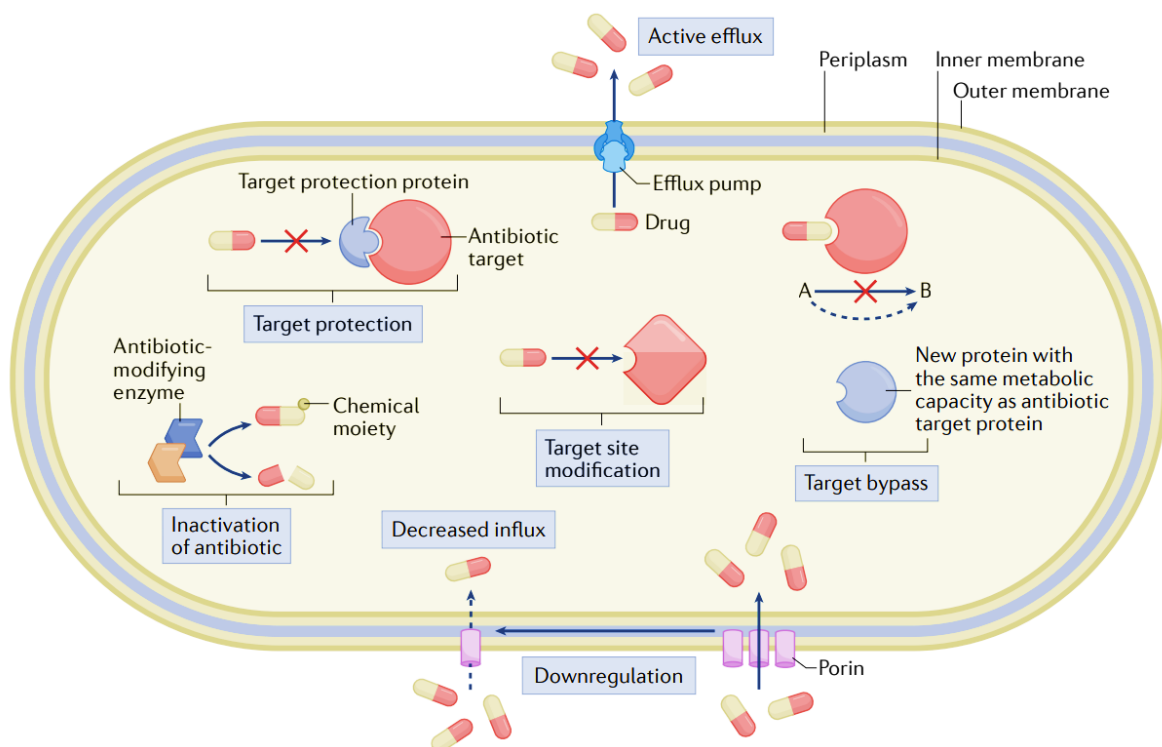


Figure 3.2: Overview of antibiotic resistance strategies in bacteria. Adapted from¹³¹.

III.2.1. Resistance through target modification and protection

One way for bacteria to hinder antibiotic binding is by modifying the target's structure or protecting it with other chemical groups, which would shield it from the antibiotic (Figure 3.2). Target protection usually involves the physical association of the antibiotic target with a resistance protein (target protection protein). This mechanism is described for the resistance to antibiotics of the tetracycline class. Tetracyclines are known for their inhibitory action in protein synthesis through bacterial translation inhibition by binding to the 30S subunit of the bacterial 70S ribosome. Numerous Tetracycline Ribosomal Protection Proteins (TRPPs) have been described to date, including Tet(O) and Tet(M). These later bind to the ribosome, at the interface of the tetracycline binding site and the 16S rRNA, and drive tetracyclines' displacement. TRPPs encoding genes are found in a diverse range of pathogens, with Tet(M) standing out as the predominant tetracycline resistance determinant in clinical isolates of staphylococci, streptococci, and enterococci ¹³².

On the other hand, target modifications are often associated with random mutations of bacterial genes on the chromosome that accumulate during growth, and even expand under drug pressure. This mechanism is described for different antibiotic classes. For example, mutation in genes encoding PBPs results in their alteration leading to decreased susceptibility to β -lactams. That is the case for *Enterococcus faecium* and *Streptococcus pneumoniae* resistance to ampicillin and penicillin, respectively ¹²⁰. Furthermore, keeping up with cell wall antibiotics, glycopeptides inhibit cell wall synthesis by binding to D-alanyl-D-alanine residues of peptidoglycan precursors. Alteration of this precursor, through D-alanyl-D-alanine changing to D-alanyl-lactate or D-alanyl-serine, prevents their recognition by glycopeptides and the subsequent inhibition of the transpeptidation. This resistance is mediated by clusters of van genes. Among the 11 clusters described to date, six encode for the key enzyme of D-alanyl-D-lactate synthesis, which confers high-level of resistance to vancomycin with MICs > 256 mg/ml. The remaining five clusters confer low-level of resistance with MICs of 8–16 mg/ml and encode for the enzyme that is responsible for the synthesis of D-alanyl-serine ¹³³.

III.2.2. Resistance through antibiotic inactivation and degradation

Another resistance mechanism in bacteria is the modification of the antibiotic (Figure 3.2) achieved by enzymes that can either cause its modification or degradation. Antibiotic-modifying enzymes have been identified for several classes of antibiotics, including aminoglycosides, macrolides, rifamycins, streptogramins, and phenicols. Aminoglycosides are known to inhibit protein synthesis by binding to the 30S ribosomal subunit.

Aminoglycoside-modifying enzymes (AMEs) can modify the hydroxyl or amino groups of the drug, which in turn reduces its affinity to the target. Different AMEs have been described, including acetyltransferases (AACs), phosphotransferases (APH), and nucleotidyltransferases (ANTs). AMEs could be encoded by the chromosome or mobile genetic elements, and are found in both Gram-positive and Gram-negative species, and in mycobacterial species¹³⁴. Examples of antibiotic degradation include β -lactamases which hydrolyse nearly all β -lactams, a class of antibiotics that target cell wall synthesis by binding to PBPs. β -Lactamases catalyse the hydrolysis of the amide bond of the β -lactam ring resulting in the drug's degradation. About 300 β -lactamases have been identified to date, and are classified into 2 main systems: either based on structural information (Ambler classification) or based on functional information (Bush–Jacoby–Medeiros classification). They are divided, on the basis of sequence, into four classes, among which three are serine hydrolases (serine β -lactamases; SBLs) for classes A, C, and D. The remaining class B contains unrelated group of zinc metalloenzymes (metallo- β -lactamases, or MBLs)¹³⁵.

III.2.3. Resistance through active efflux

Besides preventing drug entry, bacteria can actively expel antibiotics from the cell (Figure 3.2). This process is achieved by transmembrane proteins present in the cytoplasmic membrane called efflux pumps. So far, six families of pumps have been presented, including the ATP-binding cassette family (ABC) transporters which use ATP as a transport energy source. The five other transporters are powered by pumping ions out of the membrane, namely the resistance-nodulation-cell division (RND) family, the multidrug and toxin extrusion (MATE) family, the proteobacterial antimicrobial compound efflux (PACE) family, the small multidrug resistance (SMR) family, and lastly, the major facilitator superfamily (MFS) family¹³⁶. The MATE and MFS families are identified in Gram-positive bacteria, whereas the five pump families can be found in Gram-Negative Bacteria, with the RND family being the predominant pump in clinical isolates¹³⁷. Some efflux pump families are found to act independently in the inner membrane, whereas others (ABC and RND families) function in a likely coordinated way with other membrane compartments. RND pumps for instance form tripartite complexes with periplasmic adaptor proteins (PAPs) together with an outer membrane factor (OMF) that expand across the Gram-negative cell envelope^{131,136}. These machinery types constitute important mediators in Gram-negative bacteria resistance, including *E. coli*, *P. aeruginosa*, *N. gonorrhoeae*, and *Acinetobacter baumannii*. Their overexpression together with their ability to export a variety of structurally and chemically unrelated antibiotics (e.g. macrolides, and tetracyclines) significantly contributed to multidrug-resistant strains¹³¹.

III.2.4. Resistance through influx reduction and permeability control

Antibiotic entry relies on crossing the bacterial cell envelope by diffusion through the bilayer and also through porins for Gram-negative bacteria. Bacteria have evolved and acquired mechanisms that decrease antibiotic uptake, thus preventing its accumulation, and increasing its expulsion from the cell. Porins are β -barrel proteins that enable the influx of hydrophilic compounds (<600 Da), including antibiotics such as β -lactams and quinolones¹²⁰. Their expression is highly regulated in response to environmental stimuli, thus porins expression can be downregulated and limit these drugs' entry (Figure 3.2). This was described in *Klebsiella pneumoniae* isolates which presented a high-level ertapenem resistance due to a lack or altered expression of at least one of its major outer-membrane porins¹³⁸. Furthermore, hydrophobic antibiotics mainly diffuse through the bilayer. This is particularly convenient for Gram-negative bacteria that present a highly impermeable double-layered structure. Gram-positive bacteria lack the outer membrane making them more permeable to antibiotics. However, alteration of membrane structure and fluidity, through a change in lipid composition, can reduce antibiotic permeability. This was observed in *Enterococcus faecalis* clinical strains that developed resistance to daptomycin (DAP) following changes in membrane phospholipid profiles¹³⁹. Alternatives to modify antibiotic influx consist in the modification of the bacterial surface. Mycobacteria prevent antibiotic entry by producing extensive outer lipid layers and capsule-like coats of polysaccharides¹⁴⁰. In Gram-negative bacteria, LPSs play a key role in membrane permeability to hydrophobic drugs. Extensive efforts have been devoted for the development of drugs against LPS including the cationic peptides such as polymyxin, which is known as a membrane permeabilizers that binds to anionic LPS on its lipid A. This destabilization of the LPS layer leads to outer membrane permeabilization, allowing polymyxin to reach the inner membrane¹⁴¹.

III.3. LPS: a distinguishing target against bacterial resistance

LPS plays a key role in the Gram-negative bacteria permeation barrier. Targeting LPS biosynthesis and transport machinery presents a promising strategy for the development of effective drug agents against Gram-negative bacteria. As described in I.3.2, LPS biogenesis starts in the cytoplasm. Then LPS is translocated across the inner membrane, transported across the periplasm, to finally be inserted into the outer membrane's outer leaflet. A series of enzymes are involved in LPS biosynthesis and transport. Substantial efforts, from academic research and pharmaceutical companies, have been devoted to the discovery of molecules targeting LPS synthesis and transport. In this section, a summary of key LPS-directed inhibitors that have been described is provided (Table 1, 2).

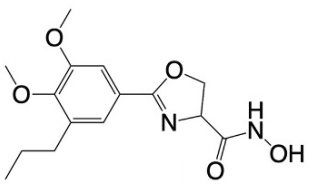
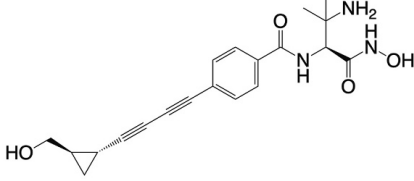
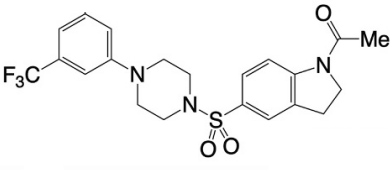
III.3.1. Targeting LPS biosynthesis

A large focus on the identification of inhibitors against LpxC family has been made. LpxC is a zinc metalloenzyme which catalyses the second and irreversible step in lipid A biogenesis (see Figure 1.4)²³. LpxC is considered a promising target for three reasons: it's a key enzyme in the Raetz pathway, it is conserved across multiple strains and species in Gram-negative bacteria enhancing its potential as a broad-spectrum target, and it lacks eukaryotic homologues enhancing its specificity as an antimicrobial agent target¹⁴². Back in the 1980s, Merck identified the first LpcX inhibitor termed L-573,655 by following its impact on LPS synthesis in *Salmonella* using a radiolabelled galactose incorporation assay¹⁴³. One LpxC inhibitor made it to the clinical trials, the small molecule ACHN-975. This molecule presented *in vivo* antimicrobial activity against *Klebsiella pneumoniae*¹⁴⁴. However, the trial was concluded and ACHN-975 was removed from the antibiotic discovery pipeline due to reported dose-limiting toxicity and inflammation at the infusion site in subjects during the Phase I clinical trial^{142,144}.

Similarly, LpxA and LpxD, two acyltransferases catalysing the first and third steps in the Raetz pathway (see Figure 1.4), respectively, were also targeted. Their early inhibitors were antibacterial peptide inhibitors discovered by phage display¹⁴². The pentapeptide, Peptide 920, bound to LpxA with a nanomolar binding affinity and displayed a competitive inhibition of the acyl chain donor in *E. coli*¹⁴⁵. Other peptide library screenings led to the identification of RJPXD33, a 12 amino acids peptide found to bind immobilized LpxD, and to also bind and attenuate the activity of LpxA with moderate affinity¹⁴⁶.

Other small molecule inhibitors blocking lipid A biosynthesis through LpxH inhibition were discovered, including the sulfonyl piperazine molecule referred to as Compound 1. This molecule inhibits LpxH in efflux-deficient *E. coli* strains, but revealed a lack of activity in the wild-type strain which limited progress of the project^{144,147}.

Table 1: Summary of LPS biogenesis inhibitors that have entered pre-clinical development.
Adapted from¹⁴⁴.

Molecule name	Structure	Mechanism of action	Stage of development
L-573,655		LpxC inhibition	Terminated due to narrow-spectrum activity
ACHN-975		LpxC inhibition in <i>K. pneumoniae</i> and other Gram-negative pathogens	Terminated after Phase I clinical trials due to toxicity (inflammation)
RJPXD33	TNLYMLPKWDIP	Binds and inhibits LpxA/LpxD in several Gram-negative bacteria	In pre-clinical development stages
Compound 1		Inhibits LpxH in efflux-deficient <i>E. coli</i>	In pre-clinical discovery stages

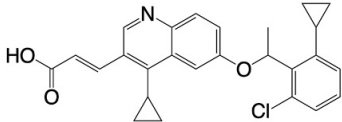
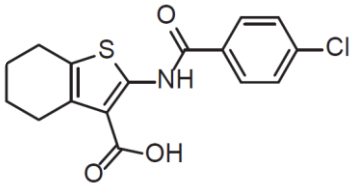
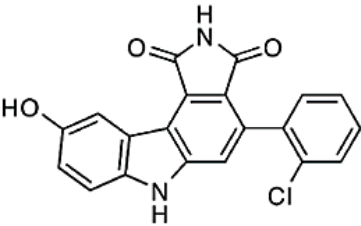
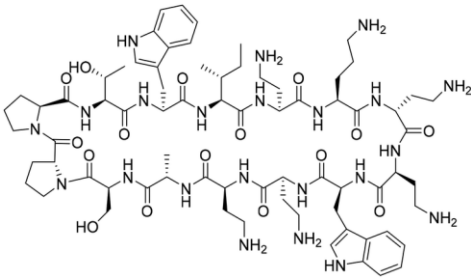
III.3.2. Targeting LPS transport

The development of LPS transport inhibitors has mainly focused on three proteins: the MsbA flippase, the ATPase LptB, and the OMP LptD (Table 2). *In vitro* screening of small molecule compounds led to the discovery of a quinolone-based compound against MsbA known as G592, which further was optimized resulting in product G907, a molecule with enhanced potency against MsbA¹⁴⁴. G907 was found to trap MsbA in an inward-facing LPS-bound conformation that blocked its ATPase activity and inhibited the flipping of core-lipid A across the inner membrane¹⁴⁸. However, G907 clinical development was terminated due to its significant binding to plasma proteins¹⁴⁴. Other inhibiting compounds have been later identified, including a molecule referred to as compound 2, which triggers the ATPase activity of MsbA but inhibits its flippase activity¹⁴⁹.

Going further in LPS transport machinery, LptB has also been subjected to drug inhibition. Screening of kinase inhibitors, mostly composed of ATP-competitive inhibitors, led to the discovery of molecules including compound 3 which was found to block the ATPase activity of LptB. However, this molecule only displayed potency against LptB in hyperpermeable *E. coli* strains but wasn't observed in the LptB₂FGC complex in whole bacterial cells¹⁵⁰. Furthermore, a small molecule IMB-88, recently discovered by yeast double hybrid assay, was found to specifically bind to LptA and block its interaction with LptC, thus exhibiting antibacterial properties¹⁵¹.

Lastly, LptDE complex, responsible for the last LPS transport step, was not spared by inhibitors targeting. The first set of LptD inhibitors was discovered after a screening of a library of peptidomimetic compounds based on the chemical structure of protegrin I, an antimicrobial peptide that disrupts membranes. This subsequently resulted in the discovery of murepavadin (POL7080), a cyclic peptide of 14-amino-acid, specifically interacting with LptD and blocking LPS passage across its lumen before it reaches its final destination, the outer leaflet of the outer membrane^{144,152}. POL7080 reached stage III clinical trials back in 2018 after showing efficient treatment of patients with ventilator-associated pneumonia during phase II clinical trials. However, due to toxicity, where 56 % of subjects suffered kidney damage, the trial was concluded in 2019, while only the inhaled form of POL7080 is still in pre-clinical development^{142,144}.

Table 2: Summary of LPS transport inhibitors in pre-clinical development. Adapted from ¹⁴⁴.

Molecule name	Structure	Mechanism of action	Stage of development
G907		Traps MsbA in inward-facing form and inhibits its flippase activity	Terminated due to significant binding to plasma proteins
Compound 2		Enhancement of MsbA ATPase activity and decoupling of protein ATPase activity from flippase activity	Terminated after Phase I clinical trials due to toxicity
Compound 3		Inhibition of ATPase activity of LptB in hyperpermeable <i>E. coli</i>	In pre-clinical development stages
Murepavadin (POL7080)		Interaction with LptD β -barrel and blockage of LPS passage to the outer leaflet	Terminated after Phase III clinical trials due to toxicity Inhaled form in pre-clinical development

III.3.3. LPS direct targeting with polymyxin antibiotics

The polymyxin antibiotics discovered back in 1947 are secondary metabolites produced by the bacterium *Bacillus polymyxa*. Up to date, up to fifteen varieties of polymyxins have been identified, of which polymyxin B and polymyxin E (colistin) are used clinically ^{144,153}. From a structural point of view, polymyxins are lipopeptides of approximately 1.2 kDa constituted of a typical ring and tail structures (Figure 3.3). The ring comprises a polycationic heptapeptide,

while the tail is made up of a tripeptide chain which in turn is linked to the N-terminus fatty acid¹⁵⁴. In the case of colistin, the fatty acid is either a 6-methyl-octanoic acid group (colistin A) or a 6-methyl-heptanoic acid group (colistin B) (Figure 3.3)¹⁴⁴.

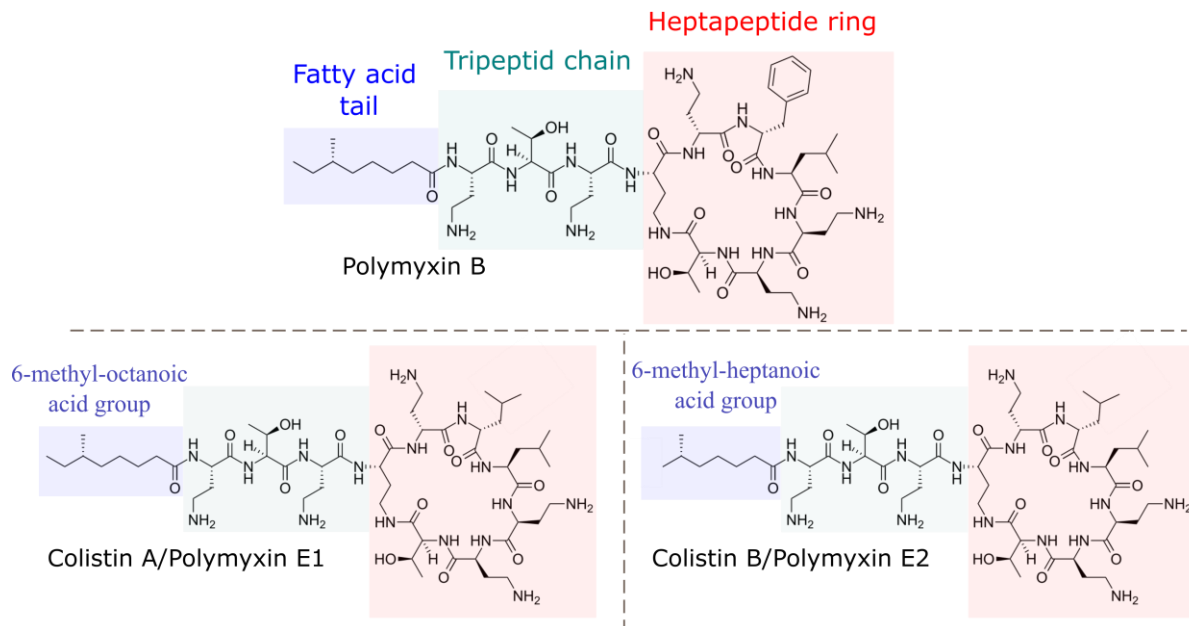


Figure 3.3: Chemical structures of polymyxin B and colistin (A and B). The heptapeptide ring is shown in red, the tripeptide chain in green, and the fatty acyl chain in blue. Adapted from¹⁴⁴.

Polymyxins target the outer membrane of Gram-negative bacteria and bind directly to LPS through electrostatic interactions. These interactions occur between the positively charged peptide ring of polymyxins on one side and the negatively charged phosphate groups of lipid A on the other side. Subsequently, membrane-stabilizing divalent cations (Ca^{2+} and Mg^{2+}) are displaced and the LPS layer is destabilized, resulting in outer membrane disruption. At this point, the permeability of the outer membrane is weakened allowing Polymyxin lipid tail to insert into the hydrophobic inner region of the membrane by interacting with LPS acyl chains and to trigger a total loss of outer membrane integrity^{144,154}. Colistin was found to also interact with LPS in the course of synthesis at the inner membrane. By doing so, it disrupts the membrane and leads to cell lysis¹⁵⁵. Furthermore, other antibacterial processes have been reported to take place after outer membrane disruption and to promote bacterial killing. These mechanisms include the inhibitory effect of vital respiratory enzymes, including NADH oxidoreductases at the bacterial inner membrane, generation of reactive oxygen species (ROS), and membrane fusion^{144,153}.

The clinical use of polymyxins is however limited by dose-limiting nephrotoxicity. However, as a “last-resort” antibiotic together with the emergence of resistant bacteria, a need for

understanding these resistance mechanisms is renewed. This would contribute to the development of polymyxin-derived molecules with improved safety profiles. Pathogenic bacteria can become resistant to polymyxins through two main routes: mutations increasing expression of LPS-modifying enzymes (such as PhoPQ/PmrAB), or acquisition of a plasmid harbouring a mobile colistin resistance gene (*mcr*)^{144,153}. In both mechanisms, the resistance is acquired by modifications of LPS net negative charge with cationic groups, thereby diminishing the electrostatic interactions with cationic polymyxins. The resistance conferred by mutation results in glucosamines modification following the expression of ArnT and EptA enzymes. ArnT mediates the transfer of L-Ara4N to lipid A at position 4'-phosphate, whereas EptA transfers a PEtN at position 1'-phosphate or to both 1'- and 4'-phosphates, in the absence of L-Ara4N^{25,34}. On the other hand, *mcr* genes, first described in *E. coli* back in 2015, encode a membrane-associated enzyme responsible for the decoration of lipid A phosphate groups with a PEtN¹⁵⁶. In line with this approach, polymyxin-resistant mutants isolated from *S. typhimurium* and *E. coli* are found to carry modifications in LPS structure by substitution of lipid A phosphates with PEtN¹⁵⁷.

Extensive efforts to improve polymyxins safety and efficiency are made. Synthesised polymyxin-derivative molecules have already been developed and some have entered clinical trials. The synthetic lipopeptide F365 (QPX9003) is developed by Qpex Biopharma includes modifications at multiple positions throughout the polymyxin scaffold. This molecule made it to phase I clinical trials with excellent *in vivo* efficiency and superior safety profiles against various lung pathogens, including *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Klebsiella pneumoniae*¹⁵⁸. Another polymyxin derivative released by Spero therapeutics, SPR206, concluded phase I clinical trials and showed impressive *in vitro* and *in vivo* activities against *A. baumannii*, *P. aeruginosa*, and other multidrug- and extensively drug-resistant strains¹⁵⁹. This compound is dedicated for intravenous administration for the treatment of lung, blood, and urinary tract infections caused by resistant Gram-negative bacteria pathogens¹⁵⁹. Finally, a class of antibiotics derived from polymyxin and murepavadin scaffolds have been developed. These chimeric antibiotics were found to target outer membrane biogenesis through binding to both LPS and the BamA machinery¹⁶⁰. Optimized derivatives demonstrated potent activity against multidrug-resistant pathogens, with the lead candidate already in the pre-clinical toxicology process¹⁶⁰.

IV. State-of-the-art approaches for LPS studies

IV.1. LPS structural characterization

IV.1.1. LPS extraction and fractionation methodologies

IV.1.1.1. LPS isolation and detection

Given the amphiphilic nature of LPS molecules, its compositional and structural characterization has often required prior isolation and fractionation steps of its different composing entities. Furthermore, LPS and LOS structures require different isolation protocols. The O-antigen portion confers higher solubility to LPS molecules which are therefore extracted into aqueous solutions. Whereas, LOS molecules lacking the O-antigen, exhibit a higher lipophilic property and are thereby extracted in organic solutions. Different extraction methods have been implemented since the discovery of LPS back in 1941, including the most widely used ones the hot phenol-water extraction, and the phenol/chloroform/light petroleum (PCP) method ¹⁶¹. The former method is efficient for LPS ¹⁶², whereas the latter is used for LOS extractions ¹⁶³. PCP method operates without cell lysis unlike hot phenol-water extraction, which then gives good LPS yields but foresees the coextraction of contaminating nucleic acids and proteins. Hence there is a need for further steps to remove contaminants (combination of enzymatic treatments, ultracentrifugation, and chromatography) ^{21,161,164}.

LPS molecules are later subjected to rapid analytical techniques to assess the composition of LPS structures. The most commonly used method is sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE). The latter allows the detection of alterations and heterogeneities within LPS structures only by assessing the electrophoretic band profiles ¹⁶¹. Different staining methods have been implemented for LPS detection, including silver staining, Zinc-Imidazole Stain, the Pro-Q Emerald 300 dye staining, and the immunoblotting methods ¹⁶⁵. In the case of silver staining, the LPS sugar reacts with silver nitrate yielding oxidation of its hexoses and formation of silver (black). With this method, less than 1 µg of LPS is sufficient for visualization in polyacrylamide gel ¹⁶¹. Given the drawbacks reported for silver stain, including toxicity and cost, a new sensitive technique has been developed, the Zinc-Imidazole Stain. This method allows the detection of LPS as transparent and colourless bands, which can be recovered from gel slices for further studies ^{166,167}. The specific Pro-Q Emerald 300 dye staining on the other hand links to LPS and glycoproteins. The sugar moieties are initially oxidized into aldehydes by the action of periodic acid. Aldehydes then interact with the dye and generate fluorescent conjugates. More importantly, only a few nanograms of LPS

are sufficient for detection using this staining method against 250–1000 ng for conventional silver staining¹⁶⁸. Lastly, the immunoblotting stain can be used to detect LPS. It exclusively relies on the use of O-antigen-specific antibodies, which are revealed through alkaline phosphatase-conjugated secondary antibody¹⁶¹.

IV.1.1.2. LPS fractionation

LPS fractionation requires different chemical treatments. LPS delipidation can be achieved by mild acid hydrolysis, with 1% acetic acid or sodium acetate pH 4.4-4.5^{161,169}. This approach selectively cleaves the lipid A-polysaccharide bond, by cleaving the linkage between the Kdo moiety and the non-reducing GlcN of lipid A (Figure 4.1). The reaction releases intact lipid A and polysaccharide parts, which are then separated using appropriate solvents, chromatographically purified, and lastly characterized with the method of choice. It was later found that following this hydrolysis, Kdo could be replaced by the D-glycero- α -D-talo-oct-2-ulopyranosidonic acid (Ko) derivative, leading to the use of higher acidic concentrations for a better lipid A removal²¹. However, this hydrolysis could result in partial degradation of the carbohydrate portion of the LPS, or partial dephosphorylation and O-deacylation of lipid A^{21,161}. Finally, Lipid A can further be extracted following the Bligh and Dyer extraction method using chloroform/methanol/water¹⁷⁰, or by triethylamine citrate (TEA-citrate) hydrolysis¹⁶⁹. The latter extraction method is tailored for micro-scale samples, especially those designed for mass spectrometry (MS) experiments. The TEA-citrate reagent does not require removal prior to analysis by MS, as it enhances mass spectra while preserving labile LPS constituents¹⁶⁹.

On the other hand, the polysaccharide portion can also be released following coupled O-deacylation and N-deacylation reactions (Figure 4.1)¹⁷¹. The O-deacylation occurs through anhydrous hydrazinolysis and removes the ester-linked fatty acids, while the following N-deacylation (using hot KOH) removes the amide-linked fatty acids of lipid A. This double alkaline treatment results in the recovery of both polysaccharide and acyl chains of lipid A which could further be separated using adequate solvents (water/chloroform). Importantly, the recovered phosphorylated polysaccharide products are not destroyed and can be subjected to further characterization studies (e.g. compositional and sequence analysis). However, phosphodiester, pyrophosphate, diphosphodiester, acetyl, and carbamoyl groups are cleaved under hydrazine/KOH treatment¹⁷¹.

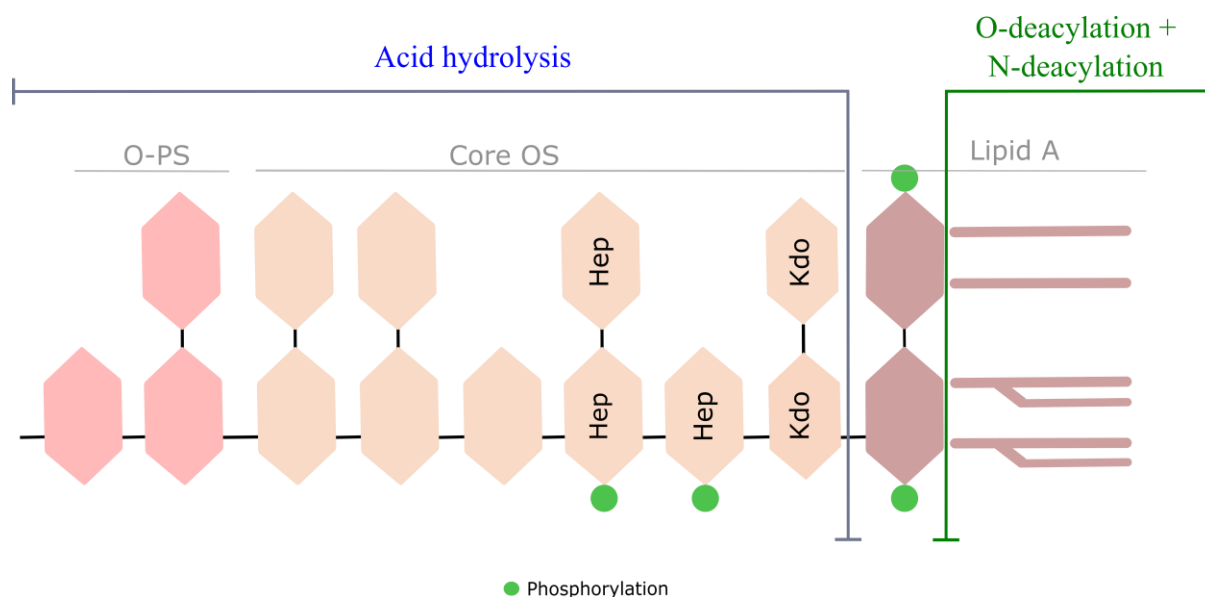


Figure 4.1: Illustration of the cleavage sites of LPS deacylation reactions. The acid hydrolysis cleaves the lipid A-polysaccharide bond. Whereas the coupled O-deacylation and N-deacylation reactions release the polysaccharide portion and N- and O-linked fatty acids.

IV.1.2. LPS compositional analysis

Gas chromatography (GC) was widely used in monosaccharide and fatty acids quantification. Coupled with mass spectrometry (GC-MS), the composition, branching points, and absolute configuration of the carbohydrates can be determined. Given the fact that lipid A fatty acyl chains and monosaccharides are not volatile, a chemical derivatization is required prior to GC-MS analysis. Fatty acids compositional analysis can be done using GC-MS. For that perspective, total ester- and amide-linked fatty acids can be directly released from LPS, as methyl ester derivatives, by acidic methanolysis¹⁷². LPS treatment with anhydrous hydrochloric methanol effectively releases both entire ester- and amide-bound total acyl chains from LPS. In these conditions, no additional esterification reaction is needed before GC-MS analysis. The analysis of the retention times of released methyl esters with those of commercially available standards, combined with analysis of their fragmentation pattern, enables the determination of fatty acids composition¹⁶⁴.

The composition of LPS in monosaccharides can be obtained with the analysis of acetylated O-methyl glycosides. These derivatives are formed following the cleavage of the glycosidic linkages with anhydrous methanol (HCl/MeOH) resulting in O-methyl glycoside monosaccharides. Following up, a peracetylation reaction with acetic anhydride in pyridine (Ac₂O/Pyr) is performed forming acetylated methyl glycosides which can be analysed by means of GC-MS. The analysis of the fragmentation pattern (m/z) of each sugar aided by

standards provides the monosaccharide type ¹⁶⁴. Alternatively, monosaccharides can be analysed by GC or GC-MS as alditol acetates. These are obtained following hydrolysis of the glycosidic linkages, reduction with sodium borohydride, and peracetylation ¹⁷³.

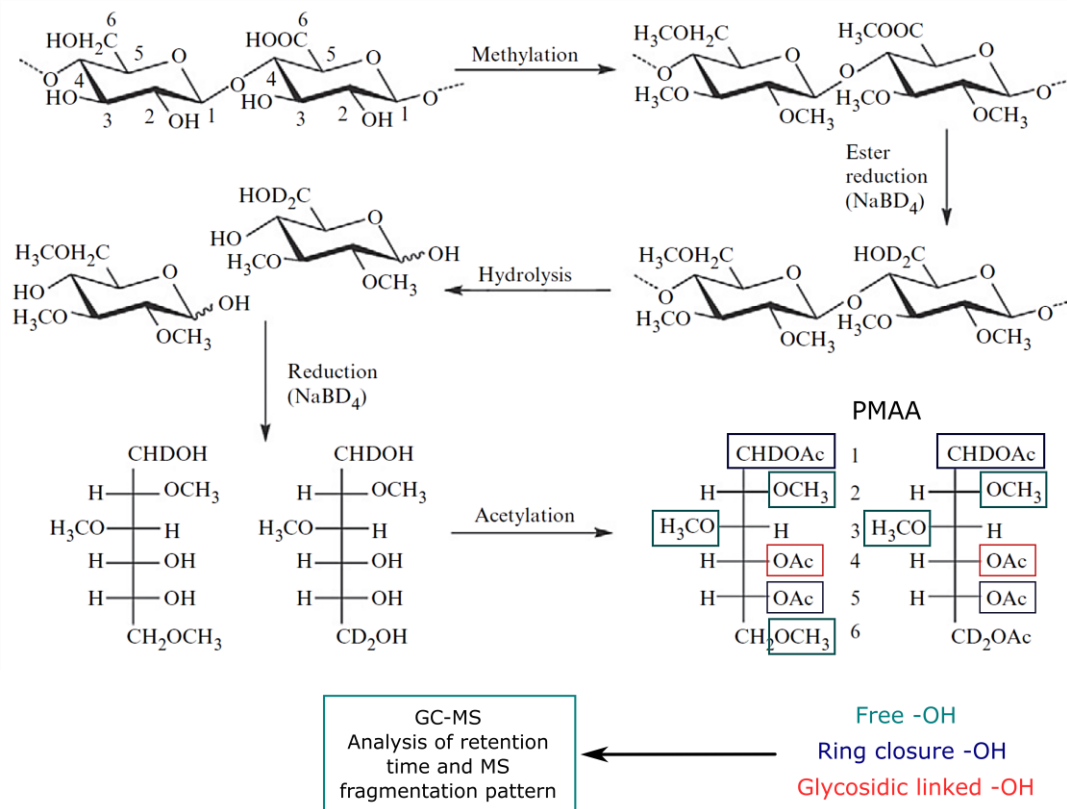


Figure 4.2: Derivatization protocol used for LPS monosaccharide branching determination. Carbon atoms are numbered throughout the protocol. By the end of the reaction, free hydroxyl functions form O-methyl groups (C2, C3, and C6 shown in green). O-acetyl groups represent the positions involved in ring closure (C1 and C5 shown in blue), and the branching position (C4 shown in red). Adapted from ¹⁶⁴.

Furthermore, the glycosidic linkage can be derived using Partially Methylated Acetylated Alditols (PMAA) derivatives ^{164,174}. This reaction consists first of methylation of free hydroxyl groups into methyl ethers in an etherification reaction using methyl iodide in anhydrous dimethyl sulfoxide with NaOH powder. Secondly, a reduction reaction follows with deuterated sodium borohydride (NaBD₄) and transforms the methyl ester function into a hydroxymethyl group with two deuterium atoms. Afterwards, the glycosidic linkages are hydrolysed resulting in the release of partially methylated monosaccharides, followed by a reduction of the anomeric position using NaBD₄. Finally, the acetylation of the remaining free hydroxyl groups (involved in ring-closing and glycosidic linkage) forms the so-called PMAA (Figure 4.2), which provides information on the hydroxyl groups involved in the glycosidic linkage through GC-MS ¹⁷⁴.

Each monosaccharide derivative unit contains two acetyl functions involved in ring closure, at positions 1 and 5 for aldopyranose sugars. The additional acetyl group on the other hand indicates the position involved in the glycosidic linkage, since the other hydroxyl groups are already methylated.

None of the previously discussed methods can ascertain the absolute configuration of the detected monosaccharides. The D and L enantiomers of methylglycosides cannot be resolved by GC, and are eluted as a single peak. This limitation can be overcome through chemical derivatization using chiral alcohol, including 2-octanol or 2-butanol¹⁷⁵. The acetylated 2-octyl or 2-butyl glycosides can be resolved under the GC-MS chromatographic conditions. One of the approaches consist in using racemic alcohol (e.g. 2-butanol) and an enantiopure monosaccharide to produce two diastereoisomers which are then used as standards. Using a D-glucose (D-Glc) and racemic 2-butanol, the diastereoisomers D-Glc-(R)-butanol and D-Glc-(S)-butanol are produced. They have the same chromatographic behaviour as their enantiomers, L-Glc-(S)-butanol and L-Glc-(R)-butanol, respectively. Analysis and comparison of the chromatograms obtained for the standards and the sample, derived with an enantiopure 2-butanol, allows the determination of its absolute configuration (Figure 4.3).

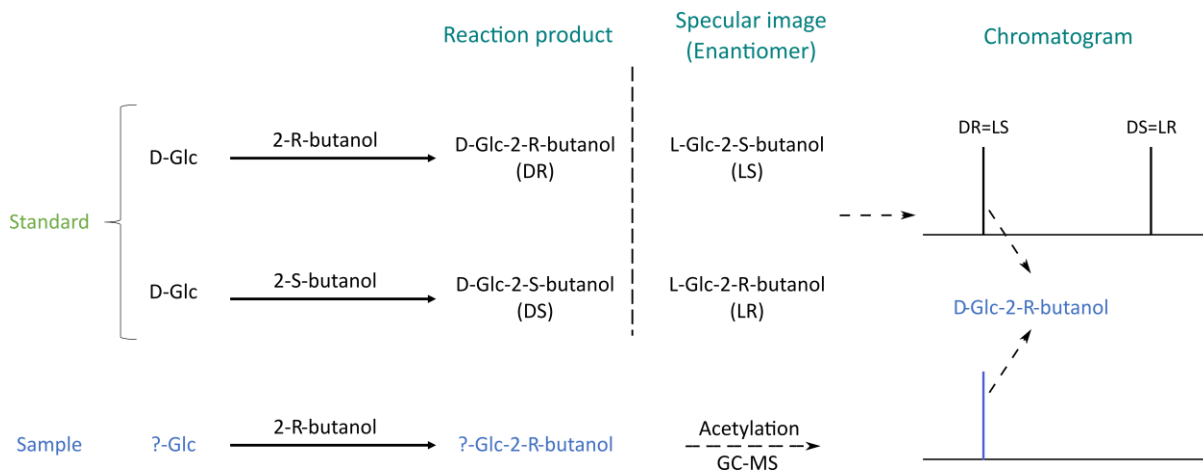


Figure 4.3: Strategy used for the determination of absolute configuration of LPS monosaccharides. Preparation of 2-butylglycoside standards using as an example D-Glc with racemic 2-butanol produces a mixture of diastereoisomers. The retention time of these compounds, which is the same as for their corresponding enantiomers, allows to determine the absolute configuration of the monosaccharide of interest (?-Glc). When the latter is prepared using an enantiopure 2-butanol, its configuration can be deduced by a simple comparison of its retention time with those of the standards. Adapted from¹⁶⁴.

IV.1.3. LPS structure elucidation

IV.1.3.1. Mass spectrometry

Mass spectrometry (MS) is a key and complementary technique widely employed in LPS structure elucidation. The use of soft ionization techniques, including matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI), together with tandem MS is typically used for the complete assignment of the lipid A portion, including the acylation and phosphorylation patterns, as well as potential substitutions in glucosamine residues¹⁷⁶. MALDI has been widely used for lipid A analysis, starting with the analysis of methylated lipid A in the positive mode¹⁷⁷, then underivatized lipid A in the negative mode¹⁷⁸, to finally be implemented for structural characterization of intact endotoxins (especially LOS forms)¹⁷⁶. MALDI-MS experiments, usually carried out in the negative mode for LPS, generate single-charged ions resulting in a relatively easy interpretation of the spectra. The improvements made in both sample and matrix preparations resulted in the acquisition of high-quality MALDI mass spectra of intact LOS without prior chemical treatments enabling the simultaneous observation of lipid A and core OS moieties (Figure 4.4)¹⁷⁹. The obtained data combined with the ones recorded on pre-treated samples (e.g. O-deacylated LOS) (Figure 4.4) facilitates the interpretation of the complex intact LOS spectra and enables easier structure determination of both core OS moiety and lipid A. This method has been applied to LOS from *Xanthomonas campestris*, *Pseudoalteromonas issachenkoni*, and *Shewanella pacifica*¹⁷⁹. On the other hand, the use of MS in intact smooth LPS analysis has been documented in only a limited number of cases. It was reported for the enteroinvasive *E. coli* O164 strain and resulted in a single peak at m/z 24.7 kDa in MALDI-MS. Given the established repeating unit mass of ~ 1 kDa, the O-antigen part was estimated to contain 24 repeating units¹⁸⁰.

MALDI- and ESI-MS-based techniques are also valuable for the analysis of core OS derived from various bacterial strains¹⁸¹. With the implementation of powerful mass analysers, such as ion-trap, time of flight (TOF), and Fourier transform ion cyclotron resonance (FT-ICR), MS stands-out as one of the leading approaches in the elucidation of LPS structures. Lastly, the combination of separation methods, including thin layer chromatography (TLC), and GC-MS allows for accurate structural analysis of purified fractions pertaining to LPS components, including the O-antigen chain, core OS, and lipid A¹⁷⁶.

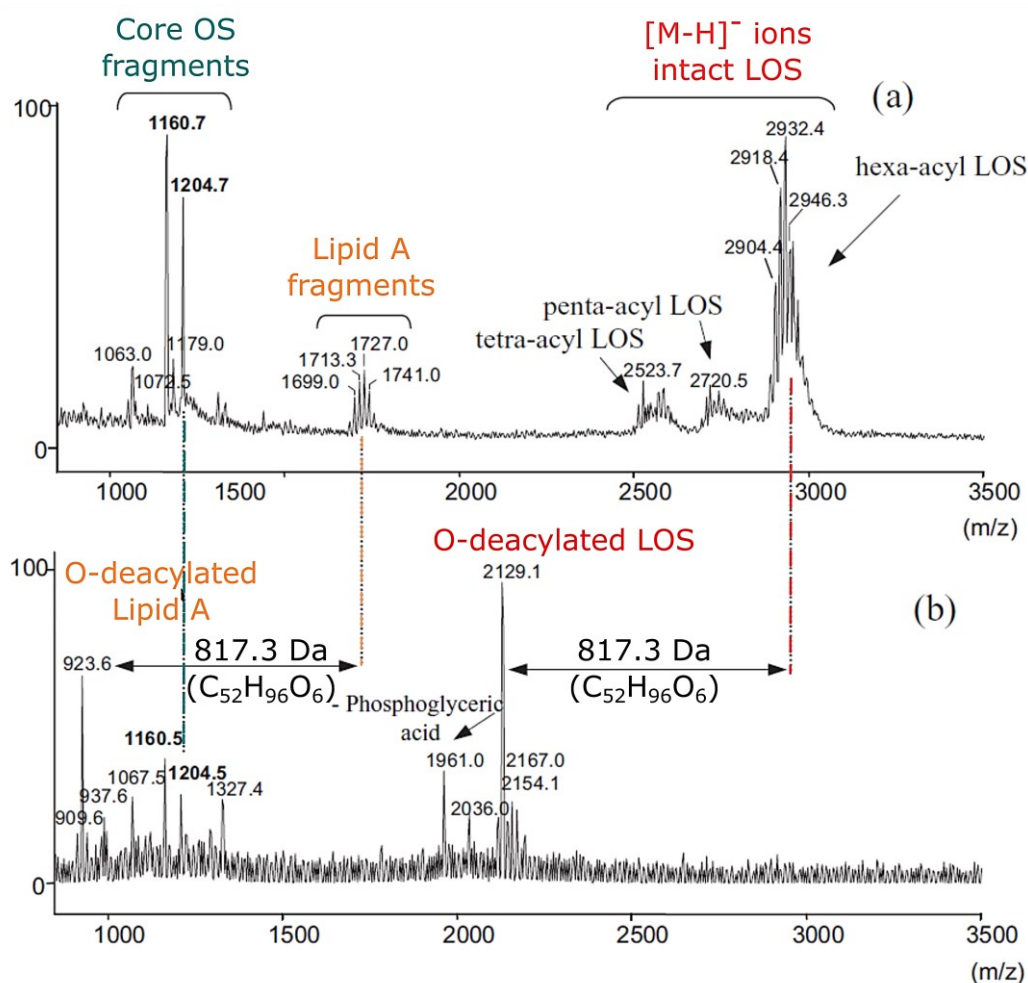


Figure 4.4: MALDI-MS spectra of (a) intact and (b) O-deacylated LOS of *Shewanella pacifica*. (a) The negative ion MALDI-MS spectrum of native LOS revealed three main fragment regions. A high molecular weight region ion corresponding to intact LOS (with some heterogeneities in the fatty acyl chains), ions corresponding to lipid A fragments, and core OS fragments. (b) In the O-deacylated LOS spectrum, fragments relative to core OS retained the same m/z ratio (m/z 1204.5 and 1160.5), while ions deriving from the O-deacylated lipid A and from the O-deacylated native LPS differ by 817.3 Da. Adapted from ¹⁷⁹.

IV.1.3.2. Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) is a leading technique for LPS complete structural elucidation without prior derivatization. For a straightforward liquid state NMR analysis, again organic chemistry is at the basis of most LPS analysis. Given the LPS amphiphilic nature, besides the use of detergent or organic solvents for LPS solubilization, the removal of lipid A is a prerequisite for solution NMR core OS analysis. A full assignment of core OS can then be completed using a combination of uni- and bi-dimensional NMR experiments (Figure 4.5).

Starting with 1D ^1H and ^{13}C spectra, one can assess the sample quality, get insight into the number of anomeric sugars, the presence of structural motifs, and further determine sugar's anomeric configuration based on their chemical shifts. The anomeric proton resonances are mainly found in the chemical shift range of 4.4-5.5 ppm, while the ring proton resonances are found in the 3-4.2 ppm range¹⁸². α -configured pyranose residues show a ^1H chemical shift at 4.9-5.6 ppm and ^{13}C chemical shifts at 95-101 ppm, whereas β -anomers exhibit a ^1H chemical shift at 4.3-4.7 ppm and ^{13}C chemical shifts at 103-105 ppm.

Bi-dimensional NMR experiments, on the other hand, are used to decipher atom connectivity. A bunch of homonuclear ($^1\text{H}/^1\text{H}$) and heteronuclear correlation experiments are routinely used. This includes correlation spectroscopy (COSY) and total correlation spectroscopy (TOCSY) for ^1H - ^1H correlations, and heteronuclear single-quantum coherence (HSQC) and heteronuclear multiple-bond correlation (HMBC) for ^1H - ^{13}C correlations. COSY and TOCSY experiments are used to determine the number and type of sugar residues through the detection of off-diagonal correlations between coupled spins for COSY (i.e. coupling of H1 with H2), and detection of off-diagonal correlations between all the spins in each residue for TOCSY (e.g. coupling of H1 with H2, H3, H4...). HSQC ^1H - ^{13}C experiment on the other hand allows the association of each ^{13}C chemical shift to its attached proton, whereas HMBC allows the detection of scalar couplings between ^1H and ^{13}C connected through the glycosidic linkage yielding to connectivity between the different sugar rings. Recording of dipolar coupling-based experiments, such as Nuclear Overhauser Effect Spectroscopy (NOESY) and Rotational Overhauser Enhancement Spectroscopy (ROESY) experiments, provide additional information regarding the OS primary sequence and sugar ring configuration. These experiments correlate protons that are close in space, and provide intra- and inter-residual NOE contacts which can complement the connectivity deduced from HMBC spectra.

The sugar configuration, in particular at the anomeric position, can be obtained through the measurement of J-coupling constants. For pyranosidic residues, a J-coupling constant $^1J_{\text{C1,H1}}$ of ~ 170 Hz is characteristic of α -anomers, whereas $^1J_{\text{C1,H1}}$ of ~ 160 Hz indicates β -anomeric sugar configuration¹⁸². The vicinal coupling constants $^3J_{\text{H,H}}$ also indicate relative proton orientations offering key configurational information. For instance, in pyranose structures, $^3J_{\text{H,H}}$ coupling constants are large (7-8 Hz) in cases where the two protons are in an axial configuration. In contrast, $^3J_{\text{H,H}}$ values of ~ 4 Hz and < 2 Hz reveal an equatorial-axial and an axial-equatorial or equatorial-equatorial configuration, respectively¹⁸².

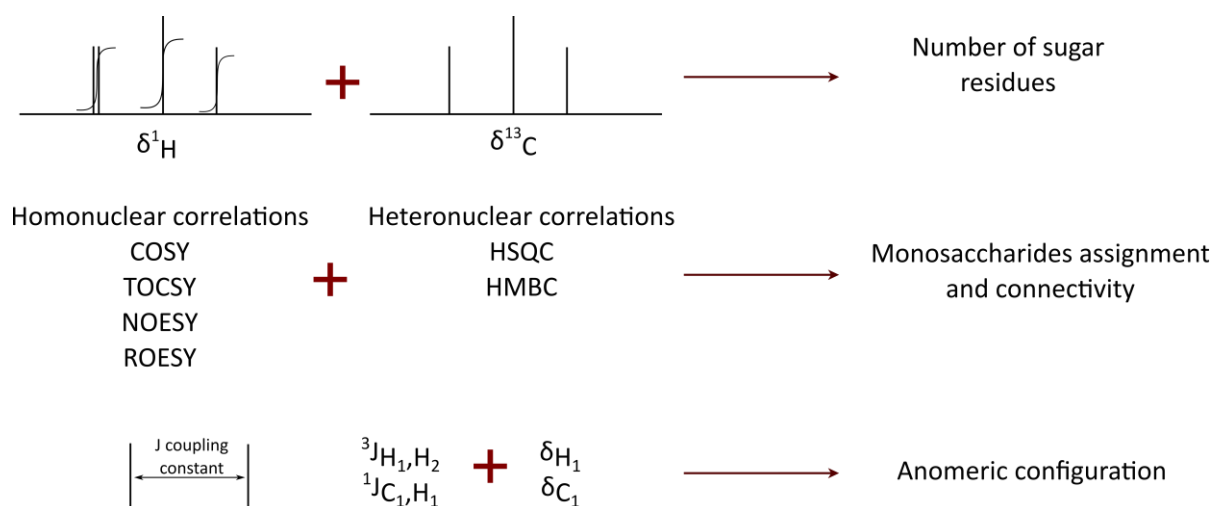


Figure 4.5: Summary of the described NMR approaches employed for core OS assignment.
 Adapted from ¹⁸².

When it comes to NMR analysis of native LPS molecules, solution NMR is no longer doable. LPS molecules form heterogeneous structures in solution depending on the glycan moiety size. LPS generally forms elongated micellar structures, while LOS assembles as vesicles ¹⁸³. Those objects are too large for solution NMR due to the severe signal broadening associated with slow rotational motion, but are well suited for Magic Angle Spinning (MAS) NMR as those effects are effectively averaged-out. In solid-state NMR experiments (ssNMR), the sample is packed into a rotor, placed into the NMR probe, and spun at the magic angle (54.74°) with respect to the static magnetic field. Following the MAS rate, anisotropic interactions can be averaged out, partially or completely (at very fast MAS) resulting in narrow lines ¹⁸⁴.

Similarly to solution NMR, a set of homonuclear and heteronuclear experiments are implemented in ssNMR, including Dipolar Assisted Rotational Resonance (DARR) and Proton Driven Spin Diffusion (PDS) exploiting dipolar based-couplings. These experiments rely on magnetization transfer from ^1H to ^{13}C nuclei, which in turn transfers it to other ^{13}C nuclei which are close in space. The obtained ^{13}C - ^{13}C cross peaks reflecting intra- and inter-residue contacts, when using long mixing times, allow for molecule assignment. This strategy is used for 3.2 mm rotors and typical MAS spinning rates of 15 KHz. Moreover, typical ^1H - ^{13}C heteronuclear ssNMR experiments relying on cross-polarization (CP), exploiting dipolar based-couplings, and insensitive nuclei enhanced by polarization transfer (INEPT), exploiting scalar couplings, are also used. They do not only associate each ^{13}C chemical shift to its attached proton, but are

also utilized to identify rigid and mobile molecular parts in the studied system by CP and INEPT-based experiments, respectively. In addition to complete structure assignment, molecular interactions with LPS are also reported. It was reported for two LPSs with different O-Antigen portions from *P. aeruginosa* interaction studies with a natural antibiotic from the aminoglycosides produced by Gram-positive bacteria *Micromonospora echinospora*, gentamicin¹⁸³. This study revealed that only one specific O-antigen portion is involved in the antibiotic binding by interpretation of signal intensities upon gentamicin binding (Figure 4.6).

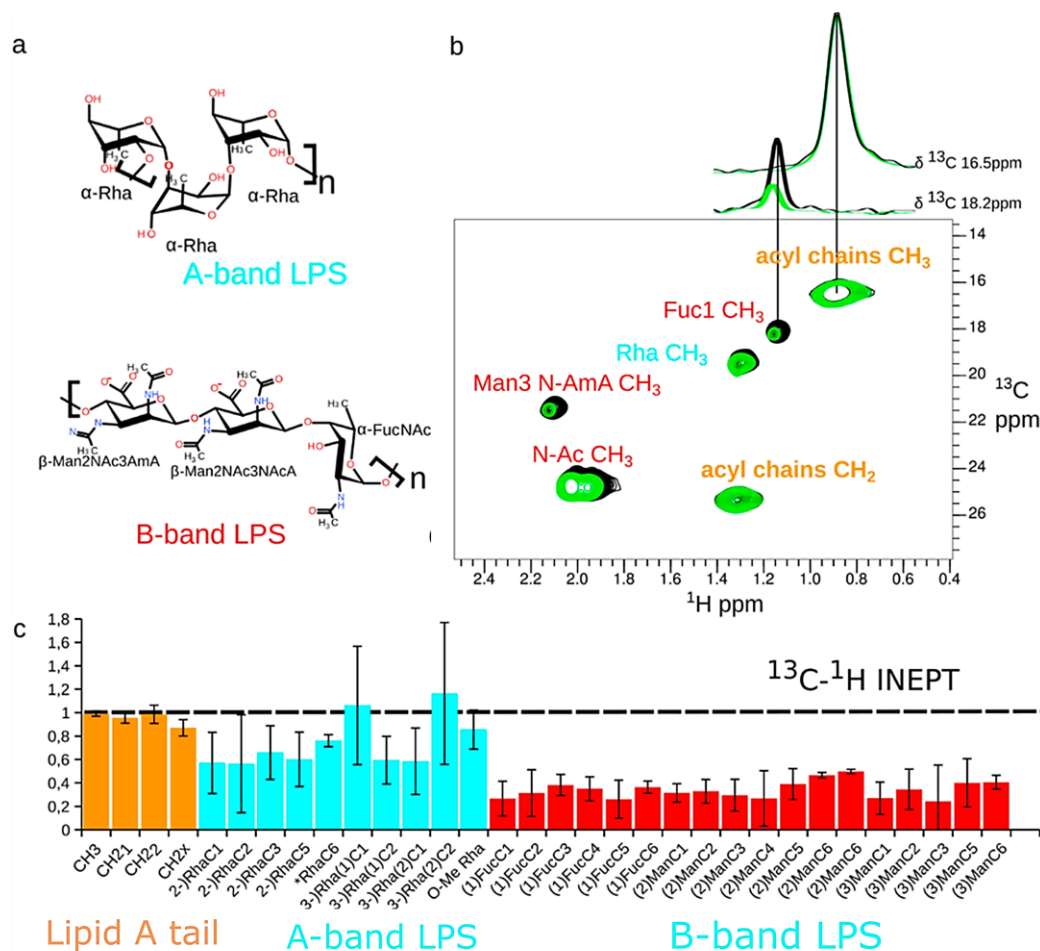


Figure 4.6: *P. aeruginosa* LPS O-antigen interaction with gentamicin. (a) Chemical structure of A-band and B-band O-antigen repeating units. (b) ¹H-¹³C INEPT ssNMR spectrum of *P. aeruginosa* LPS in the absence (black) and presence (green) of gentamicin with 1D slices of lipid A methyl and fucose methyl peaks shown above the spectrum. (c) Ratio of peak intensities ($I_{gentamicin}/I_0$) of each LPS portion measured in the ¹H-¹³C INEPT spectra showing the effect of gentamicin binding on the resonance intensities of the different LPS portions. O-antigen B-band represents a stronger gentamicin effect, and can be reflecting an interaction of gentamicin with this O-antigen. Adapted from¹⁸³.

IV.2. LPS-lectin receptor interaction studies

IV.2.1. Glycan arrays for LPS-lectin interaction screening

C-type lectin receptors family have been reported to recognise carbohydrate motifs presented on the surface of microorganisms. Based on the recognition occurring, that governs the discrimination of self- from non-self-molecular motifs, the immune response is then launched towards either activation or tolerance in case of pathogenicity or beneficial interactions, respectively ¹⁸⁵. Glycan arrays have been an important approach for the screening of glycan-binding receptors. Immobilized glycan arrays of microbial glycans, lipopolysaccharides or their synthetic versions, have been documented for interaction characterisation involving Gram-negative bacteria glycans with host immune system receptors and antibodies ^{186–188}.

Binding specificities of immune lectin receptor DC-SIGN were analysed against a glycan array comprising 140 compounds, including 12 bacterial LPS inner core glycans ¹⁸⁶. This study revealed specific recognition by DC-SIGN of several LPS inner core glycans through a terminal L-glycero-D-mannoheptose (Hep) moiety. Furthermore, in another study, human and murine Langerin binding to bacterial glycans were tested against a glycan array comprising over 300 bacterial carbohydrates ¹⁸⁸. It was found that the two homologues presented different recognition profiles towards the tested bacterial glycans. Although their binding to simple ligands, such as monovalent mannose, was found to be nearly identical, pronounced differences were found for complex bacterial glycan structures. While the murine Langerin binds to a large subset of polysaccharides of different structures originating from a wide range of bacterial species, the human Langerin only recognized few glycan structures originating from *Shigella* antigens, and derived from *Y. pestis* and *E. coli* strains ¹⁸⁸. These differences can be attributed to the slight differences in residues adjacent to the binding site between the two homologues which can alter the complex polysaccharide structures specificity. Moreover, it is noteworthy to also keep in mind that the composition in polysaccharides is not the only determinant of the binding, but their spatial arrangement within the complex structure and the context in which they are presented are also of paramount importance in this process.

Lastly, another example of a glycan array containing 60 chemically synthesized mycobacterial glycans representing the major carbohydrate structures presented on the cell surface of *Mycobacterium tuberculosis* and other mycobacteria was reported ¹⁸⁹. A subset of glycan binding receptors, including DC-SIGN were analysed and found to interact with mannose-containing caps found on the mycobacterial lipoarabinomannan (LAM), and

phosphatidyl-myo-inositol mannosides (PIMs) (Figure 4.7). Additionally, internalization mediating receptors were also found to recognise distinct sets of glycans, including mannose-binding endocytic receptors DC-SIGN and MMR which bind glycans derived from LAM and PIMs structures ¹⁸⁹.

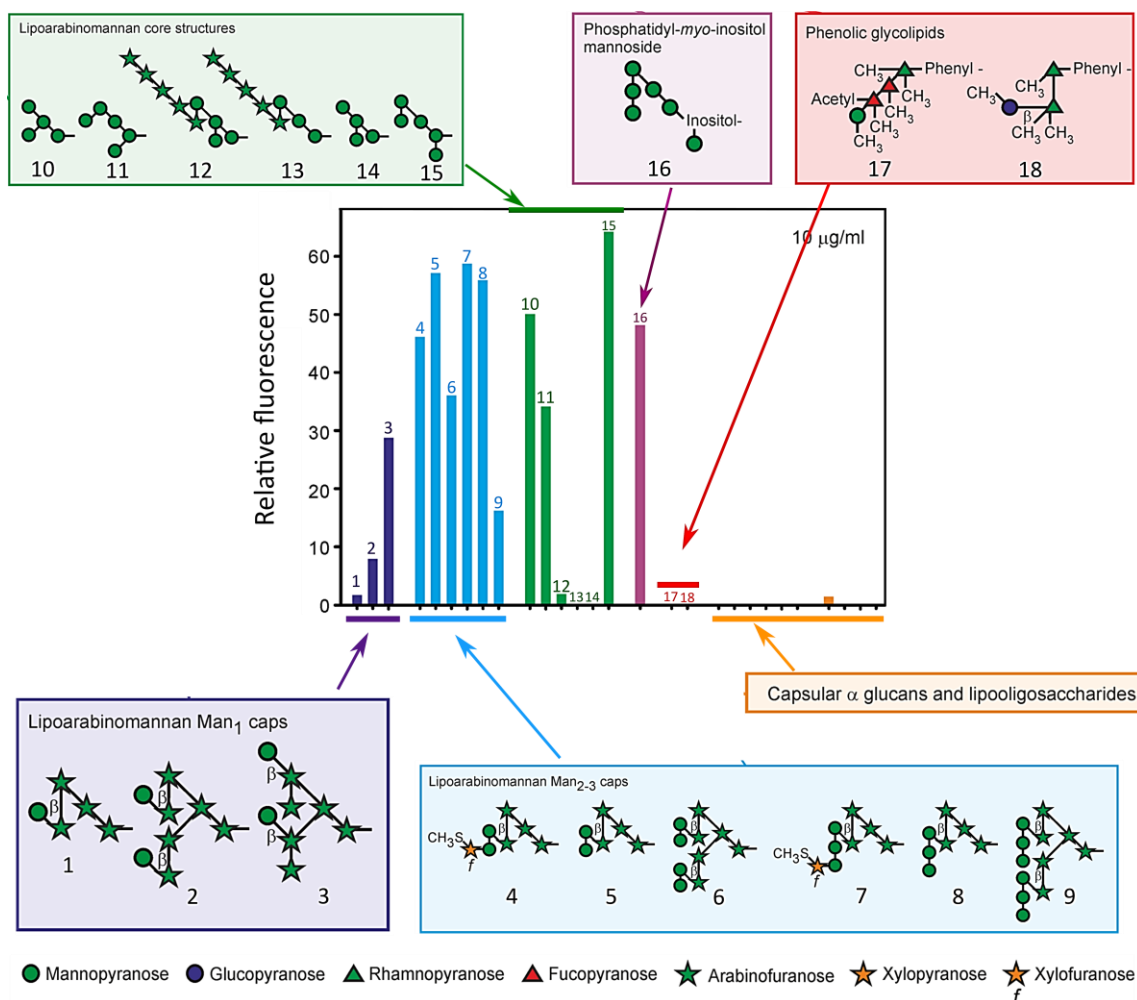


Figure 4.7: Interaction of the extracellular domain of DC-SIGN to mycobacterial glycans. DC-SIGN labelled with Alexa fluor 555 was incubated with the mycobacterial glycan array at different concentrations (the one shown in here is at 10 µg/ml). Bound DC-SIGN was afterwards detected by fluorescence, and interacting mycobacterial glycans are highlighted. The Symbol representation of the monosaccharide units is provided at the bottom. Adapted from ¹⁸⁹.

IV.2.2. Lectin arrays for host–microorganism interaction screening

Glycan arrays have been widely used to screen the specificity of lectins and other glycan-binding proteins to a large variety of immobilized glycans. Nevertheless, when it comes to studying lectin-pathogen interactions, the limited availability of pure and well-characterized glycans from microorganisms is a constraint for glycan array utilization. In that perspective, the development of a lectin array offers the opportunity to simultaneously screen a variety of lectins with a wide range of labelled, or intact microorganisms. Such a system was reported by Jégouzo et al. where a mammalian lectin array was developed using CRDs from bovine C-type lectins¹⁹⁰. The lectin array was constructed using biotin-tagged CRD fragments immobilized on streptavidin-coated wells resulting in a defined CRD orientation where the sugar-binding sites are pointed outward the wells. CRDs were expressed in *E. coli*, refolded, and purified by affinity chromatography on immobilized sugars, ensuring the proteins correct folding and sugar-binding functionality.

The binding screen was performed on multiple strains of both Gram-positive and Gram-negative bacteria, including two pathogenic *E. coli* strains: the enteropathogenic *E. coli* strain E2348/69 (O127:H6), and the enterohemorrhagic *E. coli* strain EDL933 (O157:H7). The observed binding profiles of lectins with the glycan residues of these strains are somehow consistent with what is known about lectins' interactions with defined glycan motifs. In other words, *E. coli* O127:H6 strain harbouring galactose and GalNAc repeating units in its O127 polysaccharide showed strong binding to most of the galactose-specific receptors, including MGL. In contrast, *E. coli* O157:H7, with its rhamnosamine, GalNAc, glucose, and fucose polysaccharides bound only to few receptors, even not to GalNAc-specific receptors, presumably due to the limited exposure of target sugar units (such as GalNAc, glucose, and fucose) to the tested lectins (Figure 4.8).

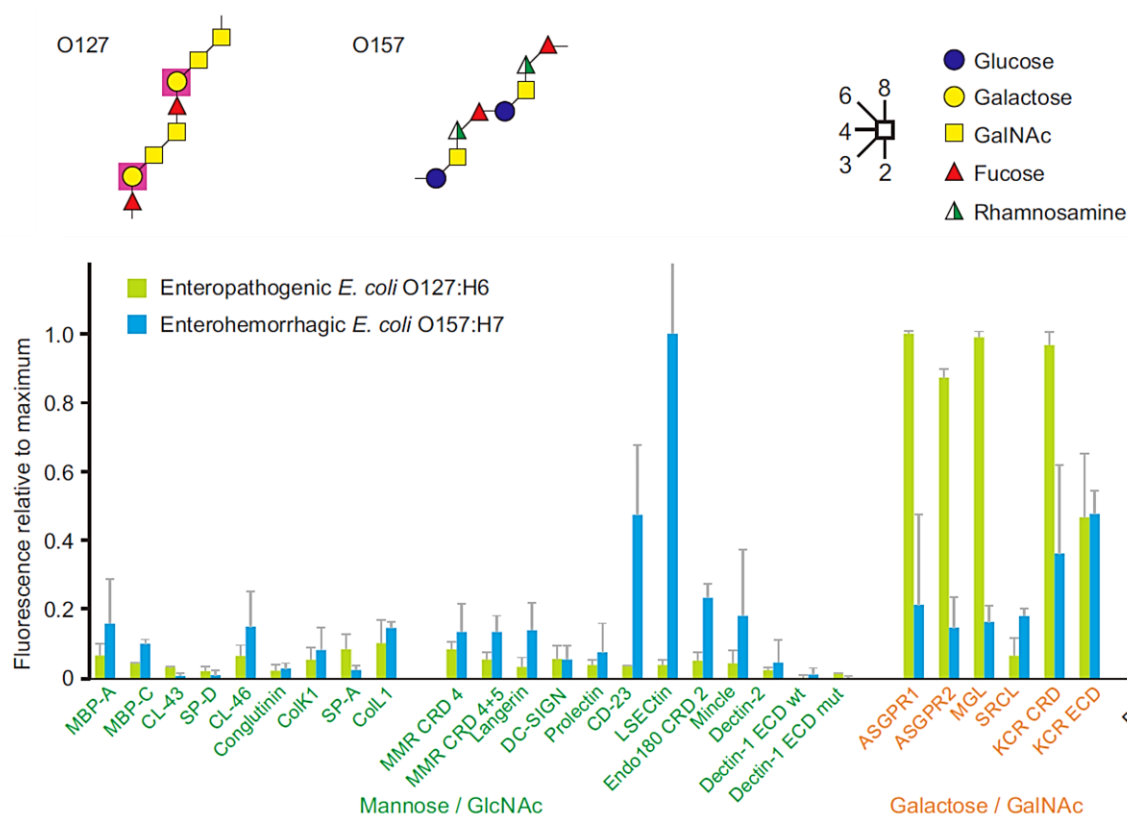


Figure 4.8: Binding of enteropathogenic and enterohemorrhagic *E. coli* strains to the lectin array. The structures of O-antigen repeating units of both *E. coli* O127:H6 and O157:H7 strains are shown on top. Purple shading in the O127 polysaccharide highlights galactose residues with free 3- and 4-OH groups for potential CRD binding. Bottom: Bacteria expressing GFP were fixed with paraformaldehyde. Fluorescence intensities were normalized to those of ASGPR1 and LSEctin for *E. coli* O127:H6 and O157:H7, respectively. Adapted from ¹⁹⁰.

Overall, such immobilization method showed interesting aspects of lectin specificity, including within receptors of similar monosaccharide spectrum targets. Similarly, other receptors presented unexpected binding profiles based on their documented ligand specificities. That is the case of LSEctin which bound to *K. pneumoniae* surface glycans despite the absence of obvious targets within the capsule structure. These differences can be attributed to the way the ligands are presented to lectins, which in this approach reflects glycan ligands in the context of intact microorganisms.

IV.2.3. Imaging approaches for LPS-lectin interaction studies

IV.2.3.1. Atomic Force Microscopy

Atomic Force Microscopy (AFM) is an advanced high-resolution imaging method known for its ability to characterize the topographical properties of a sample surfaces in both dry and hydrated settings. Particularly, with its capacity to detect forces at the piconewton range under physiological conditions, AFM has become a valuable tool for the investigation of interaction forces between ligand-receptor complexes¹⁹¹. Experimentally, the ligand is affixed to the AFM tip while the receptor is attached to a solid surface, or the other way around. By bringing the AFM probe to the surface, the ligand and the receptor are brought in contact resulting in the formation of ligand-receptor complex. Subsequently, following the tips retraction from the substrate, unbinding forces required for ligand-receptor complexes separation are measured¹⁹¹.

AFM has been used to assess different carbohydrate-lectin interactions, ranging from lectin bindings on the surface of cancerous cell lines to microorganism surfaces¹⁹². Several AFM interaction studies have been documented on the mannose-specific tetrameric plant lectin, concanavalin A (Con A). Con A-carbohydrate interaction studies by AFM using synthetic glycan structures have been carried out. Con A interactions with a set of synthetic mannose-containing glycoconjugates was reported¹⁹³. Furthermore, another study where AFM imaging and force measurement were used to probe interactions between Con A and hexasaccharides (Oligoglucose)¹⁹¹. AFM imaging was carried out in aqueous solution on functionalized solid substrates with Con A and thiol-terminated hexasaccharides. In contrast, the binding forces were measured using hexasaccharide-coated AFM tips and Con A substrates.

Besides using synthetic glycans, AFM offers the possibility to carry out interaction studies directly on cells. A study by Touhami and co-workers reported Con A interactions with yeast cells surface exposed mannose carbohydrates involved in the flocculation of yeast cells¹⁹⁴. By using functionalized AFM probes with Con A and immobilized yeast cells on a solid substrate, adhesion forces attributed to the specific binding between Con A and cell-surface exposed mannose carbohydrates could be measured. Lastly, AFM was used on Rhizobia, a Gram-negative soil bacterium producing nitrogen-fixing root nodules when engaged in symbiosis with leguminous plants¹⁹⁵. In addition to its surface imaging by AFM, binding forces involving Rhizobia and soybean agglutinin (SBA), a tetrameric Gal/GalNAc-specific lectin, were also measured. From that perspective, the AFM tip was functionalized by SBA whereas Rhizobia was attached onto the substrate. Dissociation forces were then determined, while none

could be measured when competing with GalNAc indicating that the measured unbinding forces are attributed to specific lectin–polysaccharides interactions.

IV.2.3.1. Flow cytometry and fluorescence microscopy

Flow cytometry is a powerful technique that enables rapid analysis of individual cells in solution that is used in multiple research domains, including immunology¹⁹⁶. This technique relies on light scattering and fluorescence emission measurements using two detectors: forward scatter (FSC) aligned with the laser beam that indicates cell size, and side scatter (SSC) positioned perpendicular to the laser beam which measures cell granulometry. The rapid qualitative analysis of a large number of cells together with the quantitative measurement of fluorescence intensities renders this technique a valuable approach in medical practice and cell interaction¹⁹⁷.

Flow cytometry has been employed in the detection of bacterial pathogens using lectin receptors. Hendrickson and coworkers developed a flow cytometry-based detection of pathogenic *E. coli* and *S. aureus* using plant lectins as ligand–bioreceptors. An approach that takes advantage of bacterial cell wall polysaccharides interaction with fluorescently labelled lectins, which presents an effective alternative to the antigen-antibody interaction system used for pathogenic bacteria detection^{197,198}. Furthermore, another study reported the use of flow cytometry for profiling lectins binding to bacterial isolates originating from gut microbiota¹⁹⁹. Similarly, human mannose-binding lectin (hMBL) interactions with a variety of *Hafnia alvei* strains, collected at different growth phases and expressing LPSs with different glycan lengths, were assessed by flow cytometry²⁰⁰. Lastly, the binding quantification of MGL lectin with *E. coli* strain expressing R1 core OS⁸², together with DC-SIGN binding on *E. coli* R1 cell surfaces was documented, using labelled lectins with Alexafluor647 followed by an incubation with living *E. coli* bacteria⁹⁸.

Fluorescence microscopy is an additional approach which can be used complementarily with flow cytometry in order to confirm and visualise the interactions. Among the flow cytometry examples mentioned above, fluorescence microscopy was used in addition to the flow cytometry-based detection of pathogenic strains using lectins as ligand–bioreceptors¹⁹⁷, and also in investigations of the interaction of both MGL and DC-SIGN with *E. coli* R1^{82,98}. For instance, in the case of DC-SIGN and R1 core OS investigation, fluorescence microscopy techniques ranging from epifluorescence to high-resolution microscopy beautifully completed the flow cytometry binding quantification data. The strong fluorescence observed on bacteria reflects the lectin's binding, more importantly, the fluorescence pattern observed in Stochastic Optical Reconstruction Microscopy (STORM) analysis allowed the visualization of individual

fluorophore emissions and demonstrated that this interaction undoubtedly takes place on the bacterial cell surface (Figure 4.9).

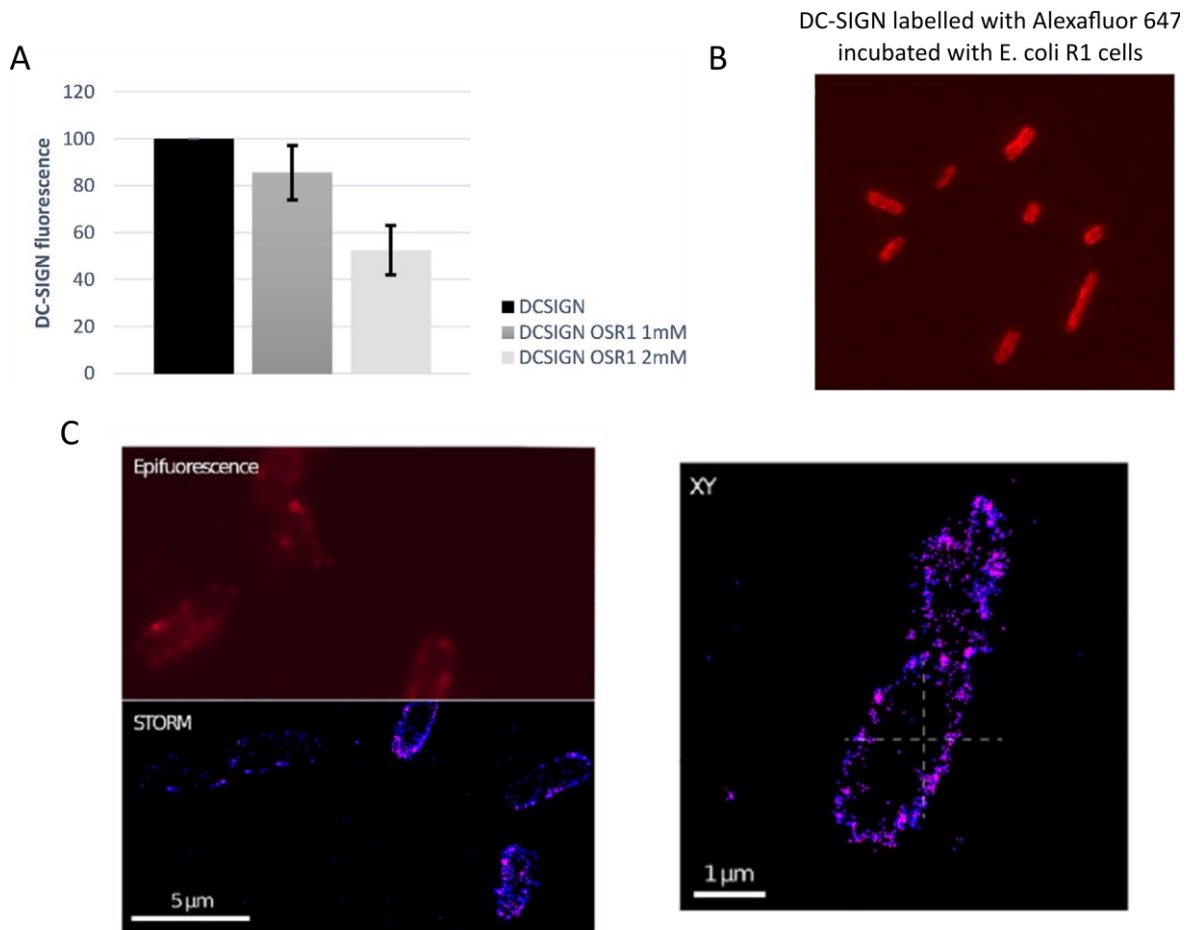


Figure 4.9: Investigation of DC-SIGN binding to *E. coli* R1 core OS by flow cytometry and fluorescence microscopy. (A) Quantification of DC-SIGN binding to *E. coli* R1 cells in the absence and presence of OS R1 as a competitor. (B) Epifluorescence imaging of preincubated R1 cells with labelled DC-SIGN. (C) Panel showing a reconstructed image of labelled R1 cells using epifluorescence and STORM (left), with a highlight of an orthogonal 2D plan projection (XY) of a single bacterium (right). Adapted from ⁹⁸.

IV.2.4. NMR spectroscopy for LPS-lectin interaction studies

NMR spectroscopy has emerged as the foremost method for gaining insights into the atomic-level view of protein-ligand interactions, and probing their respective binding affinities spanning from the nanomolar to millimolar range ²⁰¹. Different methods have been implemented for protein-ligand interaction studies, and are divided into two main categories, protein-detected and ligand-detected NMR experiments; which consist on the NMR monitoring of the protein and the ligand, respectively, in their unbound and bound states ²⁰². Interestingly, many of these techniques have been refined while investigating protein-carbohydrate interactions, notably lectin-carbohydrate complexes ²⁰³.

IV.2.4.1. Saturation Transfer Difference NMR spectroscopy

The Saturation Transfer Difference (STD) NMR spectroscopy is a ligand-detected experiment used for the study of interaction binding affinities in the range of 10^{-3} to 10^{-8} M ²⁰⁴. This method does not require any isotopic labelling as only ^1H nuclei is used. STD relies on the NOE principle where the protein is selectively saturated in a spectral range where the ligand does not contain any resonance. The magnetisation is transferred through the protein by spin diffusion and is transferred to the bound ligand by NOE. Following ligand dissociation, the saturation is then detected in the free ligand spectrum as attenuated signals. Therefore, with a rather low affinity, the ligand residence time in the bound state is shorter and the dissociation rate constant (k_{off}) is higher. This results in a more efficient saturation transfer to a large portion of the ligand and enhanced STD signal, as multiple cycles of ligand's binding and unbinding can occur during the experiment. Conversely, in the case of strong binding affinities, in the nanomolar range, where the ligand binds tightly to the protein, the residence time is longer and the k_{off} is low. This will reduce the saturation transfer effectiveness and lead to a weaker STD signal. Overall, STD experiments are generally effective when working with affinities in the low millimolar to micromolar range. At this range, there is a balance between binding and unbinding events which ensures effective saturation transfer to the ligand.

Experimentally, two sets of spectra are recorded: one with and the other without saturation of the protein signals known as the “on-resonance” and “off-resonance” spectra, respectively. Thus, ligand protons that are in contact and close in space to the protein are attenuated in the “on-resonance” spectra, while the others receive little or no saturation. Subtraction of the on-resonance and off-resonance spectra results in the final STD NMR spectrum on which only signals from bound ligands are observed, since signals of the unbound ones have been subtracted (Figure 4.10) ^{202,203}.

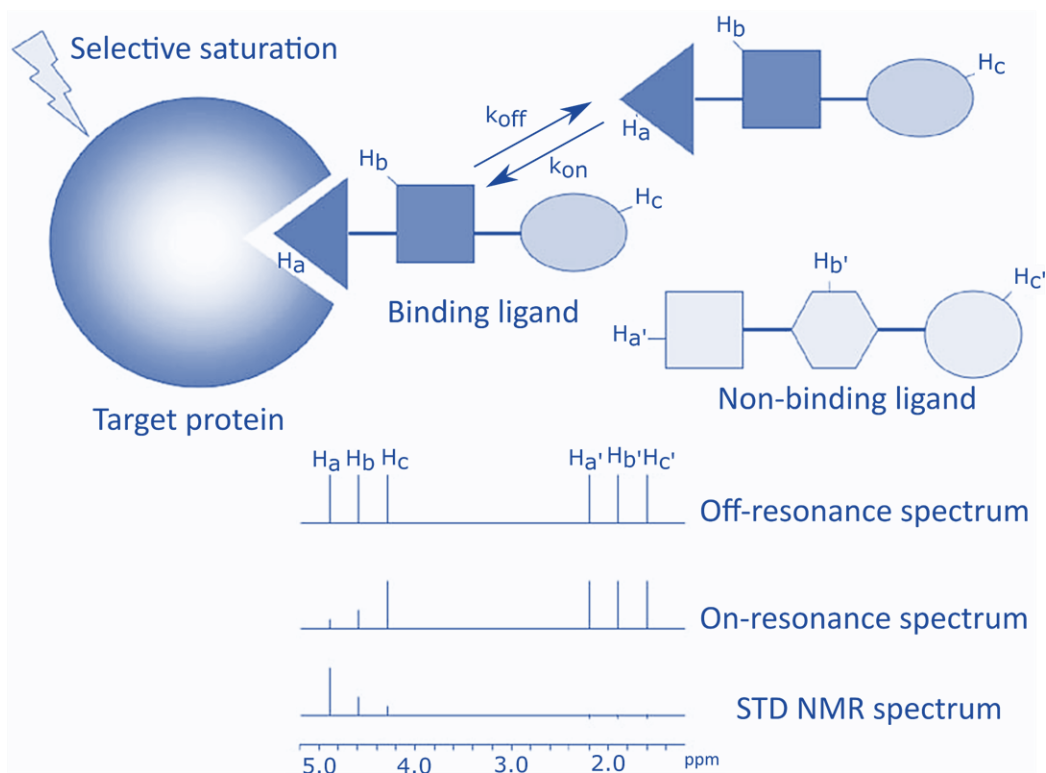


Figure 4.10: Schematic representation of the principle of STD NMR. Target protein is selectively saturated. The saturation is transferred through the protein by spin diffusion, and subsequently spreads to the bound ligand. Subtraction of on-resonance and off-resonance spectra results in the final STD spectrum, where only binding ligand signals are observed. Ligand protons in close contact with the protein exhibit the strongest signal intensities in the STD spectrum (H_a) compared with the ones that are solvent exposed which show weak STD signals intensities (H_b and H_c). Non-binding ligand protons do not show any STD NMR signals (H_a' , H_b' , and H_c'). Adapted from ²⁹³.

This technique allows the direct detection of binding components, even within a mixture of compounds. More importantly, given the fact that the magnetization transfer from the receptor to the ligand protons is dependent on the inverse sixth power ($1/r^6$) of their distances in the bound state, not all the protons exhibit the same amount of saturation (attenuation). In other words, ligand components presenting the strongest contact with the protein exhibit the most pronounced signals in the STD NMR spectrum. Taking advantage of this particularity, information about spatial proximities of the individual protons to the receptor in the bound state can be determined ²⁰⁵. That is achieved by comparing the saturation percentages received by ligand protons, after normalizing all measured STD signals ($(I_{\text{off-resonance}} - I_{\text{on-resonance}})/I_{\text{off-resonance}}$) relative to the most intense one arbitrarily set to 100% ^{205,206}.

Interestingly, STD NMR spectroscopy method was initially designed to assess the binding affinity of carbohydrate molecules to wheat-germ agglutinin (WGA), and found to be a valuable technique for the determination of the epitope mapping of the ligand²⁰⁷. Since then, it has been largely used in this research field, notably for bacterial carbohydrate (LPS)-lectin interactions, including for *E. coli* R1 core type interactions with MGL and DC-SIGN lectins. The STD NMR results confirmed DC-SIGN ECD interaction with *E. coli* R1 OS, and revealed an extended binding epitope involving the outer core region of OS R1⁹⁸. Similarly, the latter interaction with MGL lectin by STD NMR suggested that once again the outer core moiety was the one mostly engaged in the binding¹⁰¹.

IV.2.4.2. Chemical shift perturbation

Chemical shift titration is a straightforward NMR method which allows the mapping of ligand-protein interaction site on the protein, and calculation of equilibrium dissociation constant (K_D). In practice, increasing concentrations of an unlabelled ligand, consisting of a small molecule or another macromolecule, are gradually added into an isotopically-labelled protein (^{15}N). A 1D or 2D NMR spectrum of the protein is recorded for each stage of ligand titration²⁰⁴. The standard experiment employed in this context is the 2D ^1H - ^{15}N correlation experiment, HSQC. Chemical shifts are sensitive to structural and chemical environment changes, thus, the peaks that experience significant shifts are likely involved in the ligand binding. That is why it's very important to pay attention and maintain some physical parameters (pH, temperature, buffer composition) constant throughout the experiment in order to avoid the rise of any chemical shift perturbations (CSPs) other than the ones specific to protein-ligand binding²⁰¹.

For large affinities, where the exchange regime is slow, the binding event translates into intensity changes. Resonances of the protein-free state gradually disappear while the protein-ligand complex resonances simultaneously appear. For small affinities (dissociation constants typically less than 1 μM), some residues undergo ^1H - ^{15}N CSP upon ligand addition, with residues located in close proximity to the ligand being significantly perturbed. The fitting of the chemical shift variations as a function of ligand concentration can be used to obtain a value for the dissociation constant of the ligand, K_D . Additionally, in some protein-ligand interactions, the resonances follow a rather curved displacement which generally indicates complex binding. This can include the presence of at least two binding modes²⁰⁴ with different affinities. Therefore, at increasing ligand concentrations, the different sites will be occupied

sequentially and will all contribute to the overall chemical shift. Secondly, cooperative binding, either positive or negative cooperativities, where the binding of one ligand can influence the affinity of another ligand molecule can also result in curved peaks displacement. Additionally, ligand binding to its site can induce conformational changes in other sites. This allosteric effect can also be translated to non-linear chemical shift changes. Consequently, not all the peaks undergoing CSP are ultimately involved in the binding, sometimes prominent CSPs are rather induced by backbone conformation and allosteric changes in other sites upon protein complexation to the ligand ²⁰¹.

This protein-detected method has been widely used in protein-ligand interaction studies, notably in carbohydrate-lectin interactions. Binding studies of galectins to multivalent conjugates were performed using ¹H-¹⁵N-HSQC titrations, and K_{DS} in the medium–high mM range were measured ²⁰⁸. Additionally, the interaction of natural ligands with galectins was also performed by NMR, including the interaction of galactorhamnogalacturonate glycan (GRG), a large heterogeneous glycan derived from citrus pectin, with human galectin 1 (Gal-1) ²⁰⁹. Based on CSPs analysis, they could identify a much more extended GRG-binding region which included the canonical binding region ²⁰⁹. Lastly, concluding on a bacterial carbohydrate-lectin interaction example, the titration of MGL lectin with increasing concentrations of *E. coli* R1 OS core. A study carried out on both the CRD and full extracellular domains of MGL which both revealed a secondary glycan-binding site opposite to the canonical binding site for this OS core ⁸². This interaction was followed from both the ligand perspective using STD ¹⁰¹ and the protein perspective using CSPs ⁸², which highlights the valuable set of information NMR can provide.

IV.2.5. Other approaches for carbohydrate-lectin interaction studies

Carbohydrate–lectin interactions are clearly of paramount importance in a number of physiological and pathological mechanisms²¹⁰, hence there is a highest interest in deciphering their molecular determinants. Therefore, complementary to some of the previous approaches described above, other analytical techniques have been implemented for a complete investigation of these bindings. Among them, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), and bio-layer interferometry (BLI) all of which are usually employed for determining affinity, thermodynamic and kinetic parameters of the interaction.

BLI together with SPR provide real-time monitoring of the ligand association and dissociation events, giving access to kinetic parameters. SPR consists in the immobilization of one partner on a surface into a microfluidic device and the injection of the second one in solution under continuous flow on the surface. In contrast, BLI is performed in multi-well plates which contain one partner into which a biosensor tip covalently functionalized with the second partner is immersed²¹⁰. SPR detects biomolecular interaction by monitoring changes in reflected light intensity near a metal surface, whereas BLI measures shifts in interference patterns caused by biomolecular binding on the coated tips^{210,211}.

ITC on the other hand is performed in solution, without prior labelling or any partner immobilization, and provides the assessment of thermodynamic parameters and interaction stoichiometry²¹⁰. Upon ligand injection into the receptor sample cell, subsequent interaction results in temperature differences between the sample and reference cells. The energy required for maintaining constant temperature differences between the two cells, translated into heat changes, is measured and provides data on binding affinity, stoichiometry, and thermodynamic parameters²¹².

These techniques have contributed much to the study of multivalent-lectin interaction, including ITC which was reported in many of the studies involving animal and plant lectins²¹³. The binding of the plant lectin Con A to a series of mono- and oligosaccharides was investigated using ITC, and their thermodynamics and stoichiometry of the binding were determined²¹⁴. BLI was used for interaction studies between multivalent glycoconjugates (tetra- and hexadecavalent GalNAc) and *Helix pomatia* agglutinin (HPA) lectin. Kinetic parameters of multivalent interaction spanning from the micro- to the nanomolar range were measured²¹⁰. Lastly, SPR was widely used in carbohydrate-lectin binding and its use in competition-based assays for the evaluation of monovalent and polyvalent compounds was reported. A development in SPR experimental set-ups has been made and led to the design of

oriented immobilized lectin surfaces with well-oriented CRDs ²¹⁵. This approach improves the glycoconjugates binding with lectins, which are presented in the context of the cell surface, and thus improves the binding avidity. This direct interaction set-up was perceived on DC-SIGN and Langerin lectins and the binding was evaluated using a set of synthetic glycoconjugates ²¹⁵. It was later extended to bacterial carbohydrates, notably *E. coli* R1 LOS reconstituted in dodecylmaltoside (DDM) micelles for which interactions with oriented DC-SIGN surfaces were documented (Figure 4.11) ⁹⁸. An evaluation of the apparent dissociation constant (K_{Dapp}) was made, and was found to be remarkably increased compared to what was obtained on regular SPR competition assays of OS R1 binding to DC-SIGN (from IC_{50} of 1 mM on regular SPR competition assay to a K_{Dapp} of 15 μ M when using oriented lectin surfaces) ⁹⁸.

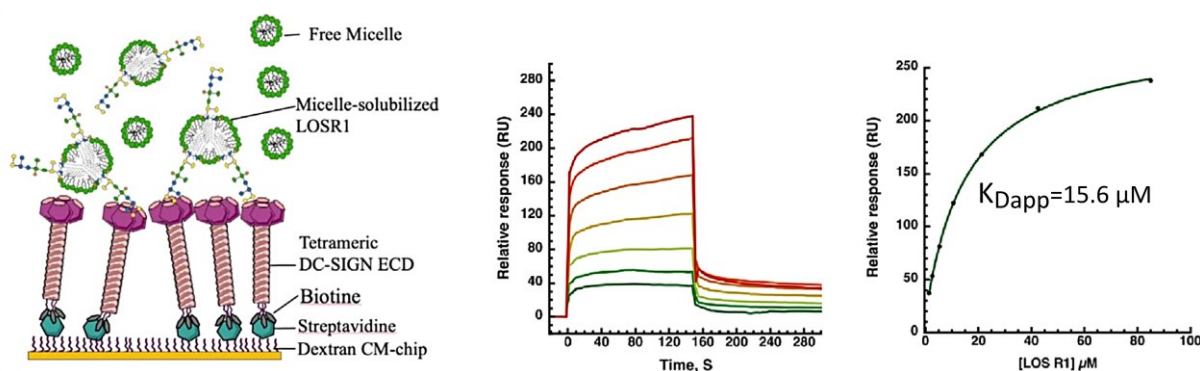


Figure 4.11: Interaction of *E. coli* R1 LOS reconstituted in DDM micelles with surface-oriented DC-SIGN lectin. Increasing concentrations of R1 LOS micelles were injected onto DC-SIGN ECD. A binding response sensorgram of increasing titration concentrations is plotted following reference-subtraction, and the titration curve was obtained by plotting the steady state binding responses against R1 LOS concentration. Adapted from ⁹⁸.

IV.3. Position of the PhD project as it relates with the state-of-the-art

The present PhD project focused on the investigation of Gram-negative bacterial LPS interaction with lectins, particularly with MGL lectin. Within the present project, LPSs from three different *E. coli* strains exhibiting variable core OS structures, and harbouring or not an O-antigen moiety were chosen. LOSs from *E. coli* R1 (F470) and *E. coli* R3 (F653) strains, and LPS from *E. coli* O157:H7 strain were considered (Figure 4.12). *E. coli* R1 and R3 strains represent the predominant core OS types observed in *E. coli* isolates, including verotoxigenic species (VTEC) ²⁶. Enterohemorrhagic and verotoxigenic *E. coli* O157:H7 strain for its part, with its smooth type LPS with an O-Antigen O157 and flagellar antigen H7, is responsible for severe food-borne infections ²⁶. Its core OS shares an identical structure with *E. coli* R3 strain core OS. Whereas, the O-Antigen portion, which has been identified by MS and NMR spectroscopy, is constituted of a tetrasaccharide repetition unit consisting of β -D-Glc, α -D-PerNAc, α -D-GalNAc, and α -L-Fuc, where D-PerNAc is 4-acetamido-4,6-dideoxy-D mannose and Fuc is a fucose unit. ²¹⁶.

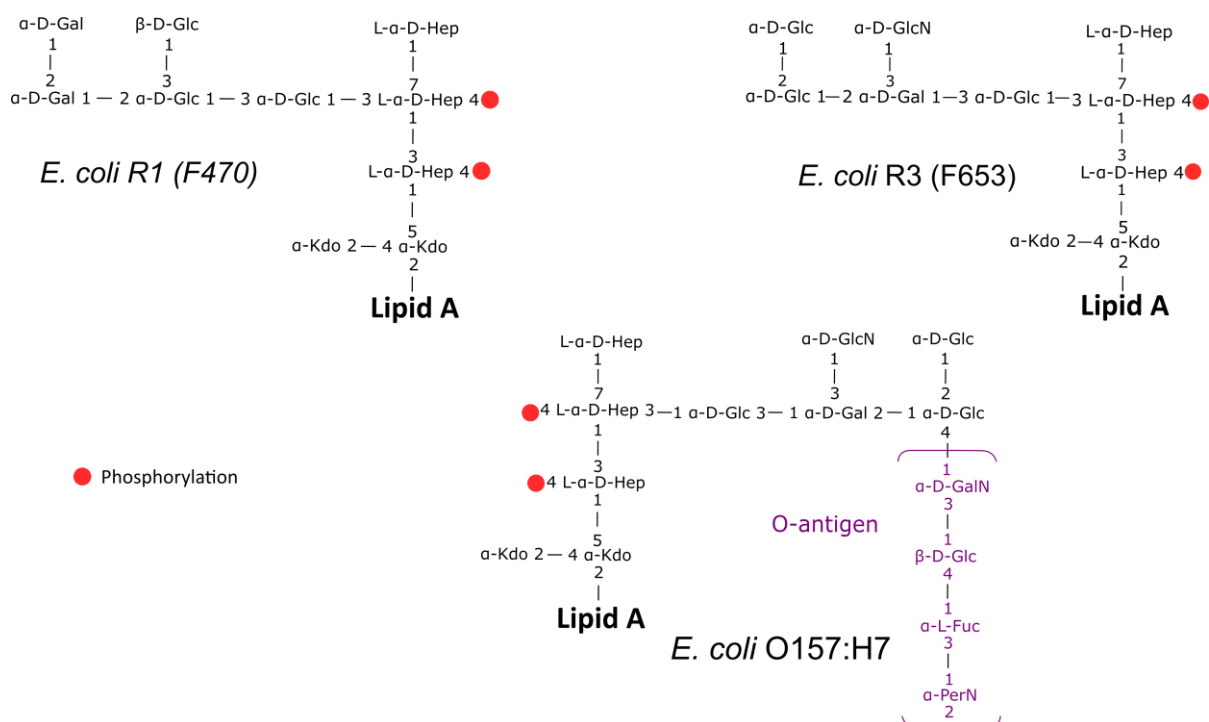


Figure 4.12: Chemical structure of the core OSs of the different *E. coli* strains used in this study.

At the beginning of this project, preliminary data regarding *E. coli* R1 OS core interactions with MGL CRD have already been collected. They pointed out a possible presence of a potential secondary binding site on MGL other than the canonical binding site to which R1 OS bound. Further investigations were then launched on MGL ECD, in its trimeric oligomerized form, and its interaction with deacylated LOS or to native LOS directly exposed on whole cells was assessed. This work is presented in chapter V, and was achieved using a combination of structural and biophysical methods ranging from the cellular level to the atomic level. MGL binding on *E. coli* cell surfaces was monitored by fluorescence microscopy and quantified by flow cytometry. Binding affinities were evaluated by SPR using oriented lectin surfaces on which LOS micelles were flown on solution. NMR spectroscopy titration of MGL with soluble ligand fractions (OS) revealed an additional glycan binding site for MGL located on the opposite surface of the canonical binding site. Finally, a 3D model of the trimeric MGL was determined using a combination of Small Angle X-ray scattering and Alphafold modelling, where the orientation of its CRDs was found to present the two glycan binding sites and ensure a tight binding to glycans on a surface.

As discussed previously, there is clearly a considerable challenge when it comes to studying LPS molecules. Their amphiphilic nature together with their different molecular assemblies in solution render LPS handling and adaptability to different biophysical methods quite impossible. This explains the prerequisite employment of chemical fractionation prior to any structural analysis. In this context, we sought to develop an innovative method for LPS reconstitution in membrane mimicking assemblies which would render possible biophysical and structural studies of intact LPS. To this end, chapter VI presents the use of Amphiphilic styrene maleic acid copolymers (SMA) polymers to form LPS nanodiscs. These polymers have been developed for membrane proteins reconstitution, and stand-out for their ability to spontaneously insert into native or reconstituted membranes and form discoidal shaped nanodiscs without prior detergent solubilization²¹⁷. SMA copolymers were thus used to form LPS nanodiscs from purified LPSs or directly extracted from bacterial outer membranes of laboratory and pathogenic *E. coli* strains (*E. coli* O157:H7 strain, *E. coli* R1 (F470), and *E. coli* R3 (F653)). The nanodiscs dimensions were assessed by Dynamic Light Scattering and AFM. They were next characterized at atomic scale by MAS NMR spectroscopy. Lastly, LPS nanodiscs were successfully used to monitor interactions with the immunity C-type lectin receptors MGL and Polymyxin B antibiotic by Quartz Crystal Microbalance with dissipation (QCM-D) and BLI.

Finally, after nanodiscs system validation, substantial efforts have been devoted for a one and final objective to construct a model of the MGL ECD arrangement on LPS membrane nanodiscs. Therefore, chapter VII describes the different advances made so far, and the reliant approaches used. For that perspective, click-chemistry is employed to graft gold nanoparticle onto MGL ECD for Electron Microscopy (EM) structural characterization of MGL-nanodiscs complexes. Finally, new MGL ECD constructions are designed for further biophysical analysis of this complex, including by solution NMR spectroscopy, or for X-Ray crystallographic structure determination of the trimeric MGL ECD.

IV.3.1. What about Styrene-Maleic Acid Lipid Particles: SMALPs

Despite the great advances made in the development of methodologies for the reconstitution of membrane components in a more-native like environment, all these methods share a common limitation: they all require a prior detergent-solubilization step. The use of Styrene-Maleic Acid (SMA) polymer has been implemented as an alternative method in membrane protein studies that does not require detergents. SMA polymers have long been used in life sciences in cancer therapy before it was found to insert in phospholipid bilayers. Consequently, new applications of SMA for the solubilization of lipid bilayers have been developed, notably for membrane proteins solubilization²¹⁸. Therefore, the resulting SMA nanodiscs have been termed SMA Lipid Particles (SMALPs).

Initial reports of SMA-solubilized phospholipid vesicles revealed the formation of discoidal assemblies. Their size distribution was thoughtfully studied using electron microscopy, size exclusion chromatography, and dynamic light scattering²¹⁸. Small-angle neutron scattering measurements of SMALP nanodiscs formed from DMPC vesicles confirmed the presence of discoidal objects with high thermal stability²¹⁹. Their diameter was found to be 9.4 nm with a diameter of the phospholipid core of 7.4 nm (Figure 4.13A). The thickness of the discs is estimated at around 4.6 nm, with a thickness of the core of the disc (lipid bilayer) of 26 ± 2 Å and 10 ± 2 Å for the lipid headgroups.

These polymers spontaneously insert into the membrane bilayer thus forming nanodiscs containing a patch of native membrane. According to the reported studies carried out on model membranes^{218,220,221}, SMA nanodiscs formation follows a three-step model (Figure 4.13B). Firstly, SMA copolymer adsorbs to the surface of the bilayer, a process modulated by electrostatic interactions between the polymer and charged lipids at the membrane surface. This step can be promoted by increasing concentrations of the polymer and the ionic strength that helps overcoming repulsive electrostatic interactions between negatively charged polymers and

anionic lipids²¹⁸. Secondly, SMA polymer inserts into the hydrophobic core of the membrane. The styrene units intercalate between lipid acyl chains, while the charged maleic acid groups interact with the lipid headgroups²²⁰. This process is determined by the physical properties of the bilayer, including lipid packing and bilayer thickness. Lastly, bilayer solubilization followed by nanodiscs formation takes place. Coarse-grained molecular dynamics simulations provided insights into molecular details of nanodiscs formation²²². This study suggests that following SMA insertion into the hydrophobic core of the bilayer, membrane defects appear, including small pore formation. SMA polymer acts as a stabilizer with its carboxyl groups faced towards the water pores while the styrene moiety intercalates between the lipid acyl chains. This stabilizing effect leads to pore growth and membrane disruption with subsequent nanodiscs formation²²².

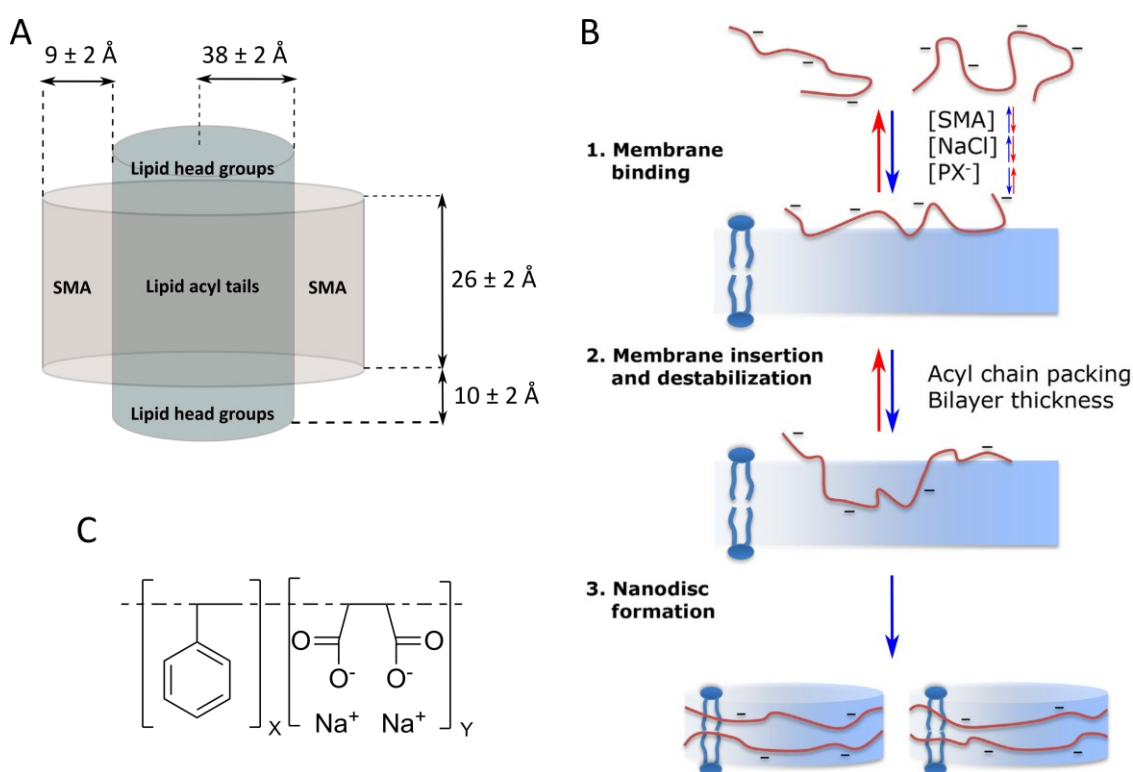


Figure 4.13: Styrene-maleic acid in membrane research. (A) Dimensions of DMPC styrene-maleic acid nanodiscs as measured from small-angle neutron scattering experiments. (B) Schematic representation of the three-step model for membrane solubilization by SMA polymer. Anionic SMA polymer adsorb onto the lipid membrane. A process modulated SMA and salt concentration, and the presence of anionic lipids (PX⁻). Insertion of the polymer into the hydrophobic core of the membrane driven by the hydrophobic effect takes place. This step leads to membrane destabilization and nanodiscs formation. (C) Chemical structure of SMA(2:1) polymer. Adapted from^{218,219,221}.

SMA polymer is synthesized following the hydrolysis of the styrene–maleic anhydride (SMA_{anh}) copolymer. This latter is in turn synthesized by radical polymerization of styrene and maleic anhydride monomers²¹⁸. Nevertheless, SMA chemistry can interfere with the properties of the resulting SMALPs. The maleic acid unit harbours two carboxyl groups with two distinct pK_a values at around 6, and 10²¹⁸. Therefore, at pH conditions below the lowest pK_a value (~6.0), where SMA acid groups are protonated, the polymer adopts globular structures and aggregates²¹⁸. In addition, a limited tolerance of SMA to divalent cations has also been proposed, and results in SMA precipitation and hampered solubilization. A process promoted by the coordination of divalent cations by the SMA carboxyl groups. These limitations are overcome by the development of new polymers with new formulations²²³. Although SMALPs have some limitations when it comes to their sensitivity to low pH conditions and presence of divalent cations, they have been successfully used for lipids and membrane protein solubilization. This includes nanodiscs formation from native membranes of T cells and from bacterial membranes including *Mycobacterium tuberculosis*^{224–226}.

Among the available polymers, SMA(2:1) and SMA(3:1), with either a 2:1 or 3:1 ratio of styrene:maleic acid, are the most frequently used ones. Investigation of the effect of styrene:maleic acid ratios on membrane solubilization were undertaken²²⁷. This study revealed that SMA(2:1) (Figure 4.13C) polymer is more efficient in the solubilization of phospholipid vesicles when compared to the more hydrophobic polymers SMA(3:1) and SMA(4:1). This characteristic is attributed to this polymer's balanced structural properties between hydrophobic and charged groups. This ratio provides efficient membrane destabilization through the convenient distribution of styrene, and ensures solubility at pH values above 5.0, the broadest pH range for solubilization among tested polymers²²⁷. Thereby, we have chosen to use this polymer for the preparation of LPS glycoconjugate nanodiscs.

Results

V. MGL binds *E. coli* surface with high avidity

Bacterial cell envelopes are decorated with a variety of glycan structures, namely LPS for gram-negative bacteria. These complex glycoconjugates are extracellularly recognised by the immune system through their Lipid A moiety *via* the LBP-MD2-TLR4 cascade⁵⁸, and intracellularly detected by the caspase system⁶⁹. On the other hand, their glycan portion is recognized by CLRs, a protein family present on antigen-presenting cells. The trimeric type II CLR MGL, expressed on the surface of dendritic cells and macrophages, with its QPD motif is reported for the recognition of glycans with terminal Gal/GalNAc. MGL recognizes numerous tumour-associated carbohydrate antigens (TACAs), including the Thomsen nouveau (Tn)-antigen (GalNAc- α -1-O-Ser/Thr) detected in cancers (90% of carcinomas)¹¹⁰. Furthermore, it was also described for microbial glycans recognition, including *S. aureus*, *C. jejuni*, *N. gonorrhoea*, and *E. coli*^{81,101,228,229}.

This chapter presents an interdisciplinary work focused on deciphering the interaction involving MGL and some *E. coli* surface carbohydrates. This project was done in collaboration with a research group in Napoli (University of Naples Federico II, Department of Chemical Sciences, Structure and Synthesis of Carbohydrates), with whom previous collaborations have highlighted the recognition of MGL by *E. coli* R1 core type¹⁰¹. Investigations of the molecular interaction behind this recognition at both the cellular and atomic levels were thus launched. Monomeric MGL CRD and MGL ECD in its trimeric oligomerized form were chosen for this study. Interactions with *E. coli* R1 and R3 OS cores or reconstituted LOS in detergent micelles were monitored. Fluorescence microscopy and flow cytometry were used to assess and quantify MGL binding onto live *E. coli* cells. SPR was used for binding affinities evaluation where LOS micelles were flown onto oriented lectin surfaces. Chemical shift perturbation analysis was used to monitor *E. coli* core OSs binding on MGL ECD and unveiled a new glycan binding site opposite to the canonical site. Lastly, a 3D model of the trimeric MGL was combining Small Angle X-ray scattering and AlphaFold, and elucidated the convenient orientation of MGL CRDs allowing high avidity glycan bindings.

The tasks to which I contributed in this work package are developed in this chapter and comprise:

- MGL ECD expression optimisation in minimum medium and purification. Different preparations were done for the different studies carried out, including a ^{15}N -labelled, a perdeuterated ^2H - ^{15}N -labelled MGL preparation, alongside with some key mutant MGL ECD constructs.
- NMR spectroscopy titrations of ^2H - ^{15}N -labelled MGL ECD with delipidated LOSs (OS R1 and OS R3) and data analysis.
- LOS reconstitution in DDM micelles and SPR experiments.

Another section discussing MGL interactions with *E. coli* R1 and R3 cells as well as *E. coli* strains harbouring O-antigen moieties using flow cytometry will be presented. This work was done during the internship of Macha Trembley, a Master 1 student I supervised, and was not included in the published article, since it was done after its submission.

V.1. Production of recombinant MGL ECD lectin

V.1.1. MGL ECD constructions

As mentioned in the introduction section, only a single gene encoding for macrophage galactose-type lectin is identified in humans. This gene is subjected to extensive splicing events resulting in different isoforms. Among the described variants, isoform 1 expressed in dendritic cells referred to here as DC-ASGPR, and isoform 2, referred to here as MGL, on which our studies focus. DC-ASGPR isoform is a longer variant of MGL with an insertion of 27 amino acids and a three amino acids deletion in the coiled-coil neck domain (Figure 5.1A) ¹⁰⁶. Both MGL and DC-ASGPR have identical CRDs, which can explain their similar carbohydrate recognition specificities.

MGL is a 292 amino acids protein sequence. It consists of a cytoplasmic domain (residues M1-H40), a transmembrane domain (residues L41-F60), a long neck domain (residues Q61-T155), and a CRD domain (residues C156-H292). Although MGL is a transmembrane protein, the choice to work only with its extracellular domain was made. If the full-length construct was considered, it would have required prior solubilization steps with detergent or reconstitution in nanodiscs which may present some disadvantages later on when interaction experiments with LPSs will be undertaken such as non-specific interactions and potential disruption of LPS nanodiscs. Therefore, the DNA fragment encoding the human MGL isoform 2 ExtraCellular Domain (ECD) (UniProt entry number Q8IUN9-2, residues Q61-H292) was subcloned into the pET-30 expression vector using Nde I and Xho I restriction sites. The cloned sequence harbours a StrepTagII, a Factor Xa protease cleavage site, and 3 glycine residues at the N-terminus of the protein (pET-30-StreptagII-Xa-3G-ECD_MGL). Thereby, the resulting expression vector expresses a 250-residue protein. The glycine residues are added in the sequence in order to introduce selective modifications on the N-terminal side of the protein, including by sorting click reactions which is notably employed for the preparation of lectin-oriented SPR surfaces. The construction of MGL ECD used during the thesis is presented in Figure 5.1B.

A

DC-ASGPR	MTRTYENFQYLENKVKVQGFKNGPLPLQSLQLRQLCSGPC	CHLLLSLGLGLLLLVIICVVGFI	60
MGL	MTRTYENFQYLENKVKVQGFKNGPLPLQSLQLRQLCSGPC	CHLLLSLGLGLLLLVIICVVGFI	60

DC-ASGPR	QNSKFQRDLVTLRTDFSNFTSNTVAEIQALTSQGSSLEETIASLKAEEVEGFKQERQAGVS		120
MGL	QNSKFQRDLVTLRTDFSNFTSNTVAEIQALTSQGSSLEETIASLKAEEVEGFKQERQA---		117

← Coiled-coil neck			
DC-ASGPR	ELQEHTTQKAHLGHCPHCVSVVHSEMLLRVQQLVQDLKKLTCQVATLNNN---	ASTE	177
MGL	-----VHSEMLLRVQQLVQDLKKLTCQVATLNNNGEEASTE		153

DC-ASGPR	GTCCPVNWEHQDSCYWFSGMSWAAEAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYT		237
MGL	GTCCPVNWEHQDSCYWFSGMSWAAEAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYT		213

→ CRD ←			
DC-ASGPR	WMGLSDPEGAWKWVDGTDYATGFQNWKPGQPDDWQGHGLGGGEDCAHFHPDGRWNDDVCQ		297
MGL	WMGLSDPEGAWKWVDGTDYATGFQNWKPGQPDDWQGHGLGGGEDCAHFHPDGRWNDDVCQ		273

DC-ASGPR	RPYHWCEAGLGQTSQESH		316
MGL	RPYHWCEAGLGQTSQESH		292

B

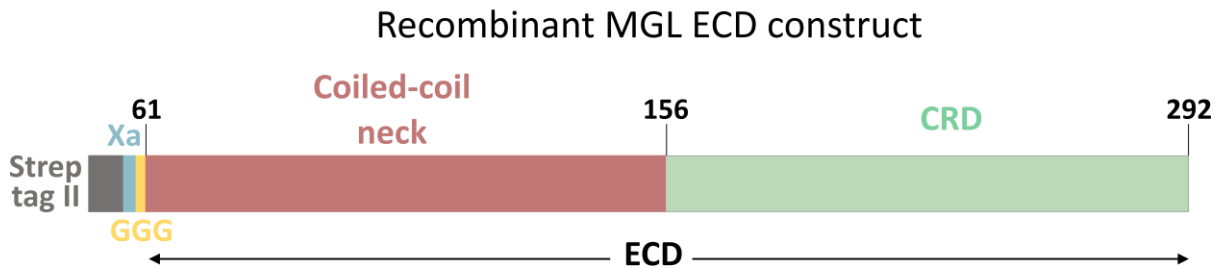


Figure 5.1: Amino acid sequences of MGL variants and the construct used in this study. (A) Sequence Alignments of DC-ASGPR (UniProt ID Q8IUN9-1) with MGL (UniProt ID Q8IUN9-2). Amino acid insertions and deletions in DC-ASGPR are shaded. The transmembrane domain (TM) is highlighted with a box. Both coiled-coil and CRD domains are underlined. (B) Construction of MGL ECD used during the thesis.

An MGL ECD mutant in a key residue of MGL CRD was designed, where a substitution of the D245 residue in the QPD motif into a histidine residue was made and referred to as MGL-ECD^{D245H}. This mutation would abolish the MGL canonical binding site. An alignment of both the MGL ECD wild type and mutant sequences used in this work is presented (Figure 5.2).

```

MGL ECDwt      MASWSHPQFEKIEGRGGGQNSKFQRDLVTLRTDFSNFTSNTVAEIQALTSQGSSLEETIA 60
MGL ECDD245H MASWSHPQFEKIEGRGGGQNSKFQRDLVTLRTDFSNFTSNTVAEIQALTSQGSSLEETIA 60
*****

MGL ECDwt      SLKAEVEGFKQERQAVHSEMLLRVQQLVQDLKKLTCQVATLNNNGEEASTEGTCCPVNWV 120
MGL ECDD245H SLKAEVEGFKQERQAVHSEMLLRVQQLVQDLKKLTCQVATLNNNGEEASTEGTCCPVNWV 120
*****

MGL ECDwt      EHQDSCYWF SHSGMSWAEAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYTWMGLSDPEG 180
MGL ECDD245H EHQDSCYWF SHSGMSWAEAEKYCQLKNAHLVVINSREEQNFVQKYLGSAYTWMGLSDPEG 180
*****

MGL ECDwt      AWKWDGTDYATGFQNWKPGQPD DWQGHGLGGGEDCAHFHPDGRWNDDVCQRPYHWVCEA 240
MGL ECDD245H AWKWDGTDYATGFQNWKPGQPH DWQGHGLGGGEDCAHFHPDGRWNDDVCQRPYHWVCEA 240
*****

MGL ECDwt      GLGQTSQESH      250
MGL ECDD245H GLGQTSQESH      250
*****

```

Figure 5.2: Alignment of MGL ECD^{wt} and MGL ECD^{D245H} mutant sequences. The point mutation is highlighted with orange. Alignment was performed on Clustal Omega.

V.1.2. MGL ECD expression in minimum media

MGL ECD is overexpressed as an intracellular protein in inclusion bodies (IB). Therefore, its purification is preceded by a refolding step. Although the refolding process was successfully employed on MGL ECD expressed in Luria-Bertani medium, the ones prepared from cultures grown in standard minimum media (intended for the purpose of isotopic labelling for NMR spectroscopy experiences) could never be refolded. These observations bring to mind the presence of compounds in the culture medium which prevent the proper refolding of the protein.

Therefore, an optimization of MGL expression in different M9 minimum media was carried out at the beginning of my thesis. To this end, three M9 media were tested, referred to here as M9⁺⁺, M9^{std}, and M9^{light}, each presenting distinct degrees of enrichment with minerals and vitamins (Table 3). The idea was to monitor the effect of the media on the protein expression level, and pick the medium allowing a descent expression with the least complex composition possible.

Table 3: minimal media composition

	M9⁺⁺	M9^{std}	M9^{light}
Additional minerals and vitamins	MgSO ₄ CaCl ₂ FeSO ₄ .7H ₂ O CaCO ₃ ZnSO ₄ .7H ₂ O MnSO ₄ .4H ₂ O CuSO ₄ .5 H ₂ O CoSO ₄ .7 H ₂ O H ₃ BO ₃ Fuming HCl	MgSO ₄ CaCl ₂ MnCl ₂ ZnSO ₄ FeCl ₃ Vitamins' mix	MgSO ₄ CaCl ₂ Thiamine
M9 salts	Na ₂ HPO ₄ KH ₂ PO ₄ NaCl NH ₄ Cl Glucose		

Regarding the expression conditions, they were all carried out in transformed *E. coli* BL21(DE3) strains with the pET-30-StreptagII-Xa-3G-ECD_MGL plasmid. The overexpression was induced by Isopropyl-β-D-thiogalactopyranoside (IPTG) at an optic density OD₆₀₀ of 0.8 at 37°C, 180 rpm for three hours. Fractions before and after induction were collected and loaded on a 15% SDS-PAGE. For the three tested media, an intense band appeared after induction at the expected molecular weight (28 kDa) and confirms the expression of MGL ECD in the different conditions (Figure 5.3). Even though, the cells density reached at the end of the expression is lower for M9^{light} compared to the other two media (OD₆₀₀ of ~2 for M9^{light} against almost an OD₆₀₀ of 3 for M9⁺⁺ and M9^{std}), the protein expression was rather nice. Thereby, M9^{light} medium was chosen for the expression of MGL ECD, which does not contain iron, manganese or other minerals that can potentially hinder the protein refolding.

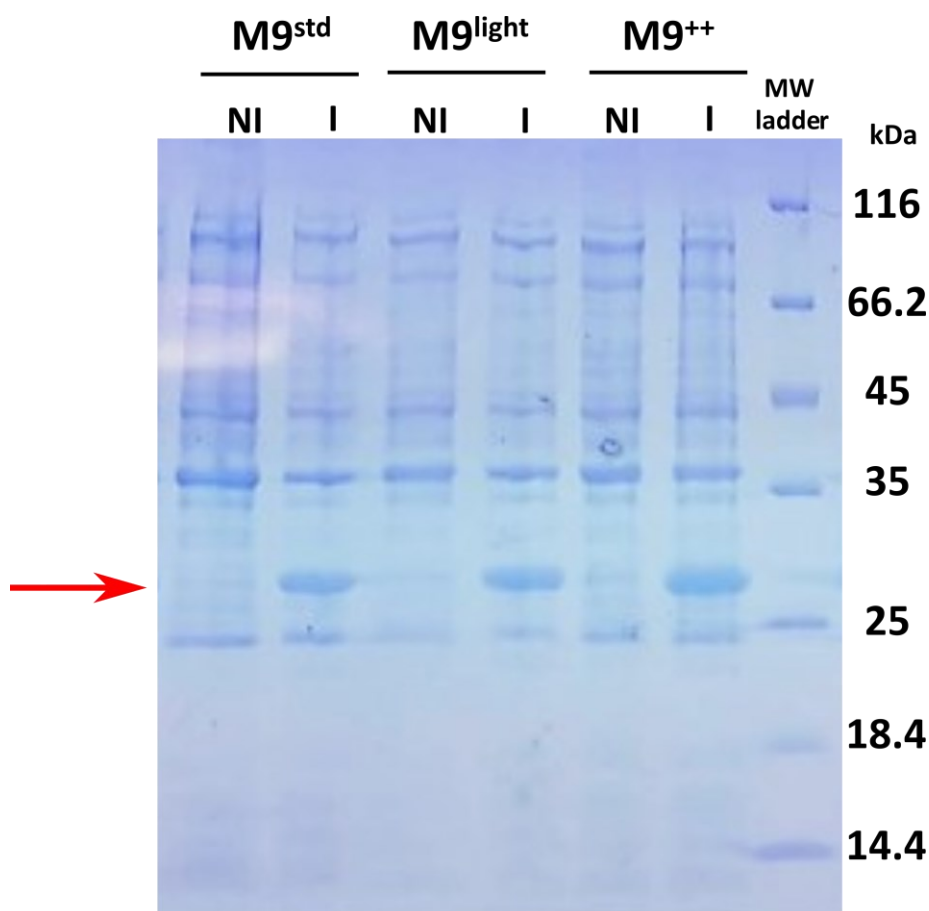


Figure 5.3: Expression test of MGL ECD in M9 minimal media. Cells were cultured in the different media described above, Protein expression was induced at an OD_{600} of 0.8 with 1mM IPTG for 3h at 37°C, 180 rpm. NI: non-induced; I: induced.

V.1.1. Refolding and purification of MGL ECD

MGL-ECD was produced with two main isotopic labelling patterns: first a uniformly ^{15}N -labeled MGL-ECD ($[\text{U-}^{15}\text{N}]$ MGL-ECD), then a uniformly ^2H , ^{15}N -labelled form ($[\text{U-}^2\text{H}, ^{15}\text{N}]$ MGL-ECD). Both preparations underwent the same purification procedure, therefore, only the purification steps corresponding to the ^2H , ^{15}N -labelled form ($[\text{U-}^2\text{H}, ^{15}\text{N}]$ MGL-ECD preparation will be presented below.

Following cell cultures and protein expression, the resuspended cells underwent sonication. Soluble proteins were removed by a first step of ultracentrifugation. Recovered IB were then resuspended in a solution containing 2 M urea and 1% triton X-100. The mild chaotropic effect of urea, along with the detergent action, allowing the removal of both membrane and membrane-associated contaminants from IBs. After a subsequent washing step, used to remove triton which is incompatible with the use of guanidine, obtained IBs were resuspended in a solution containing 6M guanidine and 0.01% β -mercaptoethanol. The strong chaotropic effect

of guanidine disrupted water molecules interactions, resulting in protein unfolding and solubilization. The β -mercaptoethanol reduced the disulfide bridges formed within IBs to fully unfold protein chains. These IBs isolation steps were monitored using 12% SDS-PAGE. A band at the expected size was only observed in the fraction resulting from the guanidine treatment, indicating successful solubilization, whereas it was absent in fractions from earlier stages aimed at eliminating contaminants and washing (Figure 5.4).

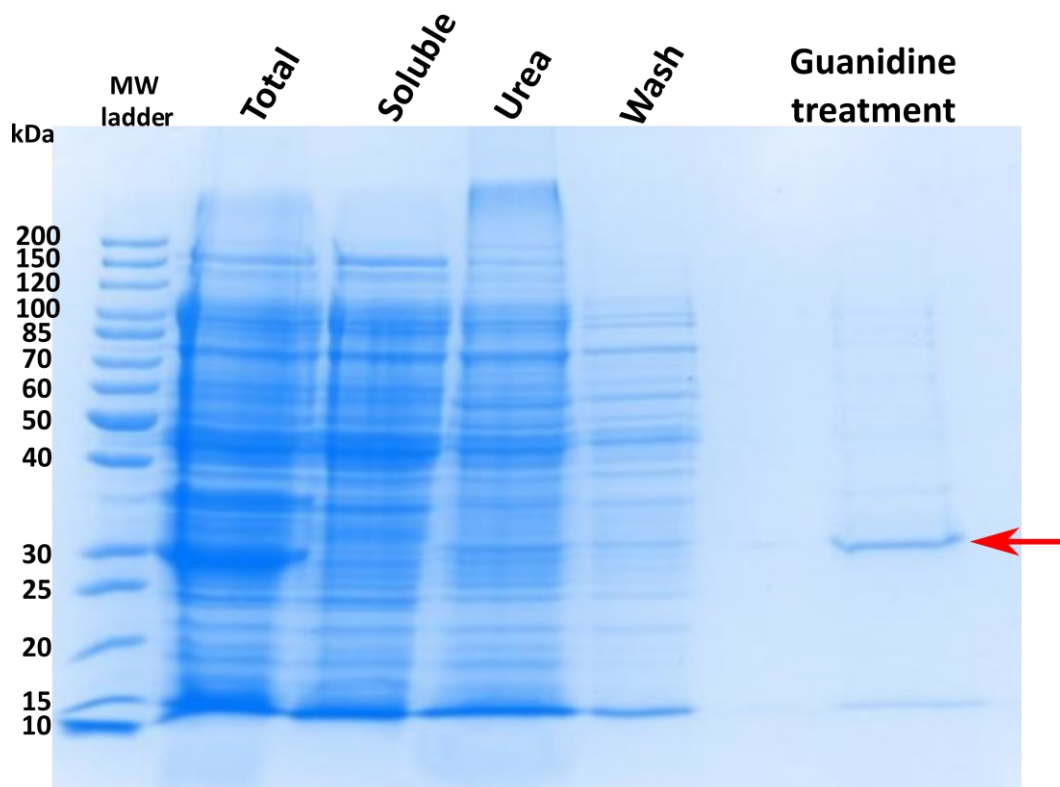


Figure 5.4: Inclusion bodies refolding steps of [U-²H,¹⁵N] MGL-ECD followed on a 12% SDS-PAGE. Cell pellets are resuspended (Total) and lysed by sonication. Inclusion bodies (IBs) are recovered by ultracentrifugation and the supernatant is removed (Soluble). IBs are resuspended in a solution containing Urea and triton detergent then subjected to a washing step. Following each step an ultracentrifugation is performed, and the supernatants were loaded in the SDS-PAGE gel (Urea and Wash). Finally, IBs were solubilized in a solution containing 6M guanidine and MGL ECD was recovered in the supernatant (Guanidine treatment, red arrow).

The sample was then diluted to 2 mg/mL (concentration estimated from MGL ECD extinction coefficient) in the solution containing 6 M guanidine and 0.01% β -mercaptoethanol to minimize protein aggregation during the refolding step. To initiate protein refolding, a rapid drop-by-drop dilution step was carried out, using buffer conditions already optimized for each lectin, containing no guanidine nor β -mercaptoethanol. Following dilution, three rounds of

dialysis were performed to decrease guanidine concentration below 100 mM to allow reoxidation of disulfide bridges. Subsequently, protein purification was achieved with a two steps protocol, first a GalNAc affinity chromatography, on a GalNAc-agarose column, followed by a size-exclusion chromatography on a Toyopearl HW-50S (Figure 5.5A). Affinity chromatography selectively isolates the MGL ECD based on its affinity for GalNAc residues, serving both purification and functional verification purposes. After affinity column elution, the protein is loaded onto the size-exclusion column to verify the sample's homogeneity and exchange the buffer for a Ca^{2+} containing buffer, essential for MGL function. Fractions below the peak are subjected to analysis on 12% SDS-PAGE, in both reduced and non-reduced conditions (with and without β -mercaptoethanol) to monitor MGL ECD's proper oligomerization state (Figure 5.5B).

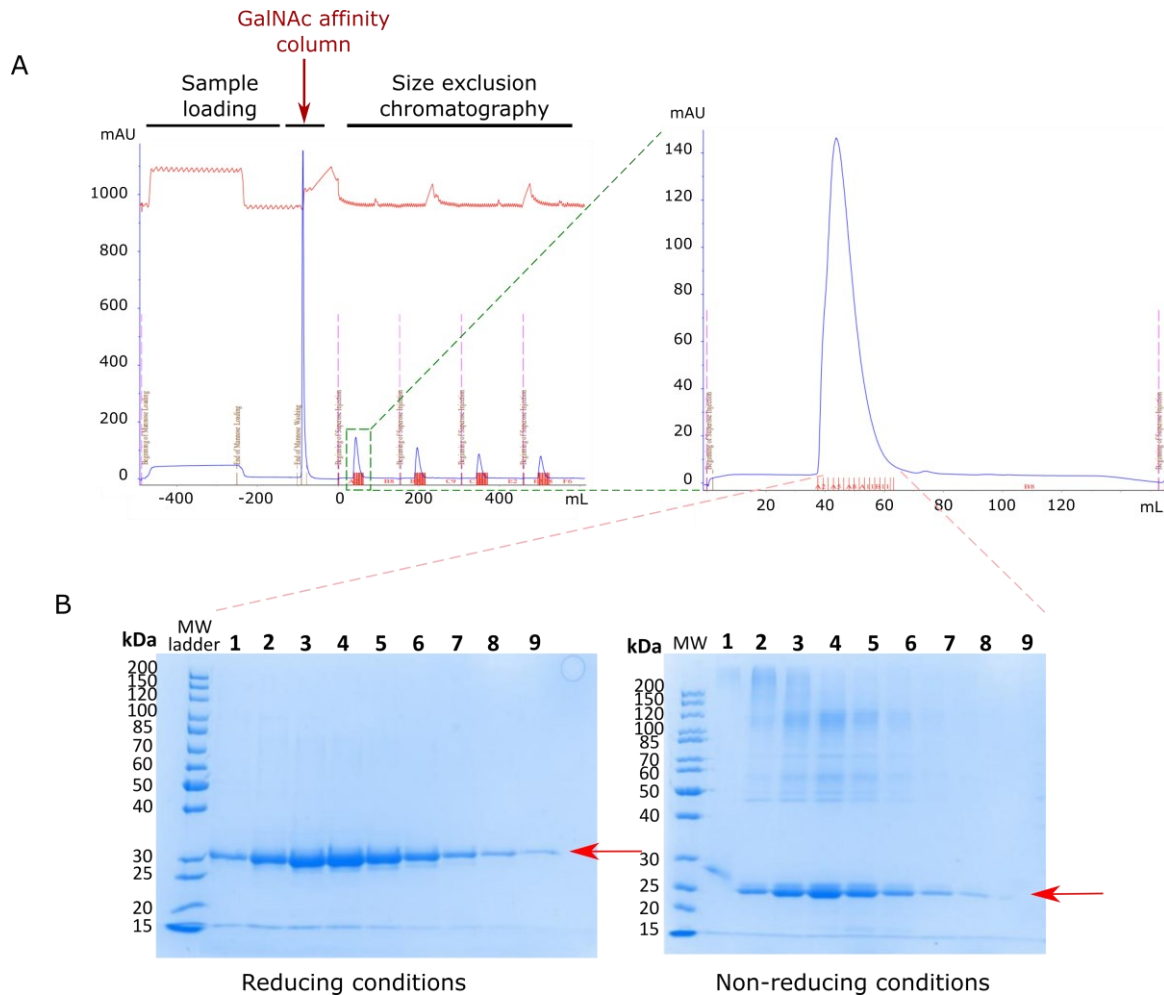
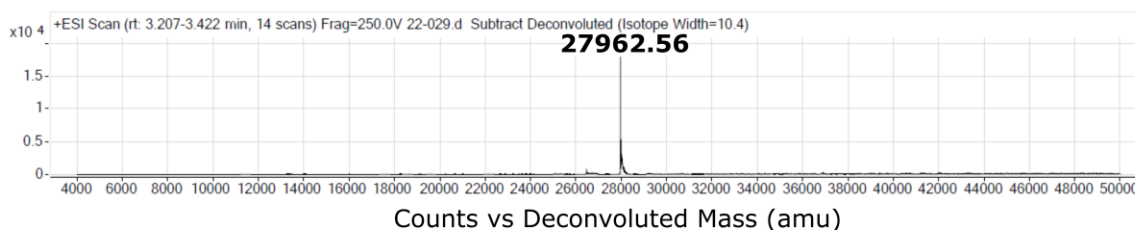


Figure 5.5: Purification of MGL ECD. (A) Full chromatogram of MGL ECD two-steps purification (left) consisting of a first GalNAc affinity chromatography followed with a size-exclusion chromatography on a Toyopearl HW-50S with a zoom on a SEC Chromatogram profile (right). (B) Representative 12% SDS-PAGE of fractions obtained following the SEC chromatography in both reduced (left) and non-reduced conditions (right).

The final sample turned out of nice purity (according to the reducing gel), but there is a proportion of oligomers which seem to be linked by disulfide bridges (according to the non-reducing gel). Species around 45 kDa which can correspond to dimers (the unreduced monomer migrating towards 22 kDa), species above 100 kDa which can correspond to tetramers, and species which do not migrate into the gel (aggregates) were observed. These species are the result of mismatches by intermolecular disulfide bridge, and could not be separated from the monomer by the Toyopearl HW-50S column. The form with intramolecular S-S bridges remains nevertheless predominant thus, fractions containing MGL ECD were pooled and concentrated.

Mass spectrometry analysis of one of the different MGL ECD preparations purified with the same procedure described above was done. The analysis was conducted on the mutant construct MGL ECD^{H262A} and confirmed the protein purity and displayed the correct molecular mass (Figure 5.6). The expected mass of this construct is 28101.86 Da (in the reduced form SH). The main mass observed is 27962.56 Da which corresponds to the protein sequence with a loss of the first methionine and an 8 Da difference with the theoretical mass corresponding to four disulfide bridges. Alongside the three disulfide bridges previously characterized within the CRD of the MGL, an additional bridge was then detected in MGL ECD in its trimeric form, presumably between the CRD segment and the coiled-coil region of the MGL ECD although no conclusions on its position can be drawn so far.

Deconvoluted spectrum



Zoom deconvoluted spectrum

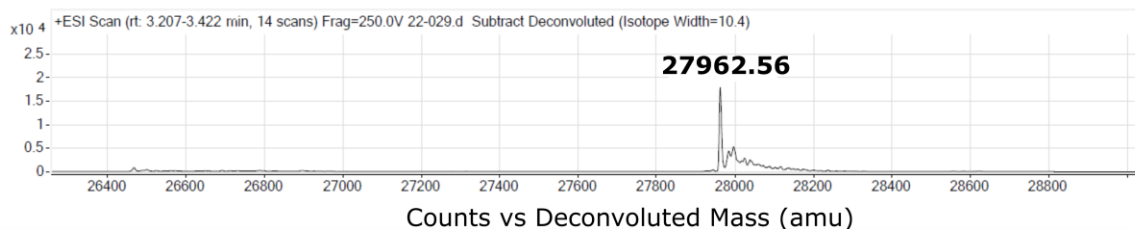


Figure 5.6: Mass spectrometry analysis of MGL ECD^{H262A}. The observed average mass of 27962.56 Da is in agreement with the protein sequence with a methionine loss and 4 disulfide bridges. Mass error: -0.11 Da, -4 ppm.

Similarly, MGL-ECD^{D245H} mutant was over-expressed in *E. coli* BL21DE3 cells in LB medium and was subjected to the same solubilization and renaturation steps described above. Since this construct has an abolished Ca²⁺ binding site and does not bind to GalNAc residues, MGL-ECD^{D245H} was instead purified using a Strep-Tactin affinity column, eluted with 2.5 mM desthiobiotin, followed with a Toyopearl HW-50S size-exclusion column operating in tandem (Annex 1).

V.2. MGL ECD binds E. coli OSs via a new binding interface

As published structural studies have been conducted on the DC-ASGPR variant, a 316 amino acids protein sequence with a long neck domain (residues Q61-T179) and a CRD domain (residues C180-H316), and given the fact that both MGL and DC-ASGPR possess identical CRDs, we have chosen to maintain the DC-ASGPR sequence numbering for the CRD part in MGL for our studies. Therefore, so that the CRD domain sequence remains between residues C180-H316, and considering the additional 24 amino acids residue in the neck domain of DC-ASGPR, sequence numbering of the MGL ECD sequence is here comprised between residues Q85-H316 (instead of Q61-H292 in the UniProt data base ID Q8IUN9-2).

MGL ECD is a large homo-trimer protein of 84 kDa and thus challenging for NMR spectroscopy. Consequently, the perdeuterated MGL ECD was used and was subjected to NMR analysis. Proteins have a high density of protons resulting in a complex and crowded spectrum with overlapping signals. By replacing hydrogen atoms with deuterium, the number of protons is significantly reduced. The remaining proton signals of exchangeable groups (e.g. amide protons) are thus detected with enhanced sensitivity, where the resulting spectrum is greatly simplified with reduced signal overlap. This atom replacement also decreases relaxation rates and homonuclear ¹H-¹H interactions (dipolar couplings), which cause severe line broadening in NMR spectra especially for large proteins, and results in spectra of improved resolution and sensitivity with narrower and sharper peaks.

The ¹H-¹⁵N correlation spectrum recorded on the uniformly ²H, ¹⁵N-labelled form ([U-²H, ¹⁵N] MGL ECD) was of high quality, given the protein's large size and elongated ECD shape, with a nice chemical shift dispersion indicative of a well-folded protein. In addition to several overlapped resonances around 8.2 ppm likely originating from the coiled-coil domain, comparison with a spectrum previously recorded on isolated CRD reveals the presence of most of the CRD resonances in the ECD spectrum (Figure 5.7). We took advantage of this close agreement between CRD signals in both the ECD and isolated CRD to assign the CRD

resonances in MGL ECD spectrum by resonances transfer. A *de novo* assignment by backbone assignment experiments would not have been possible given the low stability of MGL-ECD at the requisite temperature for triple resonances NMR spectra recording.

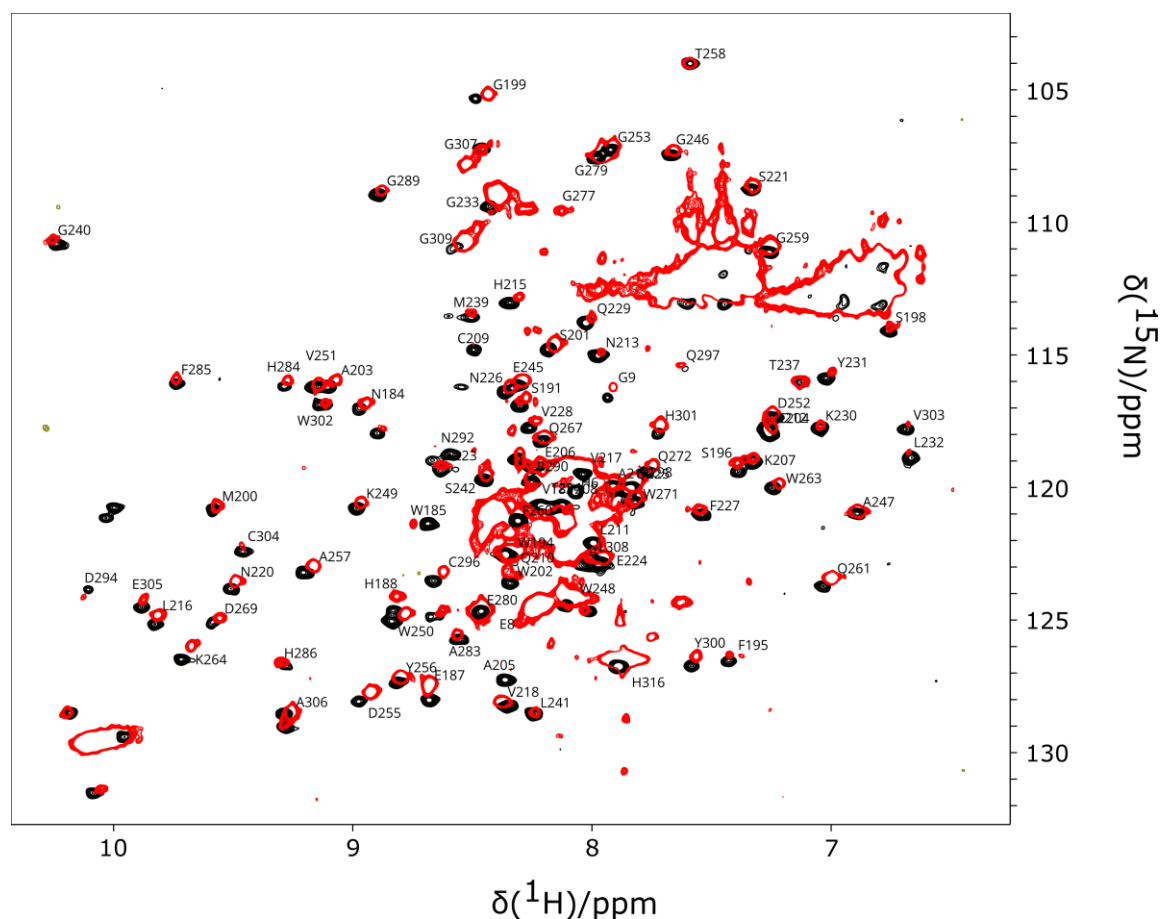


Figure 5.7: Superposition of ^{15}N - ^1H correlation spectra of MGL-ECD and MGL-CRD. MGL ECD spectrum (red) is recorded at 35°C and MGL-CRD (black) at 30°C . Assignable resonances transferred from the CRD are indicated.

Purified LOSs molecules form large vesicles in solution, which rapidly sediment. Given their unsuitability for NMR interaction studies, interaction with MGL ECD was performed using LOS derived oligosaccharides (OS) of R1 and R3 types (Figure 5.8). The experiments were performed at 35°C on an 850 MHz spectrometer. Chemical shift perturbation (CSP), corresponding to the chemical shift change in the ^1H - ^{15}N BEST TROSY spectra following OS titration were calculated as follow: $\text{CSP} = ((\Delta\delta^1\text{H})^2 + ([\Delta\delta^{15}\text{N}/6])^2)^{1/2}$. $\Delta\delta^1\text{H}$ and $\Delta\delta^{15}\text{N}$ correspond to chemical shift changes in both amide proton and amide nitrogen, respectively. Residues for which the CSP exceeds twice the standard deviation of the shift for all residues were deemed significant.

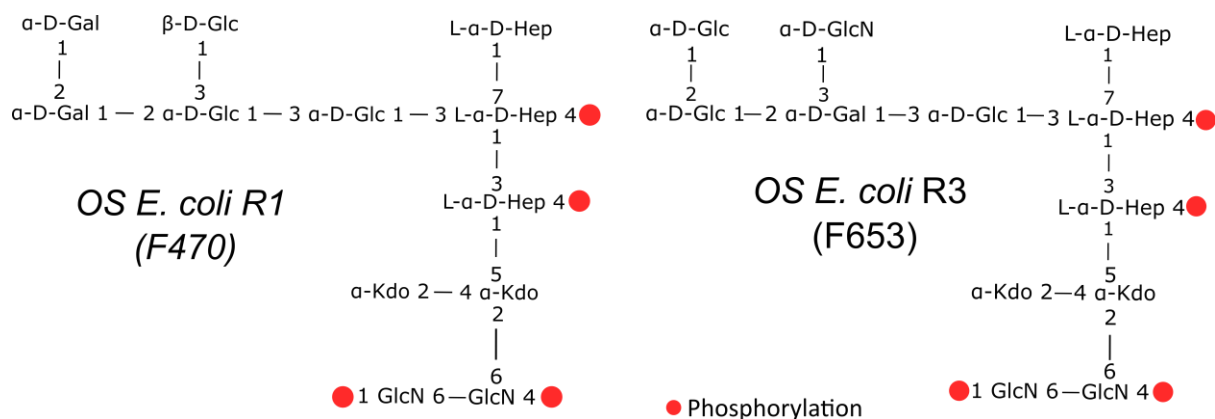


Figure 5.8: Chemical structure of *E. coli* R1 and R3 core oligosaccharides used in this study. Phosphorylation is indicated with red spheres.

The addition of increasing concentrations of OS R1 or OS R3 ligands (1 and 2 molar equivalents OS:MGL) showed CSPs for some [^1H - ^{15}N] backbone NMR resonances in a fast exchange regime. More importantly, residues experiencing strong CSPs were rather found on a surface opposite to the canonical GalNAc binding site (Figure 5.9 and Annex 2). This surface involved residues 202-216 around the $\alpha 2$ helix. CSPs induced by interactions with both R1 and R3 OSs are very similar, indicating that MGL ECD has no selectivity towards either R1 or R3 chemical structures when presented in this context (soluble derivatives). CSPs induced by the interactions with these OSs turned out to be similar to those observed on isolated MGL CRD⁸². We thus concluded that this new binding interface is involved in the binding of MGL to these *E. coli* OSs. This finding was further validated through mutagenesis experiments using two MGL mutants: MGL CRD^{D269H} and MGL ECD^{D269H}. Although their canonical QPD binding site was abolished, they still bound to both R1 and R3 glycans, as investigated by different biophysical methods including flow cytometry for MGL ECD^{D269H} and NMR for MGL CRD^{D269H}⁸². Altogether, these studies indicated that MGL's canonical binding site contributes to R1 and R3 LOS binding, but it is not its primary determinant as a new binding interface has been revealed for these *E. coli* glycans.

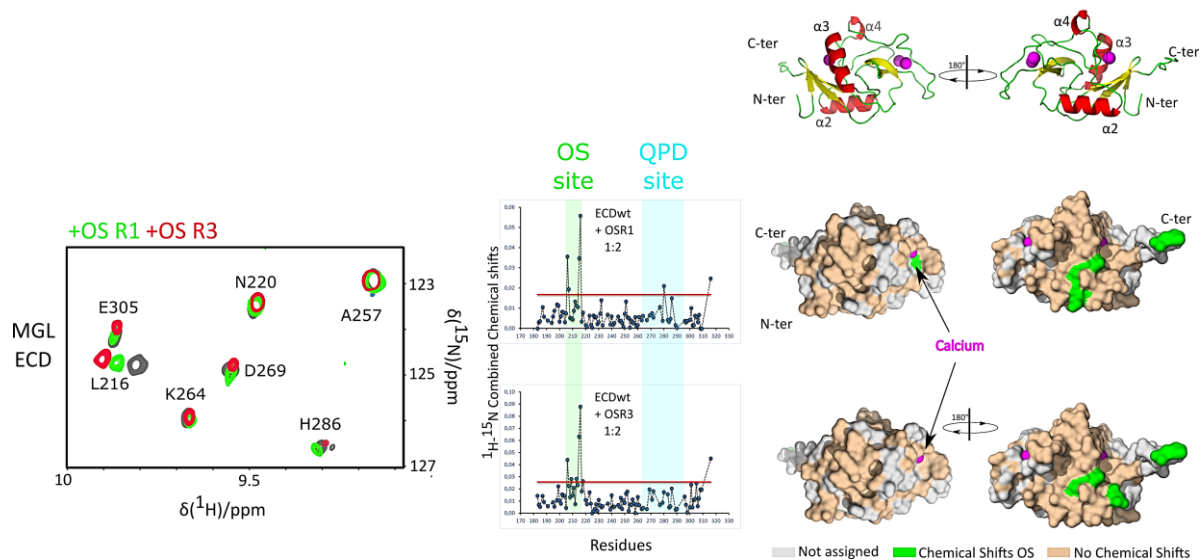


Figure 5.9: NMR spectroscopy interaction studies of MGL ECD with OS R1 and R3. Left: overlaid extract of ^1H - ^{15}N correlation spectra interaction of MGL ECD with OS R1 and OS R3. Histogram of CSPs with respect to amino acid sequence are shown in the middle. The threshold of significant CSPs is indicated with the red line. Significant CSPs are represented on MGL ECD surface upon OS R1 and R3 interaction with green. Calcium ions are indicated in magenta.

V.3. SPR interaction analysis of LOS micelles with CLRs

SPR was next used as an alternative method to further investigate MGL interaction with R1 and R3, not in their soluble OS form but as intact LOS solubilized in dodecylmaltoside (DDM) detergent micelles. Micelles were prepared after LOS R1 and R3 chemical extraction following the PCP method¹⁶⁴. LOS R1 and LOS R3 were then solubilized in DDM micelles for 15 min followed by an ultracentrifugation for insoluble material removal.

SPR direct interaction analysis were performed on MGL ECD oriented surfaces. MGL ECD was specifically biotinylated on its N-terminus using the sortagging procedure described in Achilli et al.²³⁰. The resulting protein was then grafted onto sensor chips beforehand functionalized with Streptavidin. Surface functionality was tested using Mannotriose and GalNAc conjugated to bovine serum albumin (BSA-GalNAc/BSA-Man) as negative and positive controls, respectively (Annex 3A). Injection of LOS micelles at increasing concentrations resulted in the interaction sensorgram shown below from which a titration curve could be traced using Steady State Affinity model (Figure 5.10). When using DDM-solubilized micelles of LOS R3 molecules, at higher concentrations of the analyte we observed a moderate

effect of the analyte on MGL binding around mid-association phase. This could be due to possible detergent perturbation on the coiled-coil region that influenced CRDs arrangement and interaction. However, since the binding response was re-established by the end of the association, we kept this data point for the fit. An apparent affinity constant of $\sim 15 \mu\text{M}$ for both R1 and R3 LOS was obtained. Although R1 OS has two terminal Gal residues which could be recognized by MGL, the lectin did not differentiate between R1 and R3 when presented in this context, supporting the results obtained with NMR.

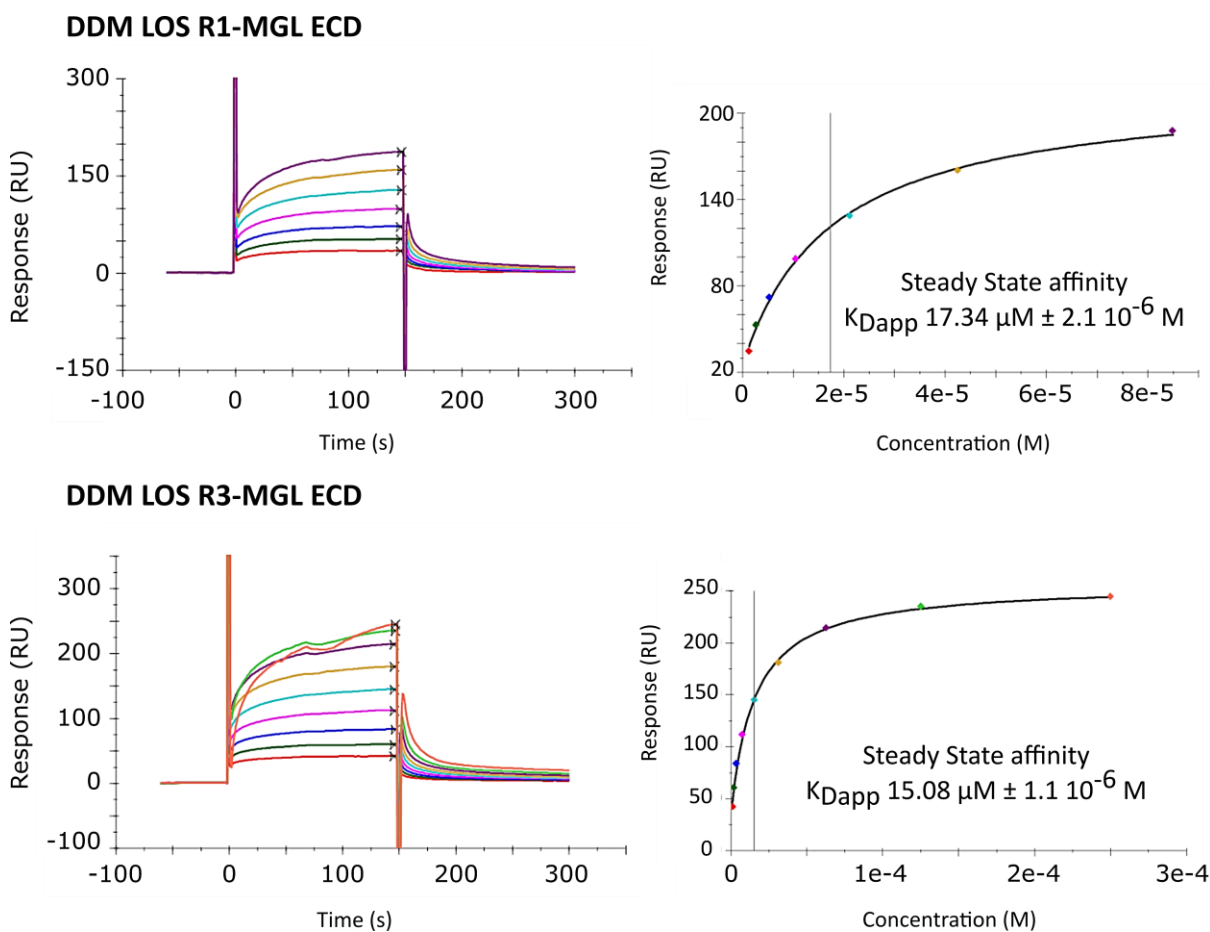


Figure 5.10: SPR direct interaction of R1 and R3 LOS reconstituted in DDM micelle with MGL ECD oriented surface. LOSs micelles are injected onto MGL ECD at increasing concentrations from 1.32 to 85 μM for LOS R1 micelles and 0.97 to 250 μM for LOS R3 micelles. The steady state binding responses were plotted against LOS concentration.

Interactions of these LOS micelles were also studied with DC-SIGN using oriented SPR surfaces. Here too, the surface functionality was tested with BSA-Man and BSA-GalNAc controls (Annex 3B). As observed for MGL, DC-SIGN showed similar binding affinities for R1 and R3, with a slightly stronger one for R3 compared with R1 ($\sim 10 \mu\text{M}$ for R3 against ~ 16

μM for R1) (Figure 5.11). Altogether, interactions of entire LOS glycoconjugates were successfully monitored on oriented SPR surfaces. Both lectins did not show any specificity towards either of the tested glycoconjugates when presented in a micellar context. However, it is important to emphasize the significant increase in the binding affinity observed compared to the results obtained with LOS derived oligosaccharides (OS) (μM using LOS micelles versus mM with isolated OSs). The increase in affinity observed with this experimental set-up can be explained by the fact that the orientation of multimeric lectins mimicking the cell surfaces can interact with multiple LOS molecules present at the micellar surface. This generated multivalent effect could be promoted through the rebinding and clustering events that could take place at the cell surface²³¹. Our finding highlights the importance of both the ligand presentation, on one hand, and the use of lectin-oriented surfaces on the other for the enhancement of the interaction strength.

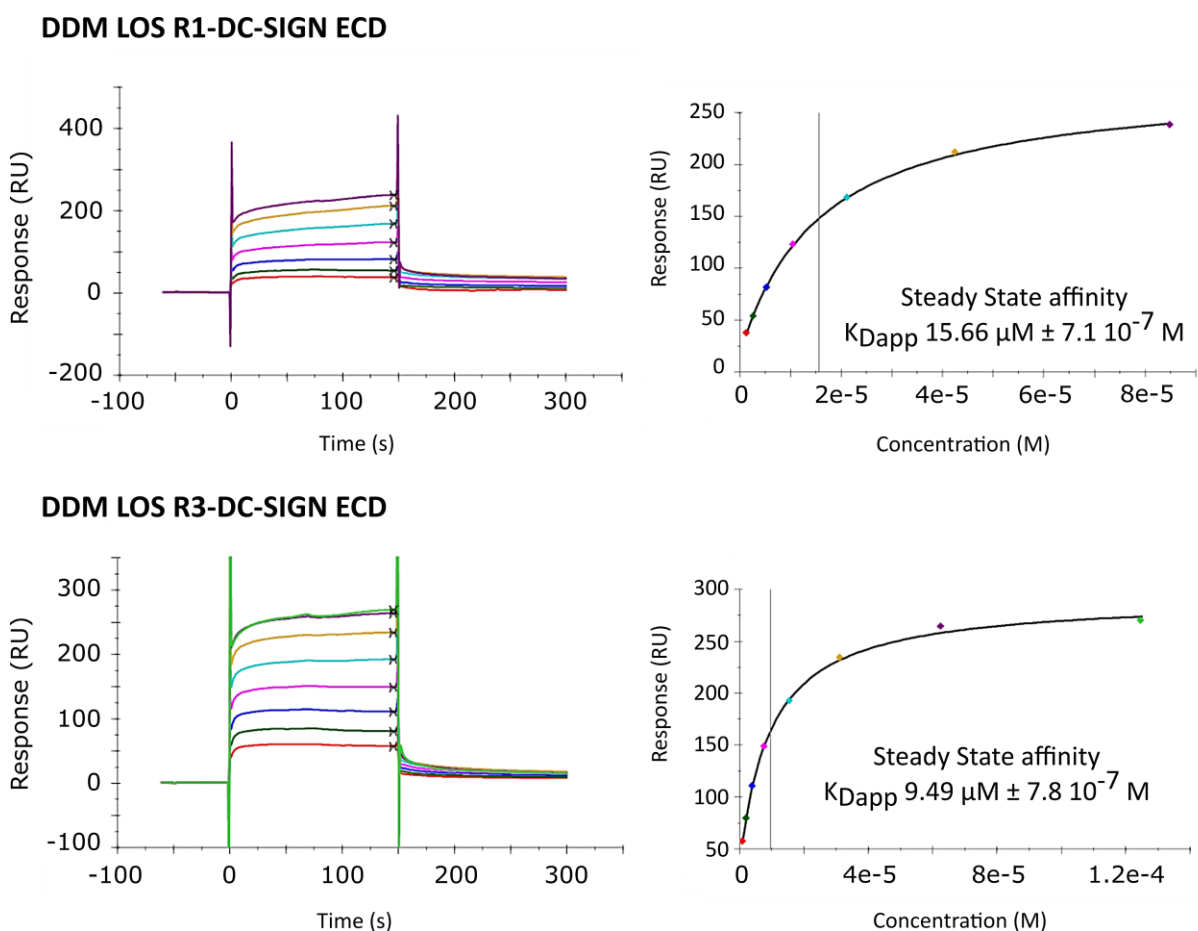


Figure 5.11: SPR interactions of R1 and R3 LOS DDM micelle with DC SIGN ECD oriented surface. LOSs micelles are injected onto DC-SIGN ECD at increasing concentrations from 1.32 to 85 μM for LOS R1 micelles and 0.97 to 125 μM for LOS R3 micelles. The steady state binding responses were plotted against LOS concentration.

V.4. Flow cytometry analysis of MGL binding onto R1 and R3 cells

MGL binding to R1 and R3 LOSs was further evaluated using whole cells and flow cytometry. Initially, MGL was first labelled with Alexafluor 647 fluorophore and then incubated with *E. coli* R1 and R3 cells. Following extensive protein washing, fluorescent cells were quantified. The fluorescence index was determined by multiplying the percentage of labelled cells by the median fluorescence, and then normalized to the OD and to 100% for MGL binding to R1 which presented maximum binding response.

Interestingly, flow cytometry data indeed reveal a difference in MGL binding preferences towards R1 and R3 core type strains (Figure 5.12, Annex 4). R1 strain has the most significant normalized fluorescence index indicative of MGL binding onto *E. coli* R1 surface. On the other hand, binding to R3 exhibited two times less fluorescence compared with R1. This suggests that the R1 core structure is better recognized than the R3 core structure, presumably due to the terminal galactoses present at the outer core of R1 placed next to each other which would enhance the binding affinity through rebinding effects.

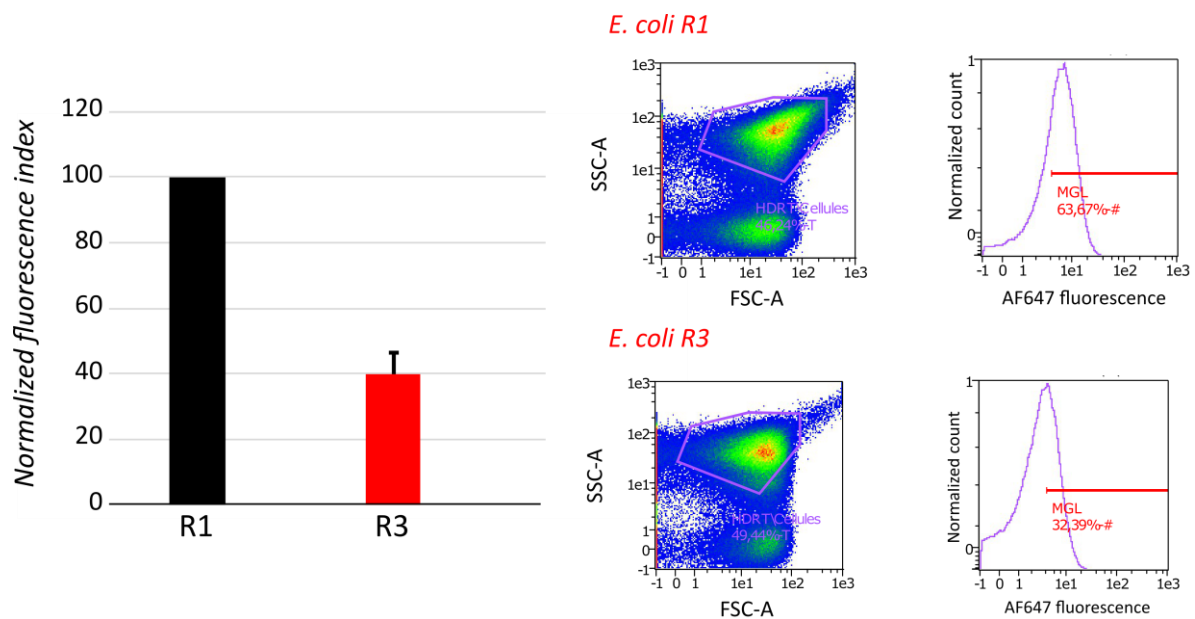


Figure 5.12: Flow cytometry quantification of MGL ECD binding onto *E. coli* R1 and *E. coli* R3 cells. MGL-bound cells fluorescence index of both *E. coli* strains is plotted and the corresponding standard deviation of the replicates is shown (left panel). Flow cytometry density plot of one of the replicates is presented in the right panel together with the count vs fluorescence plot of the selected bacterial populations. FSC: Forward scattering; SSC: Side scattering.

Given the fact that the experiments are carried out after extensive wash steps (3 washes) to which only strong binding would resist, the binding on bacteria was again quantified with and without washing steps. Similarly, the fluorescence index is calculated and normalized to 100% for MGL binding to R1. Interestingly, MGL binding to both strains without washing was quite similar, with R1 being only slightly better. However, upon washing, the binding fluorescence decreases significantly for R3 compared with R1 (Figure 5.13, Annex 5). This variation could be due to a lower affinity of the receptor for the R3 strain compared to the R1 strain and a consequently higher dissociation rate constant. This could be explained by the R3 chemical structure harbouring only a single galactose residue compared to two for R1 core OS.

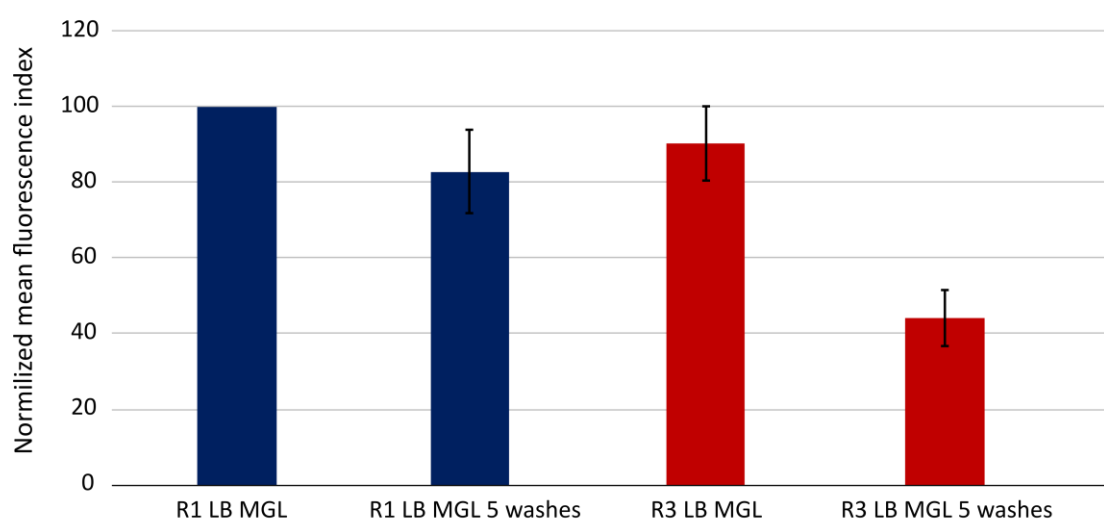


Figure 5.13: Flow cytometry quantification of MGL ECD labelled with AF647 bound to R1 and R3 cells. Quantification of MGL binding to R1 (blue) and R3 (red) cells with and without washing steps normalized to 100% for MGL binding to R1. 50% binding decreases is obtained for MGL binding onto R3 cells upon washing. Experiments were done in duplicates and standard deviations are shown.

V.5. What about MGL binding to pathogenic *E. coli* strains?

We next attempted to investigate the influence of various O-antigen structures on the interaction with MGL. Two *E. coli* strains were selected, O113:H21 and O157:H7, which present the same oligosaccharide core structures as R1 and R3, respectively (Figure 5.14). The samples were prepared similarly to the previous flow cytometry analysis, where labelled MGL ECD was incubated with *E. coli* O113:H21 and O157:H7 cells followed by excessive protein washing.

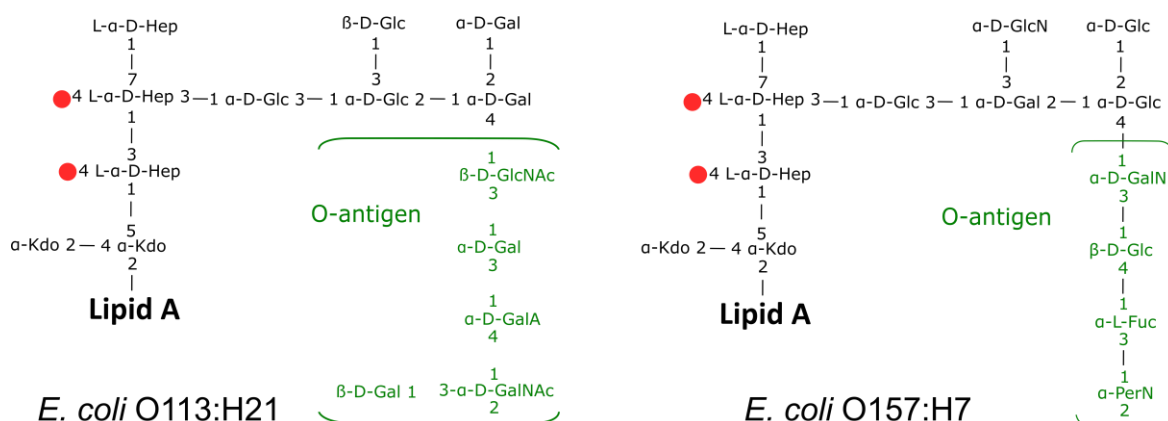


Figure 5.14: LPSs chemical structures of *E. coli* strains O113:H21 and O157:H7 LPSs used in this study. O113:H21 LPS in left and O157:H7 LPS in right. O-antigen portions are highlighted in green. Phosphorylated positions are indicated with red spheres.

Given the enrichment of *E. coli* O113:H21 O-antigen chemical structure with MGL target residues (Gal/GalNAc), we expected it to present the most pronounced fluorescence. Surprisingly, O113:H21 strain is 2 times less well recognized by MGL compared to its counterpart R1 strain (Figure 5.15, Annex 6). Different assumptions regarding this finding could be made. On one hand, wild type strains not only produce smooth type LPSs, but rather a mixture of S-LPS and LPSs with different truncations including rough type versions. This means that O113:H21 can also present a population of R1 LOSs on its surface that could be bound. This would explain the decreased interaction of MGL with this strain compared with R1 which then presents more available LOSs molecules for binding. On the other hand, if we consider only the canonical binding site of MGL and its binding to Gal/GalNAc monosaccharides, the equatorial 3-OH and axial 4-OH groups of Gal/GalNAc residues are the ones required for calcium ion binding within this binding site. However, if we take a closer look into most potential MGL ligands present in O113:H21 LPS, only the terminal Gal residue within its O-antigen moiety and one Gal residue within the outer core OS have these positions free for interaction. As these two residues are completely separated and given the fact that the O-antigen portion may completely render the core OS inaccessible for binding, this leaves MGL with only one free Gal residue. This would explain the decreased interaction of MGL with this strain compared with its counterpart R1 which presents two available Gal residues placed next to each other that would improve the binding affinity *via* possible clustering and rebinding effects. Nevertheless, in our case different parameters have to be considered: we are no longer in the mono-saccharide context, but working with bacterial surfaces covered with complex

glycans with a certain spatial arrangement. Furthermore, in addition to the canonical binding site, we also demonstrated that MGL has a second binding site which operates in a Ca^{2+} - independent manner. This site may offer MGL an expanded ligand spectrum and contribute to this strain glycans binding. As we do not know much on this site's glycan preferences, we cannot draw any conclusion regarding this point or the molecular mechanism behind this binding. We thus only assumed that the interaction response could either correspond to MGL recognition of this O-antigen residues, which is then attenuated due to the ligand spatial presentation, or represent MGL binding to the core OS which is thus significantly limited given its restricted accessibility due to the O-antigen portion.

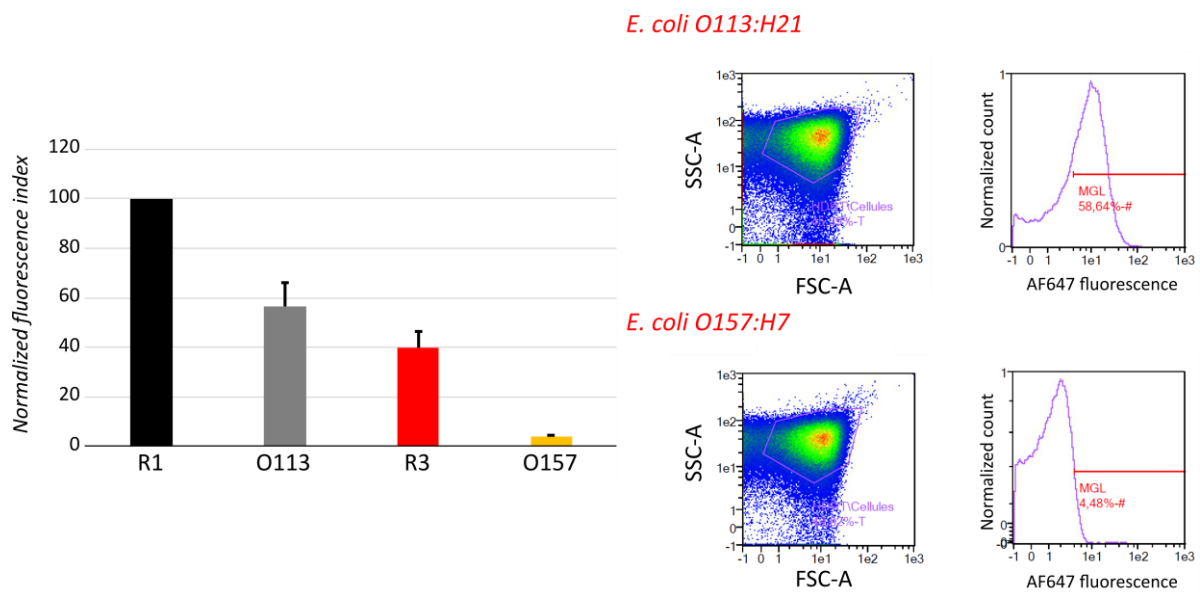


Figure 5.15: Flow cytometry monitoring of the influence of O-antigen portion on MGL ECD binding onto *E. coli* cells. (Left) Flow cytometry quantification of MGL ECD labelled with AF647 bound to R1 (black), O113:H21 (grey), R3 (red), and O157:H7 (yellow) *E. coli* cells is shown. MGL-bound cells Fluorescence index of tested *E. coli* strains normalized to 100% for MGL binding to R1 is plotted and the corresponding standard deviation of the replicates is shown. (Right) Flow cytometry density plot of *E. coli* O113:H21 and O157:H7 strains of one of the replicates together with the count vs fluorescence plot of the selected bacterial populations. FSC: Forward scattering; SSC: Side scattering.

On the other hand, the O157:H7 strain, which possess the same core structure as R3, was not bound by MGL (Figure 5.15, Annex 6). This poor recognition of the O157:H7 strain has already been reported by Jégouzo et al. using the mammalian lectin array¹⁹⁰. This can be ascribed to the poor content of residues recognized by MGL in the O-antigen portion, namely Gal or GalNAc residues. Furthermore, the presence of this long crowding sugar portion limits MGL access to the core OS part for which we know, from our studies on R3, that it is recognized by MGL. Altogether, these findings demonstrate that expression of O-antigen portions with different glycan compositions, lacking target residues or not, can influence receptors recognition, and play a protective role for bacteria to bypass immune system detection and clearance.

Implication of the O-antigen in host immune system evasion has been reported, and was found to be governed by their chemical structure variation and their surface shielding effect. Molecular mimicry between bacterial and host antigens enables the bacteria to escape from immune defences^{232,233}. This was described in *Helicobacter pylori* which displays Lewis blood group antigens in its LPS O-antigen portion which is associated with the reduced immunogenicity of its LPS resulting in immune evasion. Furthermore, the expression of modified O-antigen moieties in some bacterial strains, including *Salmonella*, was found to result in a delayed immune recognition and LPS internalization leading to bacterial survival²³⁴. Lastly, this extended polysaccharide can resist to host complement and prevent their deposition on the bacterial surface. As these components could serve as opsonin, this can result in the resistance to phagocytosis. This was observed in *Salmonella typhimurium* and *E. coli* of which, unlike wild type strains, O-antigen lacking mutants were destroyed by phagocytosis²³³.

V.6. Published article

The unique 3D arrangement of macrophage galactose lectin enables *Escherichia coli* lipopolysaccharide recognition through two distinct interfaces

Massilia Abbas^{a,1}, Meriem Maalej^{a,b,1}, Ferran Nieto-Fabregat^b, Michel Thépaut^a, Jean-Philippe Kleman^a, Isabel Ayala^a, Antonio Molinaro^b, Jean-Pierre Simorre^a, Roberta Marchetti^b, Franck Fieschi^{a,c} and Cedric Laguri^{a,*}

^aUniv. Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale, Grenoble 38000, France

^bDepartment of Chemical Sciences, University of Naples Federico II, Naples 80126, Italy

^cInstitut Universitaire de France (IUF), Paris, France

*To whom correspondence should be addressed: Email: cedric.laguri@ibs.fr

¹M.A. and M.M. contributed equally to this work.

Edited By: Josh Wand

Abstract

Lipopolysaccharides are a hallmark of gram-negative bacteria, and their presence at the cell surface is key for bacterial integrity. As surface-exposed components, they are recognized by immunity C-type lectin receptors present on antigen-presenting cells. Human macrophage galactose lectin binds *Escherichia coli* surface that presents a specific glycan motif. Nevertheless, this high-affinity interaction occurs regardless of the integrity of its canonical calcium-dependent glycan-binding site. NMR of macrophage galactose-type lectin (MGL) carbohydrate recognition domain and complete extracellular domain revealed a glycan-binding site opposite to the canonical site. A model of trimeric macrophage galactose lectin was determined based on a combination of small-angle X-ray scattering and AlphaFold. A disulfide bond positions the carbohydrate recognition domain perpendicular to the coiled-coil domain. This unique configuration for a C-type lectin orients the six glycan sites of MGL in an ideal position to bind lipopolysaccharides at the bacterial surface with high avidity.

Significance Statement

The surface of bacteria is a marker of their presence when invading a host, and gram-negative types are decorated with lipopolysaccharide (LPS) carbohydrates. In this report, the recognition of LPS from *Escherichia coli* bacteria by a sugar-binding protein (lectin) present at the surface of human immune cells is described. Using a multidisciplinary approach, the presence of an unforeseen sugar-binding site at the surface of the protein was demonstrated. The tridimensional arrangement of the lectin, determined by a combination of bioinformatics and structural biology methods, explains how its sugar-binding sites allow a very strong binding to the bacterial surface. These findings illustrate how this immunity protein can recognize pathogenic bacteria with very diverse carbohydrates at their surface.

Introduction

The outer membrane of gram-negative bacteria is compositionally asymmetric with lipopolysaccharides (LPSs) covering most of its surface (Fig. 1A), while phospholipids compose the inner leaflet. LPSs form a highly impermeable barrier and are critical in bacterial virulence (1); their structural variability and tight assembly protect bacteria against uptake of antimicrobials and enable evasion from host defenses. Constant transport and maintenance of LPS in the outer membrane are critical in the survival of bacteria. LPSs are composed of three moieties: the lipid A formed by *N*- and *O*-acylated di-glucosamine, the core oligosaccharide (core OS), and *O*-antigen polysaccharide repeat (Fig. 1B). These complex glycolipids are detected by the immune system through the lipid A via the well-described LBP-MD2-TLR4 cascade (2) and by the

caspase system in the cytoplasm (3). Antibodies directed against the glycan moieties, core OS (4), and *O*-antigen polysaccharides are also produced by the immune system to modulate bacterial infections (5). Another protein family present on antigen-presenting cells, C-type lectin receptors (CLRs), has been shown to bind sugars from the core OS of LPS (6–8). CLRs are key immunity receptors, which recognize a plethora of pathogen glycans (9), and the interaction of these CLRs with their ligands, discriminating nonself from self-molecular motifs, allows dendritic cells to modulate the immune response toward either activation or tolerance (10). Macrophage galactose-type lectin (MGL) is a trimeric type II CLR expressed on the cell surface of macrophages and dendritic cells (Fig. 1D). It mediates interactions between endothelial and cancer cells (11) but also recognizes microbial glycans. Its

Competing Interest: The authors declare no competing interest.

Received: March 25, 2023. **Accepted:** September 14, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

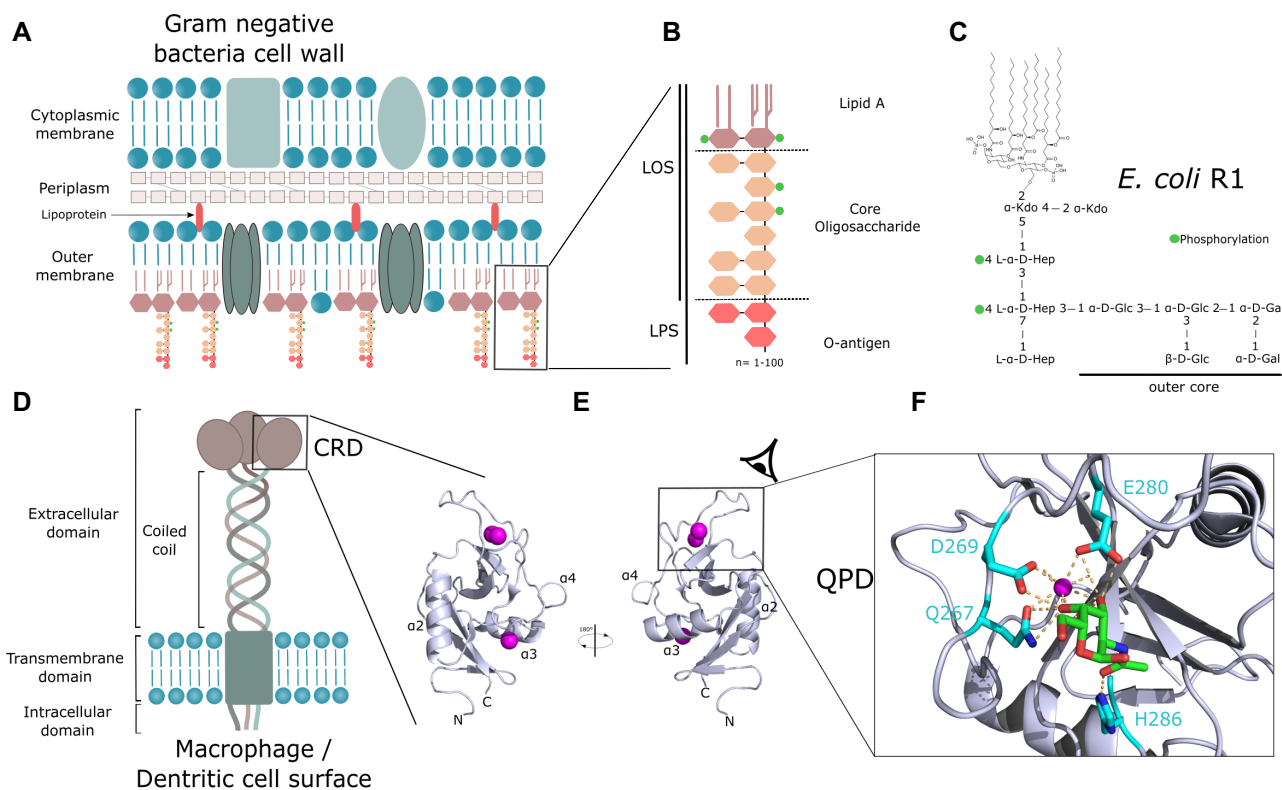


Fig. 1. Organization of gram-negative bacteria cell wall and of MGL. A) General structure of gram-negative bacteria cell wall. B) LPS composing the outer leaflet of the outer membrane. C) Structure of *E. coli* R1 LOS mainly used in this study. D) Domain organization of MGL at the surface of antigen-presenting cells. E) Structure of CRD domain of MGL. F) Close-up view on GalNAc sugar bound to the calcium-binding site (PDB:6PY1). Calcium ions are shown as spheres.

main role appears to be an immunomodulatory activity, reducing excessive inflammatory responses. So far, MGL has been described to recognize *Staphylococcus aureus*, *Campylobacter jejuni*, *Klebsiella pneumoniae*, *Neisseria gonorrhoeae*, *Bordetella pertussis*, and *Mycobacterium tuberculosis* (12–15).

MGL is a transmembrane protein composed of an intracellular signaling domain, a transmembrane domain, a coiled-coil trimerization domain, and a C-terminal carbohydrate recognition domain (CRD) (Fig. 1D). The CRD fold is highly conserved in C-type lectins and is organized as a double-loop structure (Fig. 1E) stabilized by at least two conserved disulfide bridges. The overall domain is a huge loop in itself with its N and C terminus joined together, thanks to the first disulfide bridge, which contains another loop (the so-called long loop region) also stabilized by the second conserved cysteine bridge (16). Some C-type lectin domains, including MGL, possess an additional N-terminal β -hairpin that is stabilized by a third cysteine bridge conserved in these long-form subtypes of CRDs. The domain presents a mixed α/β -fold and a large proportion of loops with undefined secondary structures (Fig. 1E). For most of the CLRs reported, glycan-binding site is calcium dependent and characterized by a tripeptide motif (EPN/QPD) and residues from the adjacent β -strand that assume metal coordination (17). MGL possesses a QPD (267–269) motif characteristic of recognition of glycans with terminal galactoses (Fig. 1F). The X-ray structure of human MGL-CRD (18) in complex with galactose-containing ligands shows two galactose ring hydroxyl groups 3 and 4 bound to the calcium ion. Additionally, H286 is proposed to be responsible for selectivity toward *N*-acetyl through a water-mediated hydrogen bond (19). MGL binds preferentially to terminal *N*-acetylgalactosamine residue and presents, for a

C-type lectin, an unusually low (μM) dissociation constant for the monosaccharide (20). The interaction of MGL with terminal galactoses from the core OS was shown for *C. jejuni* LPS (14) and for *E. coli* R1 type core OS (Fig. 1C) (21).

In this work, we have investigated MGL binding to OSs isolated from deacylated LPS or to native LPS directly exposed on whole cells. Our results show that in the trimeric oligomerized form, the CRD of MGL adopts a specific 3D arrangement that allows a unique presentation of its six glycan-binding sites (two per CRD), composed of the canonical QPD calcium-binding motif and a newly described interaction site.

Results

MGL extracellular domain strongly binds to bacterial surface, independently of the QPD motif

MGL extracellular domain (ECD) was shown by NMR to interact with the terminal galactoses of *E. coli* R1 type core OS. To establish MGL binding in the context of R1 OS assembled at the cell surface, interaction of MGL-ECD was tested with live bacteria. *Escherichia coli* bacteria exhibit variable structures of the core OS, so we chose to compare R1 and R3 types (Fig. 1C; Fig. S1) because they represent together more than 80% of *E. coli* strains including enterohemorrhagic species (22). Two bacterial strains carrying R1 and R3 core OS structures but no O-antigen, respectively F470 and F653, were thus compared for MGL interaction. MGL-ECD was labeled with Alexa Fluor 647 (AF647), incubated with *E. coli* bacteria, and excess protein was washed. Bacteria were imaged by fluorescence microscopy. F470 bacteria were significantly labeled at their surface by MGL while F653 showed no labeling, confirming that

MGL can recognize R1 core OS on cells (Fig. 2A). In order to ascertain that the interaction with the LPS observed was specific, the interaction with R1 cells was reproduced in presence of 10 mM GalNAc, that possess a low micromolar affinity for MGL, as a competitor and quantitatively assessed MGL binding by flow cytometry (Fig. 2B; Fig. S2). We found that GalNAc at high concentration could not significantly compete to the binding of MGL to R1 presenting cells.

The inability of the GalNAc monosaccharide to compete with the MGL binding to F470 cells could be ascribed to the multivalency of the interaction between the MGL trimer and R1 OSs presented on the cell surface. We thus designed a mutant of a key residue of MGL-CRD that would abolish MGL carbohydrate-binding capacity. D269, part of the conserved QPD motif (Fig. 1F), is involved in calcium-mediated binding of GalNAc to MGL (20), so we decided to produce a D269H mutant to have a steric and electrostatic inhibition of the interaction with Ca^{2+} ion in canonical carbohydrate-binding site. MGL-ECD^{D269H} labeled with AF647 was thus incubated with F470 cells and imaged (Fig. 2C). We surprisingly found that MGL^{D269H} was still able to significantly bind bacteria. This was quantified by flow cytometry that showed only a 30% decrease in binding of the D269H variant to cells (Fig. 2B; Fig. S2) with little additive effect upon addition of 10 mM GalNAc. The behavior of this variant and the inability of GalNAc to inhibit significantly the binding suggest that, while the QPD motif is contributing to the interaction with R1 at the cell surface, it is not the main determinant of the interaction.

MGL strongly binds to R1 core OS on cells and, while the integrity of the QPD motif contributes to the interaction with R1 core OS, it is not the main determinant of the interaction. We thus hypothesized the existence of a secondary glycan-binding site in MGL and investigated its localization by NMR.

MGL-CRD binds to LPS-derived OSs through a new binding surface

MGL-CRD and its binding to GalNAc and tumor-associated glycopeptides were previously characterized by NMR, X-ray crystallography, and molecular dynamics. Those studies show a clear involvement of the QPD motif, with a particular contribution of H286 in the recognition of the N-acetyl moiety (18, 20). MGL-CRD^{wt} and MGL-CRD^{D269H} have been produced and analyzed by ¹H-¹⁵N NMR spectroscopy to localize the binding site of LPS-derived OS. Wild-type MGL-CRD shows a spectrum similar to the one already published. D269H variant ¹H-¹⁵N correlation spectrum is also characteristic of a well-folded protein and comparable to the wild-type spectrum (Fig. S3). Backbone resonances of wild-type and D269H variant were assigned and used to predict their secondary structure content. It confirmed that MGL-CRD^{D269H} contains the same secondary structure elements than the wild-type protein (Fig. S3). The mutation, by abolishing the proper coordination of the calcium ion, probably destabilizes the whole GalNAc-binding site. Therefore, the resonances from residues 265–282 remained unassigned in D269H variant.

First, the binding to GalNAc sugar was assessed for both proteins. 2D ¹H-¹⁵N correlation experiments show resonances, each one corresponding to the amide frequencies of individual amino acids. Addition of a ligand perturbs the amide frequencies at the vicinity of the binding site and can be good reporters of both the affinity and the amino acids involved in the binding. As reported by Diniz et al. (20) MGL-CRD binds strongly to GalNAc in the characterized binding site between residues 264 and 296, with strong chemical shift perturbations (CSPs) of D269 and H286 amide

resonances (Fig. 3A; Fig. S4). MGL-CRD^{D269H}, as predicted, does not show any CSP upon binding to GalNAc (Fig. S5), consistent with its inability to bind to GalNAc affinity column during purification.

Lipooligosaccharides (LOS) assemble into large vesicles in solution that rapidly sediment and are not suitable to perform interactions by NMR. Soluble LOS-derived OSs of R1 and R3 types (Fig. 3B; Fig. S1) were then produced by chemical deacylation of LOS (21). The interaction of CRD^{wt} and CRD^{D269H} was then tested with R1 and R3 OSs (Fig. 3; Figs. S4, S5, and S8). Interaction with OS R1 showed CSP of the CRD^{wt} ¹H-¹⁵N resonances on a fast exchange regime with respect to NMR timescale with no saturation of the binding even at high OS concentration, suggesting a weak affinity ($K_d \geq 5$ mM). Furthermore, residues of the CRD experiencing high CSP upon OS R1 binding lie on a surface opposite to the GalNAc-binding site, in green in Fig. 3A, and involve residues 202–216 around the $\alpha 2$ helix. The same interaction performed with the D269H variant showed a very similar interaction site opposite from the QPD motif. We thus postulate that the new interface perturbed by OS R1 is responsible for the binding of MGL to F470 *E. coli*. As a control, we also tested the binding of the CRD^{wt} and D269H variant to OS R3 in the same conditions (Fig. 3; Fig. S8). OS R3 caused very similar CSP at the surface of CRD^{wt} and CRD^{D269H}. The new interaction surface of MGL involved in glycan binding does not show specificity for R1 core OS on the contrary to results obtained on cells. The configuration of the NMR interactions is very different from the *in vivo* experiments; the CRD domain is used instead of the ECD, and the OSs are free in solution and are not presented on the cell surface as multivalent ligands. In order to confirm that the binding observed on the isolated CRD also applies to the CRD in the context of the trimeric ECD, we investigated the ECD by NMR.

The ECD of MGL is a large protein (homotrimer of 84 kDa) for NMR spectroscopy due to signal broadening arising for long molecular tumbling correlation times. The protein was thus expressed and purified as a perdeuterated version. ¹H-¹⁵N correlation spectrum of ²H, ¹⁵N-labeled MGL-ECD is of high quality considering the protein size and elongated shape and is characteristic of a well-folded protein. When comparing the ¹H-¹⁵N resonances observed on spectra recorded with isolated CRD, it is apparent that the footprint of the CRD domain is present in the ECD of MGL (Fig. S6). Several additional overlapped resonances can be observed around 8.2 ppm in the proton dimension and probably arise from the coiled-coil domain. The low stability (several days) of MGL-ECD at the temperature needed to record NMR spectra (above 35°C) did not allow its *de novo* assignment by backbone assignment experiments. The good ¹H and ¹⁵N agreement between CRD signals in ECD and isolated CRD permitted the transfer of most assignments from CRD to ECD (Fig. S6). MGL-ECD was thus titrated with increasing concentrations of OS R1 and OS R3. CSPs induced by the interaction with R1 or R3 OS are very similar to those observed with isolated CRD, though the surface involved is not as extended (Fig. 3A; Fig. S7). This suggests that the assembly of the CRD domain in the full-length ECD has no influence on the selectivity of MGL toward either R1 or R3 chemical structure when interaction occurs with isolated OSs. To confirm these observations with an alternative method, R1 and R3 LOS were solubilized in dodecylmaltoside (DDM) detergent micelles and flowed over MGL-ECD specifically oriented by immobilization through its N terminus by surface plasmon resonance (SPR) (Fig. 3B; Fig. S7). Fitting of the respective sensorgrams at equilibrium produced an apparent affinity constant of about

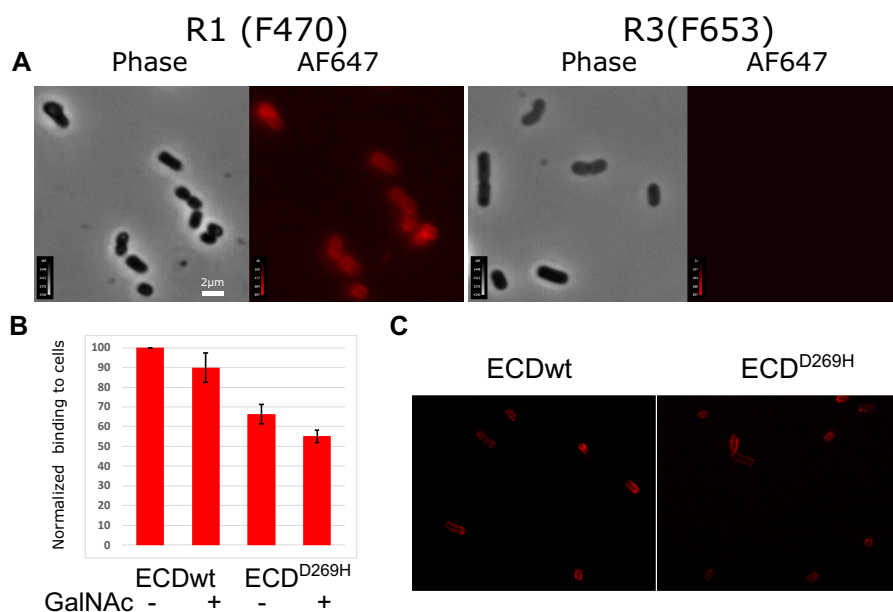


Fig. 2. MGL-ECD binds specifically to R1 presenting *E. coli* cells independently of the QPD motif. A) Phase contrast and epifluorescence microscopy images of AF647-labeled ECD incubated with R1 (left) or R3 (right) presenting *E. coli* cells. B) Flow cytometry quantification of MGL-ECD wt and D269H variant labeled with AF647 bound to R1 cells in the presence or absence of 10 mM GalNAc competitor. C) Confocal fluorescence image corresponding to conditions in B), showing strong MGL-ECD^{D269H} binding to cells.

15 μ M for both R1 and R3 LOS, confirming the lack of specificity toward the OSs when not presented as a surface.

The presence of two different glycan-binding sites at the surface of the MGL-CRD on two opposite surfaces is unprecedented in C-type lectins. It suggests that in the ECD, both sites are accessible to bind their ligands, and we thus investigated the global arrangement of the CRDs in MGL-ECD.

MGL-CRDs are oriented perpendicular to the coiled-coil domain and can present six sugar-binding sites to bacterial surfaces

The structure of MGL-ECD is unknown, and we studied its overall structure by small-angle X-ray scattering (SAXS). This method enables to assess the size and shape of a macromolecule in solution, at a low resolution. MGL-ECD SAXS curve confirms the presence of a trimeric protein with an estimated molecular weight (MW) of 94 kDa (vs 84 kDa theoretical MW) and a gyration radius of 5.6 nm, suggesting an elongated protein (23). Calculation of pairwise distribution, $P(r)$, showed a maximum interatomic distance of 17 nm (Fig. S9) consistent with the expected elongated shape of the ECD. $P(r)$ was used to calculate an envelope of MGL-ECD (see Materials and methods section). The envelope (Fig. 4A) is characterized by an elongated structure, corresponding to the coiled-coil domain, with three large bulges on its side that can be ascribed to the CRDs. The SAXS-derived envelope does not allow to orient at an atomic scale the CRD, but the location of the bulges suggests that the CRD domains are perpendicular to the coiled-coil domain. This orientation would be significantly different from an about 120° angle observed between coiled-coil and CRD domain observed for langerin or MBP trimers (24, 25).

To position the CRD into the SAXS envelope, MGL-ECD models were generated with the AlphaFold structure prediction protocol (26, 27) (see Materials and methods section). This method has provided atomic-scale prediction of protein structure of unprecedented accuracy with a combination of machine learning and

evolutionary data. The models show a long N-terminal coiled-coil domain (N86-N169) followed by the CRD (C181-H316). The arrangement of the CRD relative to the coiled-coil domain is variable and allows to sort the models into two clusters. The lack of well-defined interdomain contacts can be explained by low AlphaFold per residue score (pLDDT) at the interface and little interactions predicted in the prediction alignment error matrix (Fig. S10). One new disulfide bond is nevertheless predicted in all models between coiled-coil (C162) and CRD (C180) (Fig. 4D; Fig. S10). This disulfide bond is consistent with mass spectrometry analysis of MGL-ECD, which displays an 8 Da difference with the theoretical mass, corresponding to a total of four disulfide bonds (Fig. S11). The two cysteines involved are also strictly conserved in the MGL family in mammals (Fig. S12; Table S1), and the disulfide bond at the corresponding position was shown experimentally in the homologous protein asialoglycoprotein receptor 1 (28).

The two clusters of models are different in the orientation of the CRD domains with an almost 180° rotation around the C160–C182 disulfide bond (Fig. S10). To determine which cluster corresponds better to the conformation in solution, models were evaluated against experimental SAXS data. SAXS curves were back-calculated from the models and compared with the experimental one (Fig. 4B; Table S2 and Fig. S13). Cluster 2 structures show systematically a better fit compared with cluster 1. The best matching structures of each cluster can also be inspected visually by adjusting the structures into the SAXS-derived envelope (Fig. 4C; Fig. S14). The cluster 2 models are in the best accord to the SAXS data, and the structure with the lowest χ^2 with respect to the SAXS curve was retained for analysis (Fig. 4B–D).

The two glycan-binding sites of the CRD, the canonical QPD motif and the newly described OS-binding site, can be represented on the surface of the MGL model (Fig. 5A) in cyan and green, respectively. The orientation of the CRD is such that QPD and OS sites from two neighboring CRDs face each other. If we consider that the most likely configuration of MGL binding to the bacterial surface would be perpendicular to the membrane, the CRDs are

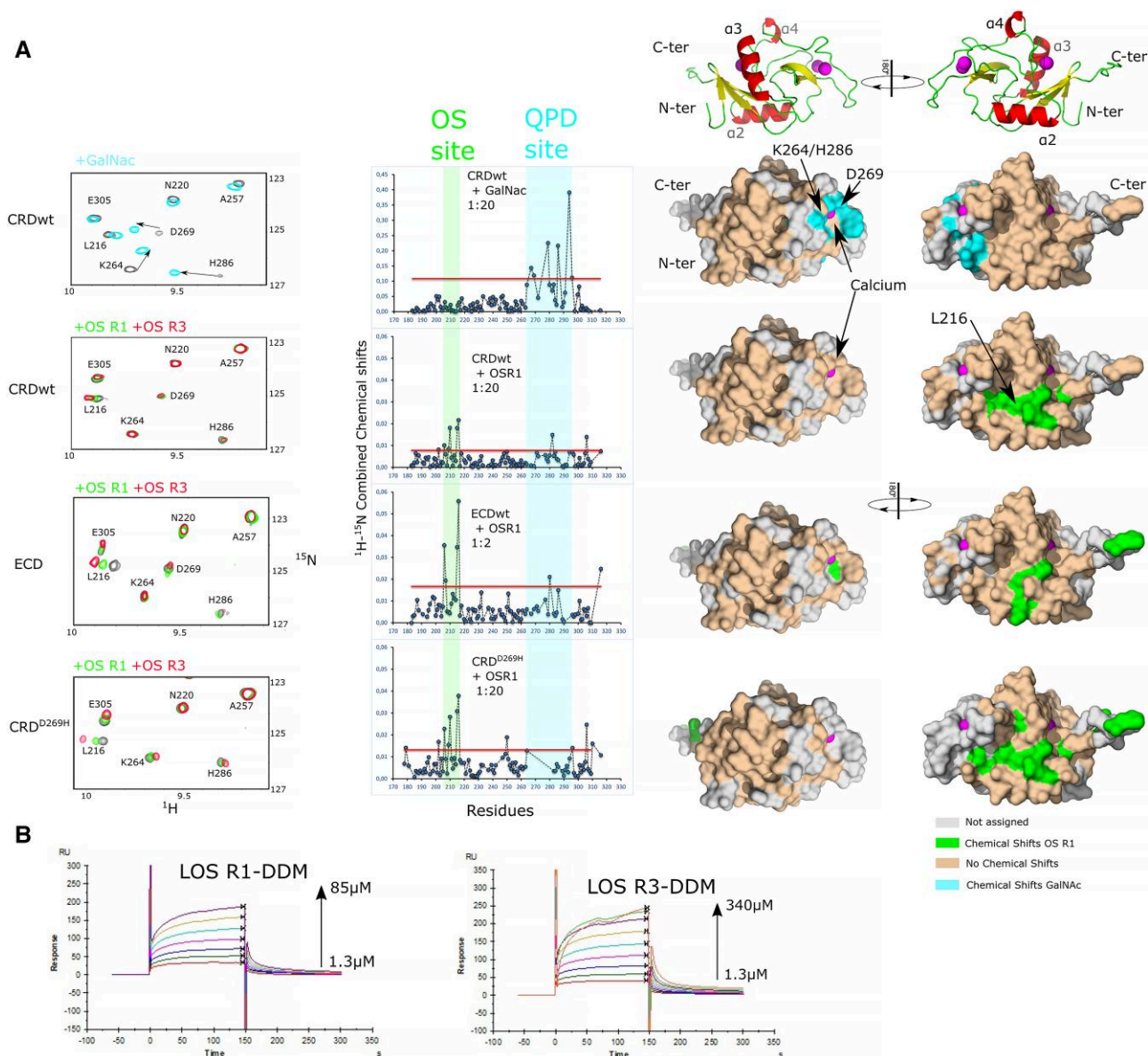


Fig. 3. GalNAc- and LPS-derived OSs interact on two opposite surfaces of MGL. **A**) Left: extracts of ^1H - ^{15}N correlation spectra of the CRD, CRD^{D269H}, and the CRD in the full ECD upon interaction with GalNAc, OS R1, or OS R3. Middle: CSP of the corresponding interactions represented with respect to amino acid sequence. The red line marks the threshold of significant CSP. Right: Significant CSP represented on the CRD surface upon interaction with GalNAc or OS R1. **B**) SPR interaction of LOS R1 and LOS R3 in detergent micelles with immobilized MGL-ECD.

able to present up to six glycan-binding sites to LPS core OS (Fig. 5B). In that configuration, even if the affinity of MGL for isolated core OS is low, the avidity of the interaction would ensure a tight binding to the surface, consistent with our observations on bacteria presenting R1 core OS.

Discussion

Glycoconjugates are present at the surface of most cells, as well as in extracellular matrices and biofilms. In complex multicellular organisms, the sugar environment is very rich and heterogeneous. The immune system must recognize friends from foes and clear pathogens but also tailor its response to avoid excessive inflammatory response. The recognition of pathogens vs commensals is critical and also relies on subtle variations of microbial glycome. MGL has been reported so far to recognize several bacterial

pathogens, with different cell wall structures, through their surface glycans.

While MGL attaches strongly to *E. coli* surface presenting R1 type core OS, this binding is largely independent of the QPD GalNAc-binding site. We could show that a second interface, opposite to the QPD-binding site, binds LPS core OS. Several examples exist of secondary binding sites in C-type lectins. They can be located adjacent to the conserved calcium-binding site to extend the binding interface and confer specificity toward a given ligand like observed for trehalose dimycolates for Mincle (29) or on a more remote site like for heparin for langerin (30) and through cooperativity for DC-SIGN (31, 32). The presence of a second binding site completely opposite to the canonical binding site is nevertheless unusual. We suggest that this is correlated with the peculiar 3D arrangement of MGL-CRDs compared with other multimeric C-type lectins. Other trimeric C-type lectins

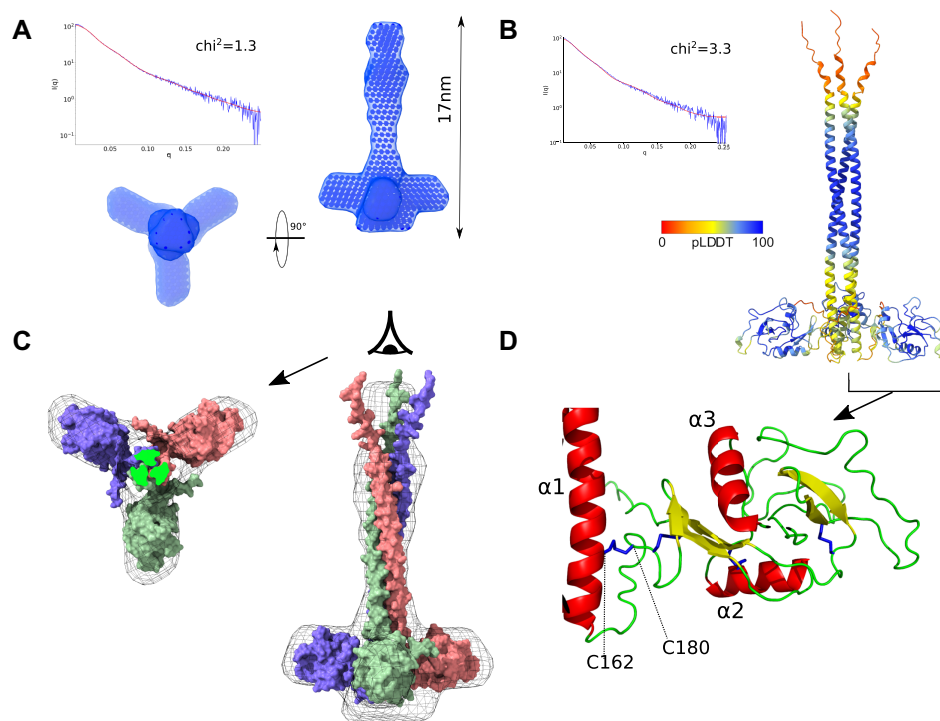


Fig. 4. Combined SAX-AlphaFold model of MGL-ECD. A) SAXS of MGL-ECD with SAXS curve (top left blue) and the corresponding fit (red) of the SAXS envelope, calculated from $P(r)$ distribution, shown as surface from side and N terminus of coiled-coil view. B) AlphaFold model with the best correspondence to SAXS curve, colored by pLDDT score. The calculated SAXS curve of this model is shown (red) compared with the experimental curve (blue). C) Best AlphaFold model of MGL-ECD adjusted into the SAXS envelope (in mesh) in side and from the N terminus of coiled-coil view. D) Close-up view on the C162–C180 disulfide bond orienting the CRD in the best AlphaFold model.

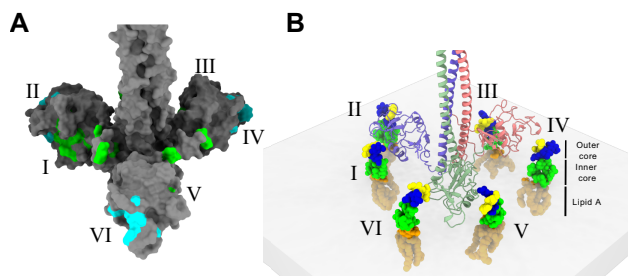


Fig. 5. Combination of SAXS and AlphaFold defines a CRD arrangement that presents up to six accessible glycan-binding sites (I to VI). A) Representation of the two glycan-binding sites, with the NMR CSP of GalNAc (cyan) and OS R1 (green) on the best SAXS-AlphaFold MGL structure. B) Schematic view of six R1 LOS molecules, with an orientation similar to what is found at the bacterial surface, facing the six glycan-binding sites of MGL-ECD.

like langerin or mannose-binding protein (MBP) (24, 25) adopt a compact arrangement of their CRDs (Fig. 6) with their canonical binding sites accessible at the extremity of the proteins. Their calcium-binding sites lie within 50 Å of each other compared with about 80 Å for MGL. It allows MGL to target surfaces with much distant glycan epitope. Furthermore, this extended conformation makes the C-terminal loop of the coiled-coil neck domain, connecting it to the CRD, accessible at the surface and could contribute to glycan binding (Fig. 6). This region of the protein varies between isoforms 1 and 2 of human MGL (Fig. 6; Fig. S12) with insertion of three additional residues (G171–E172–E173) in isoform 2 (this study). These residues could participate to the interaction of MGL with a bacterial surface but could also be important for the orientation of the CRD. The reduction of

the coiled-coil CRD linker in isoform 1 would alter, in turn, the orientation of the CRDs by likely leading them to rise upward. Thus, while these three residues' insertion, from isoform 1 to 2 of MGL, does not modify the glycan-binding specificities of their CRDs, it might impact drastically the relative geometry of the CRDs in both trimeric isoform and thus their specificity toward different glycan landscape. The conserved disulfide bond positions the CRD domain perpendicular to the coiled-coil axis and has important implications with respect to glycan binding. As we have recently shown on another CLR, thanks to molecular dynamic studies, DC-SIGN can adapt to various distance distribution of glycan epitope presentation, thanks to a rather large flexibility between the neck and the CRD domains (33). Here, a different situation occurs in the case of MGL. The presence of the newly identified C162–C180 bridge strongly constrains the extension capabilities of CRDs from the neck (Fig. 4D). However, the CRD domains show here no extensive contacts with the coiled-coiled neck domain, and subtle variations of the CRD orientation through rotation around the disulfide bond axis might allow plasticity in the presentation of the binding sites. However, the limitation in distance is compensated here, in MGL, by the presence of the additional noncanonical OS-binding site on the opposite side, within the CRD, of the Ca^{2+} -dependent QPD site. This, combined to CRD subtle rotation, might provide a large set of potential adaptation to different surfaces.

Here, the CRD orientation makes both QPD- and OS-binding sites accessible for binding glycans assembled on a surface. The presence of these six binding sites highly increases the multivalency of the interaction and probably explains the very broad pathogen-associated molecular patterns that MGL is capable to recognize, from both gram-positive and gram-negative bacteria,

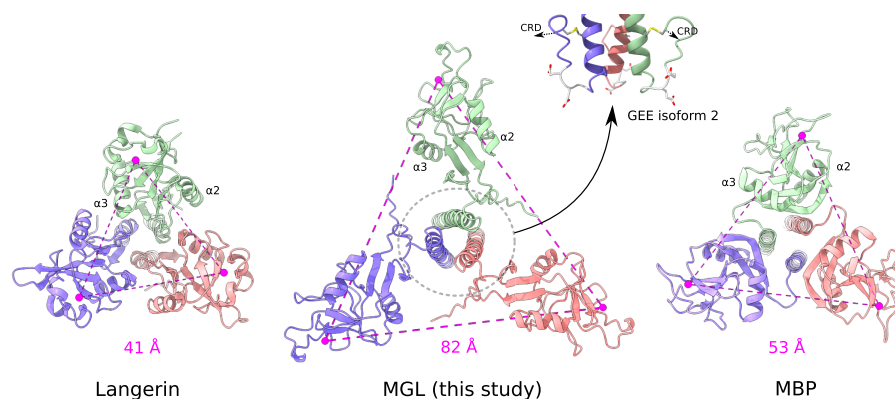


Fig. 6. Comparison of the MGL model with other trimeric C-type lectins. The MGL model, langerin (PDB:3KQG), and mannose-binding protein A (PDB:1KWW) are shown from the C-terminal side of the coiled-coil domain. The calcium ions of the canonical-binding sites are indicated as well as the distances in magenta between adjacent sites. The C terminus of the coiled-coil domain of MGL is surface accessible (gray circle), and the MGL isoform 2 studied here has an additional GEE tripeptide (gray). The C162–C180 disulfide bond is in yellow.

as well as *M. tuberculosis* (12–15). We can hypothesize that the mode of recognition by MGL of teichoic acids, which are polysaccharides assembled similarly as LPS at the surface of *S. aureus* (12), resembles that of LPS.

The 3D arrangement of MGL can explain its recognition of different cell surface glycans. Nevertheless, its binding to R1 and R3 core OS highly differs between isolated ligands and ligands presented at the cell surface. In our experimental conditions, cells are extensively washed with buffer before imaging. Thus, only the very stable interaction with R1 core OS is detected, and we hypothesize that the interaction of MGL with R3-producing cells is more labile. The avidity between MGL and *E. coli* surfaces presenting R1 is certainly key to the interaction. As already shown for MGL, a 150-fold affinity increase is measured between CRD and ECD binding to GalNAc-conjugated bovine serum albumin (BSA) (11). While the difference in avidity toward R1 or R3 might be linked to the protein, it should also be considered that the presentation of the LOS core OS on the surface of cells might differ. Furthermore, it is still unknown how the strength of the interaction of MGL with R1 or R3 presenting bacteria will relate to the function of the immune cell recognition and how it will, in turn, affect the adhesion, signaling, or antigen uptake.

So far, we have examined the binding of MGL to LPS that do not contain O-antigens. Most clinically relevant gram-negative bacteria possess O-antigen of very variable compositions and length (34). This dense and long (~10–40 nm) layer of polysaccharides could be either recognized by MGL, thanks to its ability to bind various glycans, or, on the other hand, it could block access to the core OS and prevent recognition. This should be the focus of future studies on the role of MGL in the recognition of gram-negative bacteria and the subsequent implications in the regulation of the immune response.

Materials and methods

Protein expression and purification

Human MGL isoform 2 ECD (residues Q85–H316 Uniprot Q8IUN9-2) with an N-terminal Strep-tag II and a factor Xa cleavage site (MASWSHPQFEKIEGRGGG) was expressed and purified as already reported (21). Briefly, MGL-ECD was over-expressed in *E. coli* BL21(DE3) cells in inclusion bodies. Inclusion bodies were solubilized in guanidine buffer (25 mM Tris, pH 8, 150 mM NaCl, 6 M guanidine, and 0.01% B-mercaptoethanol). MGL-ECD was subsequently refolded using a drop-by-drop dilution in renaturation buffer

(100 mM Tris, pH 8, 150 mM NaCl, and 25 mM CaCl₂) and was subjected to two purification steps: a GalNAc–agarose affinity column (Sigma), eluted with EDTA buffer (150 mM NaCl, 25 mM Tris, pH 8, and 10 mM EDTA) followed by a Toyopearl HW-50S gel filtration column (Tosoh Bioscience). MGL-ECD was also produced as perdeuterated ²H, ¹⁵N-labeled form ([U-²H, ¹⁵N] MGL-ECD) in 95% D₂O with D-glucose-d7 as glucose source as described (35). MGL-ECD^{D269H} mutant was over-expressed in *E. coli* BL21(DE3) cells in LB medium as inclusion bodies, which were subjected to the same solubilization, and renaturation steps described above. MGL-ECD^{D269H} was purified using an AktaXpress with a Strep-tag affinity column eluted with 2.5 mM desthiobiotin followed by a Toyopearl HW-50S gel filtration column (Tosoh Bioscience).

MGL-CRD and MGL-CRD^{D269H} (C181–H316 Uniprot Q8IUN9-2) with N-terminal His-tag and TEV cleavage site (HHHHHH IEGRRGGGG) were expressed and purified as described (11) in M9 minimal medium as ¹³C, ¹⁵N-labeled proteins. An MGL-CRD^{D269H} binding assay was performed on the GalNAc–agarose affinity column used for ECD purification to assess its affinity for GalNAc, which revealed it did not bind to GalNAc affinity column.

Fluorescence microscopy and flow cytometry

MGL-ECD and MGL-ECD^{D269H} were labeled with Alexa Fluor 647-NHS (Invitrogen). Briefly, MGL at 5 mg/mL in PBS buffer was incubated in 200 mM sodium bicarbonate and 0.4 mg/mL AF647-NHS for 1 h. Excess dye was removed with G25-PD10 desalting column (GE Healthcare), and MGL fractions dialyzed further against PBS buffer and concentrated. *Escherichia coli* R1 bacteria carrying R1 core OS (F470, derivative from *E. coli* O8:K27) and R3 (F653, derivative from *E. coli* O14:K7) (22) were grown in LB at 37°C under agitation up to 0.9 OD_{600 nm}. Cells were collected by centrifugation, washed in cold PBS, and incubated with 670 nM MGL-AF647 in PBS and 2 mM CaCl₂ buffer for 15 min. Cells were washed five times with cold PBS and imaged. For each sample, 2 μL of cells in suspension was mounted between a glass slide and a 1.5H 170 μm thick glass coverslip and observed using an inverted IX83 microscope, with the UPLFLN 100x oil immersion objective from Olympus (numerical aperture 1.49), using a fibered Xcite Metal-Halide excitation lamp in conjunction with the appropriate excitation filters, dichroic mirrors, and emission filters specific for AF647 (4X4MB set, Semrock). Acquisitions were performed with Velocity software (Quorum Technologies) using a sCMOS 2,048 × 2,048 camera (Hamamatsu ORCA Flash 4, 16 bits/pixel) achieving a final magnification of 64 nm per pixel.

Flow cytometry was performed on a VYB device (Miltenyi biotech) and analyzed with Macsquant software. Cells (50 μ L) grown in LB at $DO_{600\text{ nm}} = 1$ were resuspended in presence of 670 nM MGL-AF647 (wt or D269H variant) in PBS, 2 mM $CaCl_2$ with/without 10 mM GalNAc for 15 min, centrifuged twice to remove excess protein, resuspended in 150 μ L, and injected for FACS analysis until 200,000 events were recorded. MGL-ECD binding to cells was expressed as % population \times mean fluorescence (cy5 channel) and normalized to 100% for MGL-ECD wt binding.

LOS and OS preparation

F470 and F653 cells were grown in LB. LOS were extracted following the phenol–chloroform–petroleum ether (PCP) method and de-N- and O-acylated as already described (21, 36). LOS R1 (0.84 mM) and LOS R3 (0.6 mM) were solubilized in DDM micelles by addition of 150 mM of DDM in HBS-N and 2 mM $CaCl_2$ for 15 min. Insoluble material was discarded by ultracentrifugation at 100,000 *g* for 30 min, and sample homogeneity was checked by dynamic light scattering.

SPR experiments

SPR interaction was performed using oriented surfaces of ECD-MGL, specifically N-terminally biotinylated, thanks to a sort-tagging procedure (37). Streptavidin at 100 μ g/mL in 10 mM NaOAc, pH 4, was immobilized on sensor chip S Serie CM3 (Cytiva). Biot-ECD was diluted at 0.5 μ g/mL in running buffer (HBS-N [cytiva], 2 mM $CaCl_2$, and 300 μ M DDM) and injected at 5 μ L/min until 125 RU capture. For interaction measurements, LOS R1 or LOS R3 solubilized in DDM was injected at increasing concentrations in running buffer at 20 μ L/min. Streptavidin flow cell surface was used as reference for correction of the binding response. Regeneration of the surfaces was achieved by 50 mM EDTA, pH 8. Binding curves were analyzed using Biacore T200 Evaluation Software 3.2.1 (GE Healthcare), and data were fit using steady-state affinity model.

NMR titrations

Human ^{15}N -labeled MGL-CRD^{wt} or MGL-CRD^{D269H} at 50 μ M in 25 mM Tris, pH 8, 150 mM NaCl, and 4 mM $CaCl_2$ was titrated with increasing concentrations of GalNAc, R1, or R3 OSs up to 20 molar equivalents of glycan:CRD. 1H - ^{15}N -BEST-TROSY correlation experiments were recorded at 30°C on an 850, 700, or 600 MHz Bruker NMR spectrometer equipped with a cryoprobe at each OS addition. NMR titration experiments with MGL-ECD were performed at a concentration of 600 μ M of the 2H , ^{15}N MGL-ECD with 1 and 2 molecular equivalents of either OS R1 or R3 ligands. 1H - ^{15}N -BEST-TROSY correlation spectra were collected at 35°C on Bruker Avance spectrometer at 850 MHz. All spectra were processed using TopSpin 3.5 software and analyzed using CcpNmr analysis 3.0 software. CSPs, corresponding to the chemical shift change in the 1H - ^{15}N BTROSY spectra upon addition of ligands, were calculated as $CSP = ((\Delta\delta^1H)^2 + ([\Delta\delta^{15}N/10])^2)^{1/2}$, where $\Delta\delta^1H$ and $\Delta\delta^{15}N$ are chemical shift changes in amide proton and amide nitrogen, respectively. CSPs higher than twice the standard deviation of all chemical shifts were considered significant.

SAXS

SAXS data have been recorded on MGL-ECD domain at 1 mg/mL in 25 mM Tris, pH 8, 150 mM NaCl, and 4 mM $CaCl_2$ buffer at 25°C at the European Synchrotron Radiation Facility (ESRF) BM29 Biosaxs beamline (Grenoble). Automatic frame selection and buffer subtraction were performed by ISPyB (38). SAXS data were analyzed

with Atsas 3.1.3 (39) and BioXTAS RAW (40). *P*(*r*) distribution function was used as input for DAMMIF online, doing five runs including P3 symmetry and prolate anisotropy. The five solutions were sorted by DAMAVER as two clusters, and the most representative envelope of the best cluster is presented. The AlphaFold multimer program was run with the entire sequence of the MGL-ECD construct expressed, and as a trimeric protein as input. Twenty-four models have been generated, and the 10 best ranked models according to their DockQ score were retained for further analysis (41).

Acknowledgments

We would like to thank A. Imberty and N. Thielens for stimulating discussions. We thank Rose-Laure Revel-Goyet, Françoise Lacroix, Oleksandr Glushonkov, and Jean-Philippe Kleman (IBS, Grenoble) for the support and access to the Cell Imaging platform. This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-02).

Supplementary material

Supplementary material is available at PNAS Nexus online.

Funding

We thank the Agence Nationale de la Recherche (ANR) PIA for Glyco@Alps (ANR-15-IDEX-02) and their support of M.M. and C.L. and the Mizutani Foundation for Glycosciences 30th Research grant for funding C.L. M.A. received funding from GRAL, the Grenoble Alliance for Integrated Structural and Cell Biology, a program of the Chemistry Biology Health Graduate School of Université Grenoble Alpes (ANR-17-EURE-0003). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement No 851356 to R.M. This work was granted access to the CCRT High-Performance Computing (HPC) facility under the Grant CCRT2022-lagurie awarded by the Fundamental Research Division (DRF) of CEA.

Author contributions

C.L., J.P.-S., F.F., R.M., and A.M. designed the research; M.A., C.L., J.P.-S., F.F., R.M., and A.M. wrote the paper; M.A., M.M., C.L., M.T., I.A., F.N.-F., and J.-P.K. performed the research; M.A., M.M., M.T., C.L., J.P.-S., and J.P.-K. analyzed the data.

Preprint

This manuscript was posted on a preprint: <https://doi.org/10.1101/2023.03.02.530591>.

Data availability

All data required for main findings of this manuscript are included in the article and Supplementary material.

References

- Di Lorenzo F, et al. 2022. A journey from structure to function of bacterial lipopolysaccharides. *Chem Rev.* 122:15767–15821. <https://doi.org/10.1021/acs.chemrev.0c01321>

- 2 Ryu JK, et al. 2017. Reconstruction of LPS transfer cascade reveals structural determinants within LBP, CD14, and TLR4-MD2 for efficient LPS recognition and transfer. *Immunity* 46:38–50. <https://doi.org/10.1016/j.immuni.2016.11.007>
- 3 Yi YS. 2017. Caspase-11 non-canonical inflammasome: a critical sensor of intracellular lipopolysaccharide in macrophage-mediated inflammatory responses. *Immunology* 152:207–217. <https://doi.org/10.1111/IMM.12787>
- 4 Reinhardt A, et al. 2015. Antigenic potential of a highly conserved *Neisseria meningitidis* lipopolysaccharide inner core structure defined by chemical synthesis. *Chem Biol.* 22:38–49. <https://doi.org/10.1016/j.chembiol.2014.11.016>
- 5 Rollenske T, et al. 2018. Cross-specificity of protective human antibodies against *Klebsiella pneumoniae* LPS O-antigen. *Nat Immunol.* 19:617–624. <https://doi.org/10.1038/S41590-018-0106-2>
- 6 Geissner A, et al. 2019. Microbe-focused glycan array screening platform. *Proc Natl Acad Sci USA.* 116:1958–1967. <https://doi.org/10.1073/pnas.1800853116>
- 7 Hanske J, et al. 2017. Bacterial polysaccharide specificity of the pattern recognition receptor langerin is highly species-dependent. *J Biol Chem.* 292:862–871. <https://doi.org/10.1074/jbc.M116.751750>
- 8 Jégouzo SAF, et al. 2020. Mammalian lectin arrays for screening host-microbe interactions. *J Biol Chem.* 295:4541–4555. <https://doi.org/10.1074/jbc.RA120.012783>
- 9 Mayer S, Raulf M-K, Lepenies B. 2017. C-type lectins: their network and roles in pathogen recognition and immunity. *Histochem Cell Biol.* 147:223–237. <https://doi.org/10.1007/s00418-016-1523-7>
- 10 Mnich ME, van Dalen R, van Sorge NM. 2020. C-type lectin receptors in host defense against bacterial pathogens. *Front Cell Infect Microbiol.* 10:1–15. <https://doi.org/10.3389/fcimb.2020.00309>
- 11 Bulteau F, et al. 2022. Targeting tn-antigen-positive human tumors with a recombinant human macrophage galactose C-type lectin. *Mol Pharm.* 19:235–245. <https://doi.org/10.1021/acs.molpharmaceut.1c00744>
- 12 Mnich ME, et al. 2019. The C-type lectin receptor MGL senses N-acetylgalactosamine on the unique *Staphylococcus aureus* ST395 wall teichoic acid. *Cell Microbiol.* 21:13072. <https://doi.org/10.1111/cmi.13072>
- 13 Naqvi KF, et al. 2021. Novel role for macrophage galactose-type lectin-1 to regulate innate immunity against *Mycobacterium tuberculosis*. *J Immunol.* 207:221–233. <https://doi.org/10.4049/jimmunol.2001276>
- 14 van Sorge NM, et al. 2009. N-Glycosylated proteins and distinct lipooligosaccharide glycoforms of *Campylobacter jejuni* target the human C-type lectin receptor MGL. *Cell Microbiol.* 11:1768–1781. <https://doi.org/10.1111/j.1462-5822.2009.01370.x>
- 15 van Vliet SJ, et al. 2009. Variation of *Neisseria gonorrhoeae* lipooligosaccharide directs dendritic cell-induced T helper responses. *PLoS Pathog.* 5:e1000625. <https://doi.org/10.1371/JOURNAL.PPAT.1000625>
- 16 Zelensky AN, Gready JE. 2005. The C-type lectin-like domain superfamily. *FEBS J.* 272:6179–6217. <https://doi.org/10.1111/J.1742-4658.2005.05031.X>
- 17 Valverde P, Martínez JD, Cañada FJ, Ardá A, Jiménez-Barbero J. 2020. Molecular recognition in C-type lectins: the cases of DC-SIGN, langerin, MGL, and L-sectin. *ChemBioChem.* 21:2999–3025. <https://doi.org/10.1002/cbic.202000238>
- 18 Gabba A, et al. 2021. Crystal structure of the carbohydrate recognition domain of the human macrophage galactose C-type lectin bound to GalNAc and the tumor-associated Tn antigen. *Biochemistry* 60:1327–1336. <https://doi.org/10.1021/acs.biochem.1c00009>
- 19 Marcelo F, et al. 2019. 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.* 294:1300–1311. <https://doi.org/10.1074/jbc.RA118.004957>
- 20 Diniz A, et al. 2019. The plasticity of the carbohydrate recognition domain dictates the exquisite mechanism of binding of human macrophage galactose-type lectin. *Chemistry* 25:13945–13955. <https://doi.org/10.1002/chem.201902780>
- 21 Maalej M, et al. 2019. Human macrophage galactose-type lectin (MGL) recognizes the outer core of *Escherichia coli* lipooligosaccharide. *ChemBioChem* 20:1778–1782. <https://doi.org/10.1002/cbic.201900087>
- 22 Amor K, et al. 2000. Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun.* 68:1116–1124. <https://doi.org/10.1128/IAI.68.3.1116-1124.2000>
- 23 Smilgies DM, Folta-Stogniew E. 2015. Molecular weight–gyration radius relation of globular proteins: a comparison of light scattering, small-angle X-ray scattering and structure-based data. *J Appl Crystallogr.* 48:1604–1606. <https://doi.org/10.1107/S1600576715015551>
- 24 Feinberg H, Powlesland AS, Taylor ME, Weis WI. 2010. Trimeric structure of langerin. *J Biol Chem.* 285:13285–13293. <https://doi.org/10.1074/jbc.M109.086058>
- 25 Ng KK-S, et al. 2002. Orientation of bound ligands in mannose-binding proteins. *J Biol Chem.* 277:16088–16095. <https://doi.org/10.1074/jbc.M200493200>
- 26 Jumper J, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- 27 Tunyasuvunakool K, et al. 2021. Highly accurate protein structure prediction for the human proteome. *Nature* 596:590–596. [doi:10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1)
- 28 Ruiz NI, Drickamer K. 1996. Differential ligand binding by two subunits of the rat liver asialoglycoprotein receptor. *Glycobiology* 6:551–559.
- 29 Furukawa A, et al. 2013. Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc Natl Acad Sci USA.* 110:17438–17443. <https://doi.org/10.1073/pnas.1312649110>
- 30 Chabrol E, et al. 2012. Glycosaminoglycans are interactants of langerin: comparison with gp120 highlights an unexpected calcium-independent binding mode. *PLoS One* 7:e50722. <https://doi.org/10.1371/JOURNAL.PONE.0050722>
- 31 Medve L, et al. 2019. Enhancing potency and selectivity of a DC-SIGN glycomimetic ligand by fragment-based design: structural basis. *Chemistry* 25:14659–14668. <https://doi.org/10.1002/CHEM.201903391>
- 32 Wawrzinek R, et al. 2021. A remote secondary binding pocket promotes heteromultivalent targeting of DC-SIGN. *J Am Chem Soc.* 143:18977–18988. <https://doi.org/10.1021/jacs.1c07235>
- 33 Porkolab V, et al. 2023. Powerful avidity with a limited valency for virus-attachment blockers on DC-SIGN: combining chelation and statistical rebinding with structural plasticity of the receptor. *ACS Cent Sci.* 9:709–718. <https://doi.org/10.1021/ACSCENTSCI.2C01136>
- 34 Whitfield C, Williams DM, Kelly SD. 2020. Lipopolysaccharide O-antigens—bacterial glycans made to measure. *J Biol Chem.* 295:10593. <https://doi.org/10.1074/JBC.REV120.009402>
- 35 Jean NL, et al. 2014. Elongated structure of the outer-membrane activator of peptidoglycan synthesis LpoA: implications for PBP1A stimulation. *Structure* 22:1047–1054. <https://doi.org/10.1016/j.str.2014.04.017>

- 36 de Castro C, Parrilli M, Holst O, Molinaro A. 2010. Microbe-associated molecular patterns in innate immunity. Extraction and chemical analysis of gram-negative bacterial lipopolysaccharides. *Methods Enzymol.* 480:89–115. [https://doi.org/10.1016/S0076-6879\(10\)80005-9](https://doi.org/10.1016/S0076-6879(10)80005-9)
- 37 Achilli S, et al. 2020. Tetralec, artificial tetrameric lectins: a tool to screen ligand and pathogen interactions. *Int J Mol Sci.* 21:1–20. <https://doi.org/10.3390/ijms21155290>
- 38 de Maria Antolinos A, et al. 2015. ISPyb for BioSAXS, the gateway to user autonomy in solution scattering experiments. *Acta Crystallogr D Biol Crystallogr.* 71:76–85. <https://doi.org/10.1107/S1399004714019609>
- 39 Manalastas-Cantos K, et al. 2021. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Crystallogr.* 54:343–355. <https://doi.org/10.1107/S1600576720013412>
- 40 Hopkins JB, Gillilan RE, Skou S. 2017. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr.* 50:1545–1553. <https://doi.org/10.1107/S1600576717011438>
- 41 Basu S, Wallner B. 2016. Dockq: a quality measure for protein-protein docking models. *PLoS One* 11:e0161879. <https://doi.org/10.1371/journal.pone.0161879>

Supplementary information for Abbas, Maalej et al.

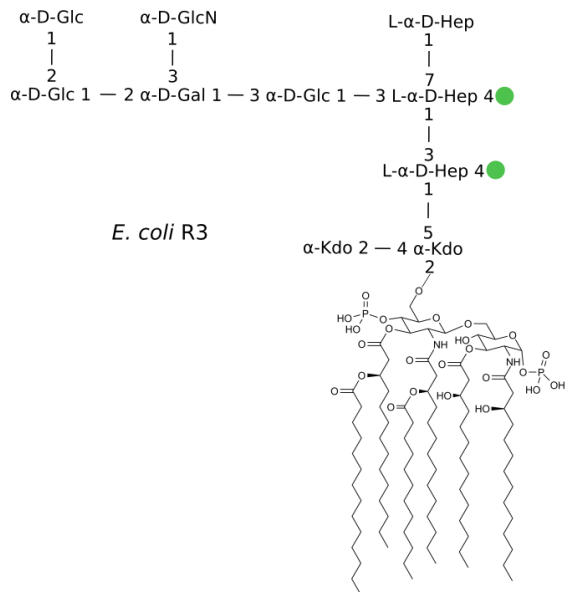


Figure S1: Structure of the main form of *E. coli* R3 type LipoOligoSaccharide

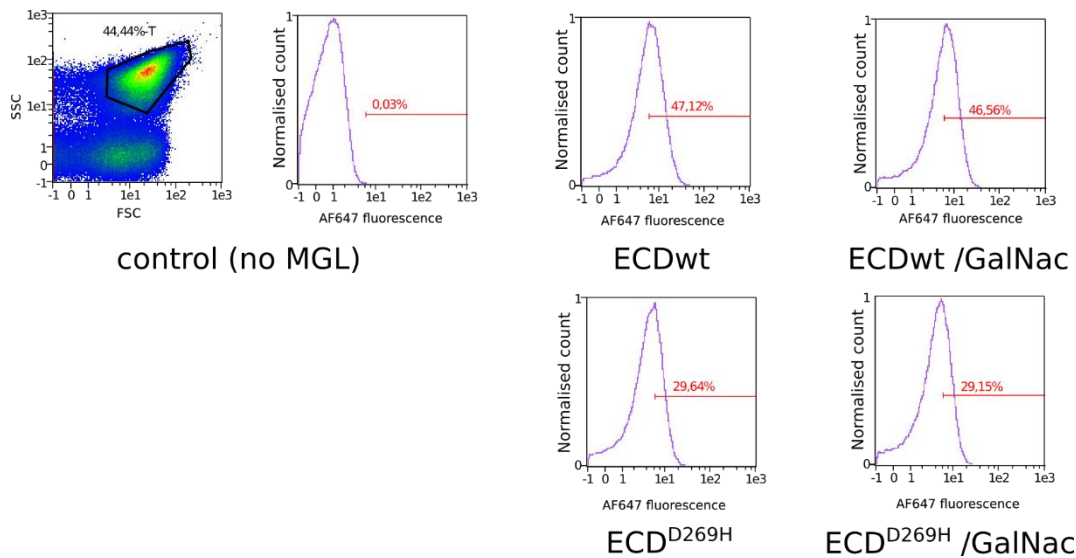


Figure S2 : Flow Cytometry of F470 cells with/without ECDwt and ECD^{D269H} labelled with Alexafluor647. Left selection of bacterial population on 2D representation of Forward (FSC) and side (SSC) scattering and on the right count vs fluorescence of the selected bacterial population. Only the population within the red range was considered to have significant fluorescence. Associated % of population is indicated (before normalization to 100% for ECDwt shown in Fig 2).

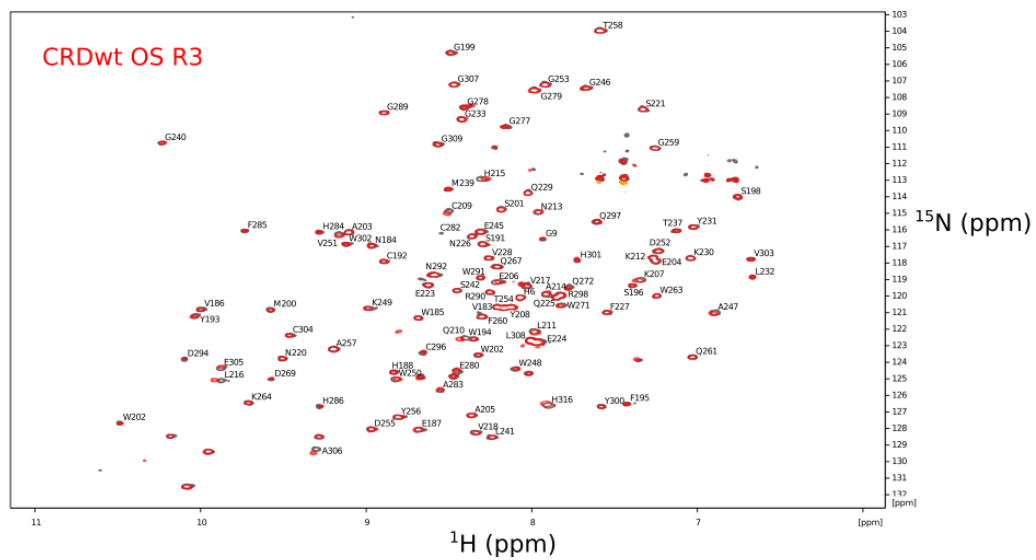
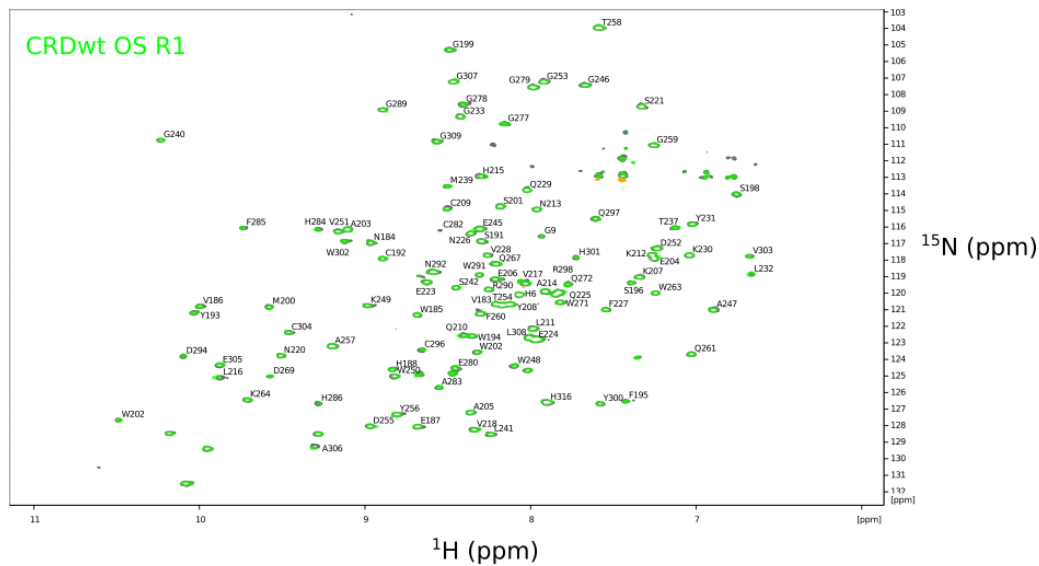
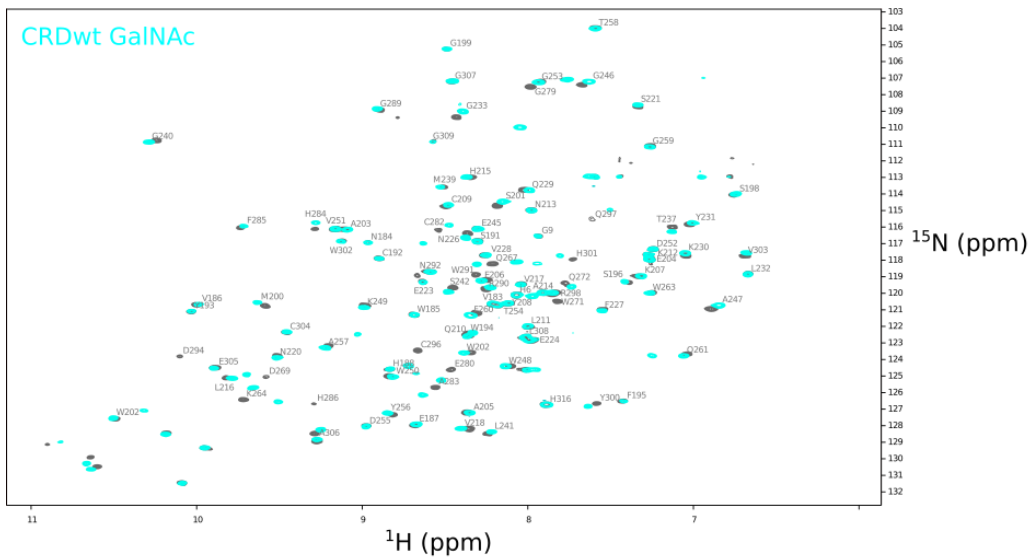


Figure S4. ^{15}N - ^1H correlation spectra of MGL-CRD wt before(black) and after addition of 20:1 Glycan/protein ratio.

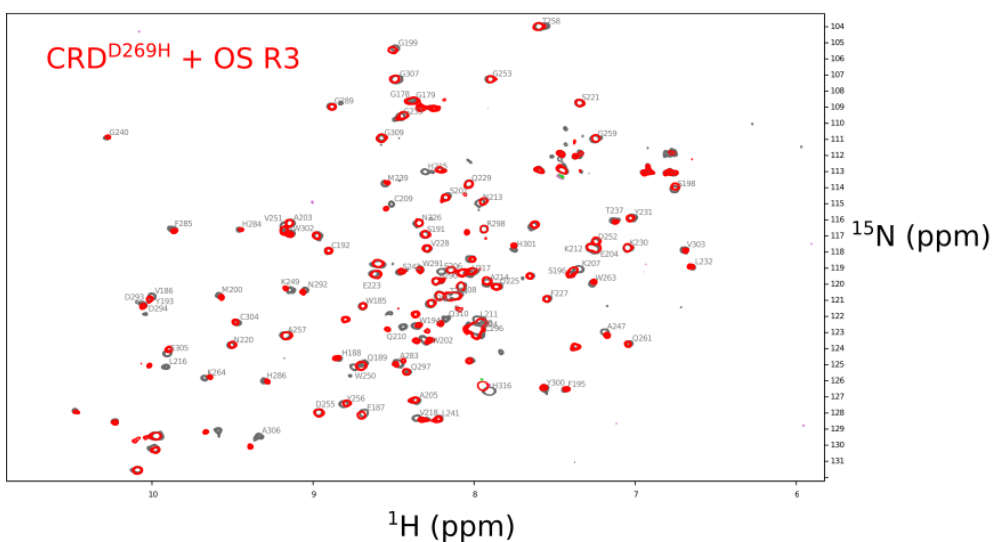
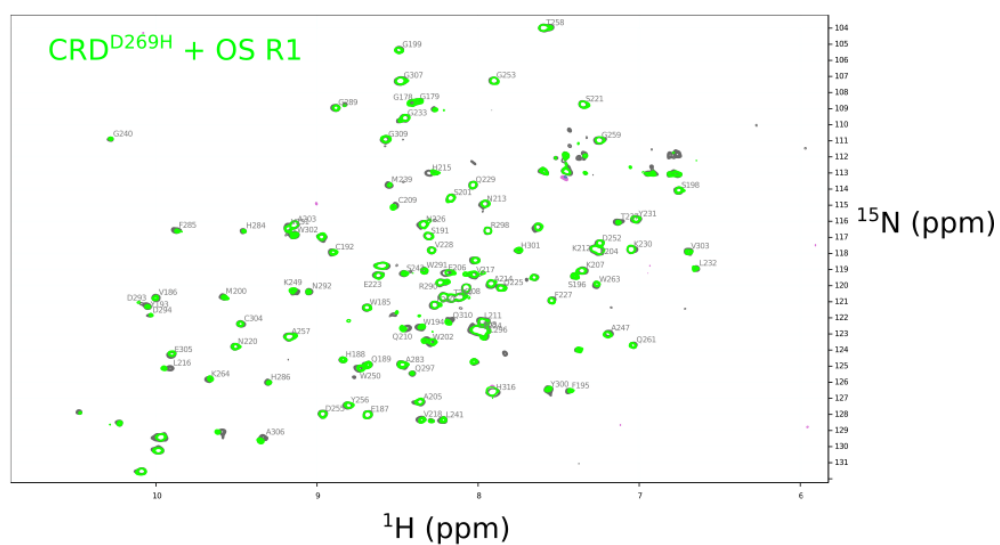
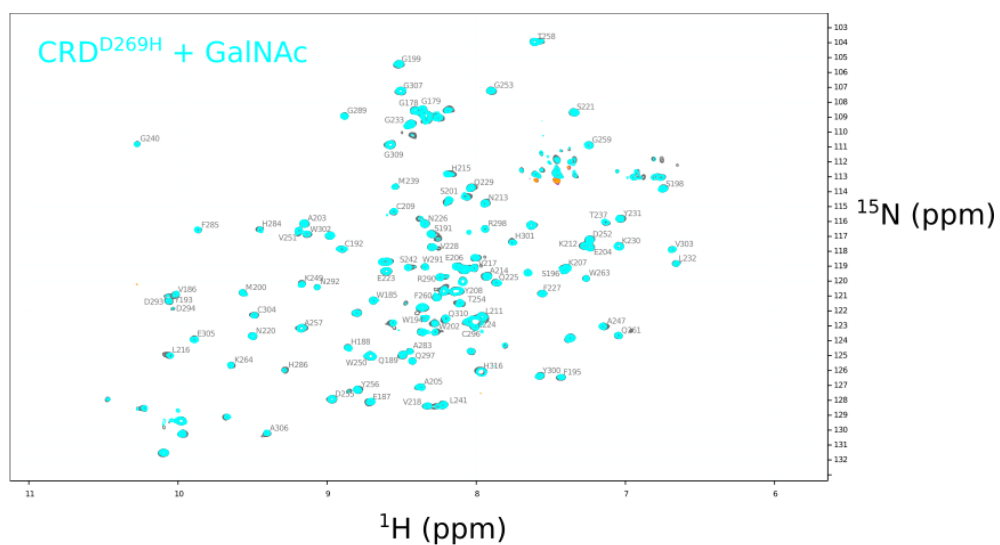


Figure S5. ^{15}N - ^1H correlation spectra of MGL-CRD^{D269H} before (black) and after addition of 20:1 Glycan/protein ratio.

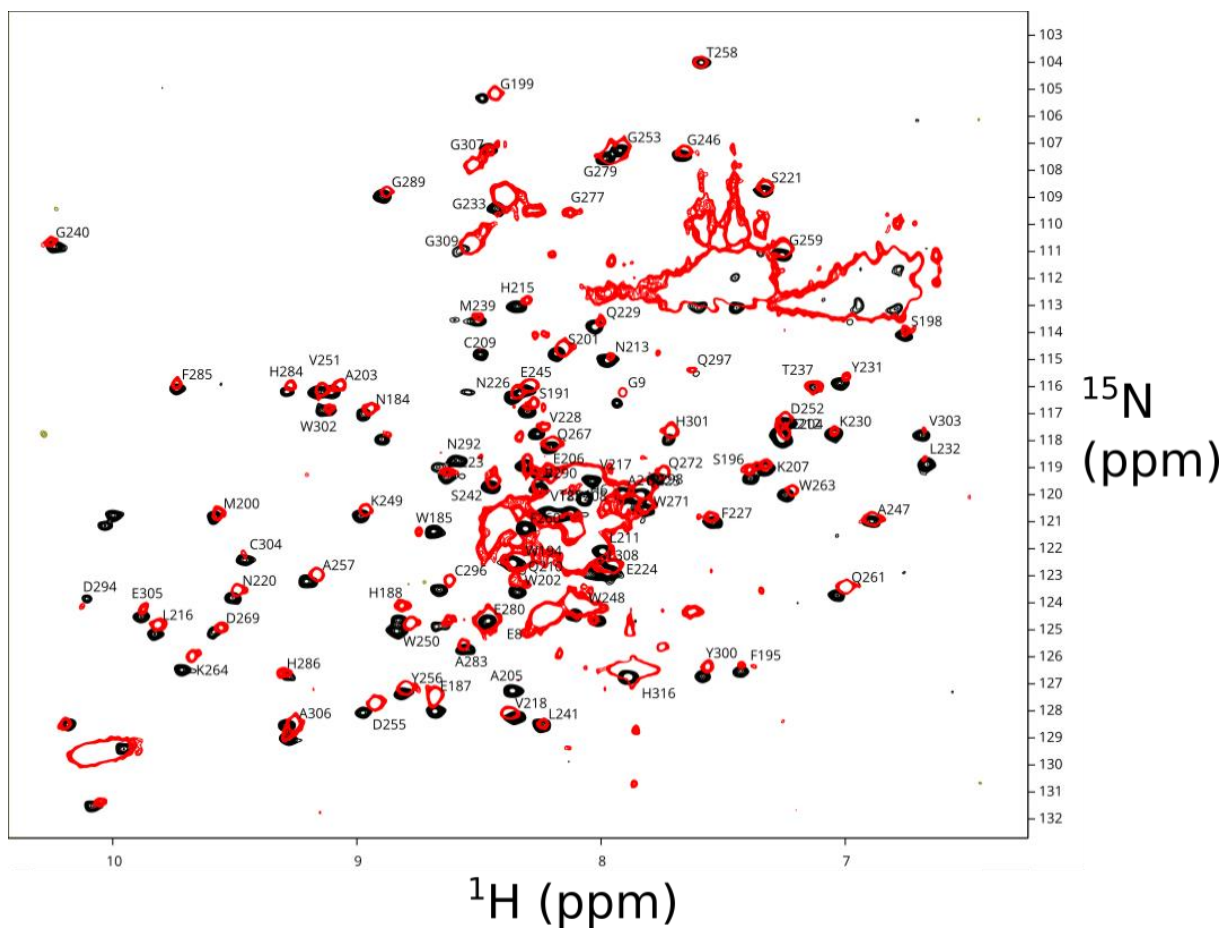


Figure S6 Overlay of ^{15}N - ^1H correlation spectra of MGL-ECD (red) at 35°C and MGL-CRD (black) at 30°C. The assignments of the CRD that could be transferred by superimposition are indicated.

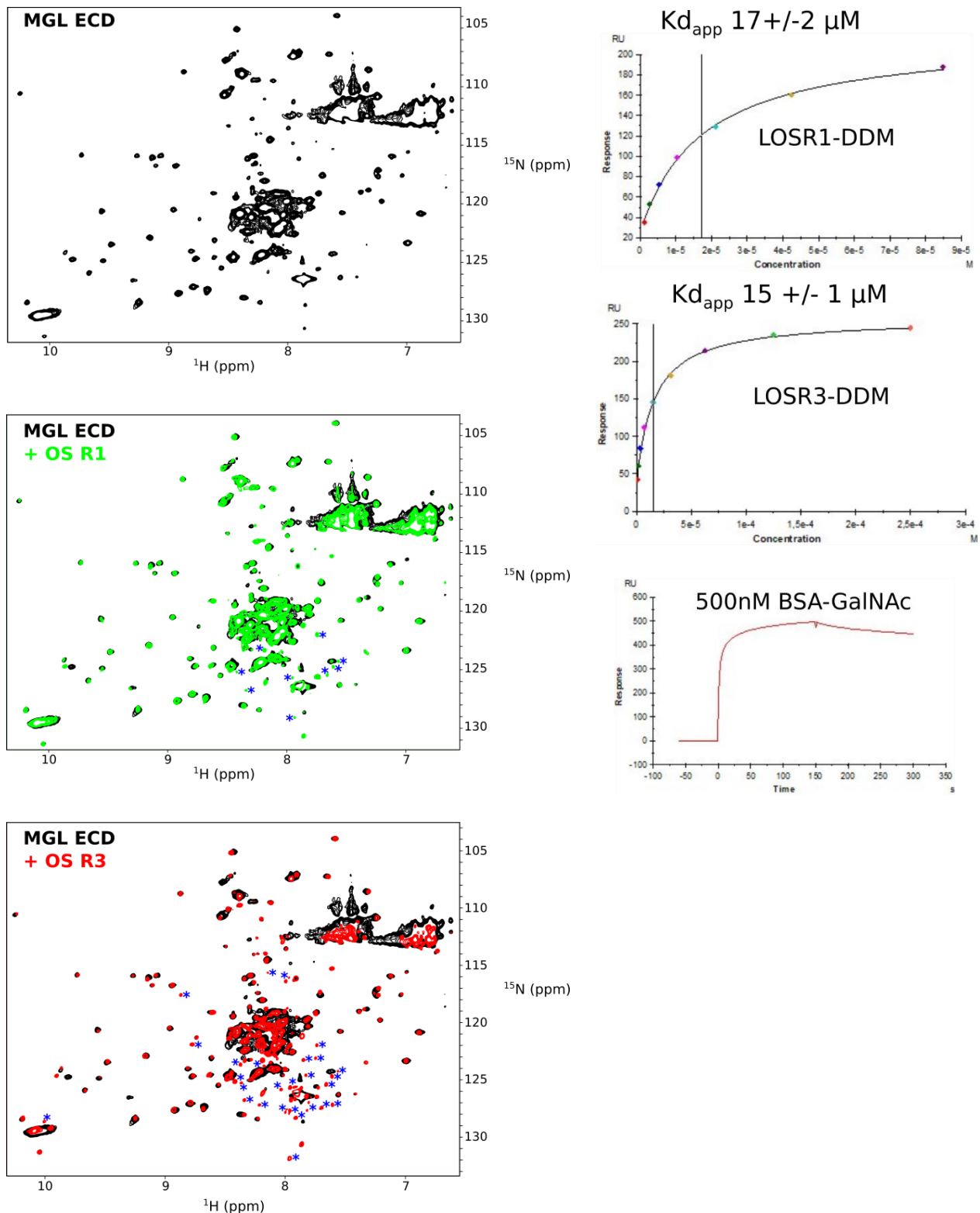


Figure S7. (left) ^{15}N - ^1H correlation spectra of MGL-ECD wt before (black) and after addition of 2:1 Glycan/protein ratio. Blue asterisks correspond to new peaks appearing from minor protein degradation over time. (right) Steady-state K_d determination of LOS R1 and LOS R3 in DDM micelles in interaction with immobilized MGL-ECD by SPR. The control with BSA-GalNAc is included at the bottom.

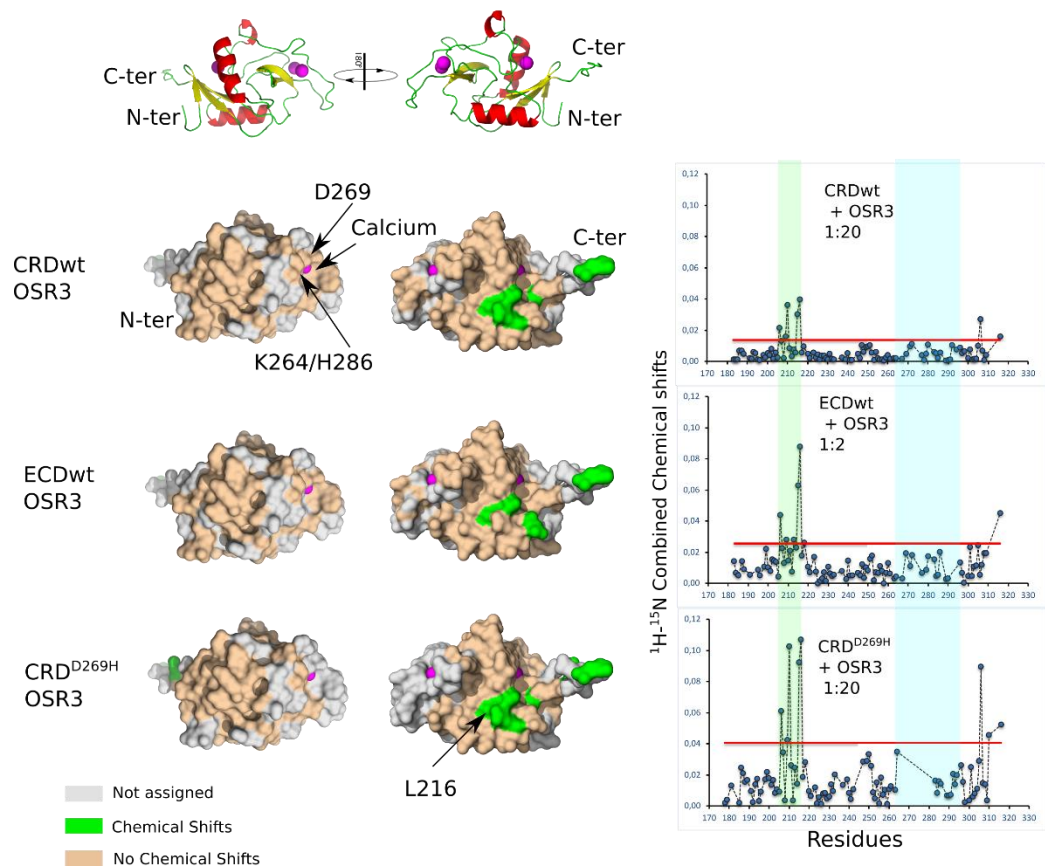


Fig S8. Chemical shifts Perturbations represented on MGL-CRD surface of MGL-CRD wt, D269H and ECD upon OS R3 interaction with the same color codes and orientations as in Fig 3. Histogram of CSP depending on residue number is shown on the right.

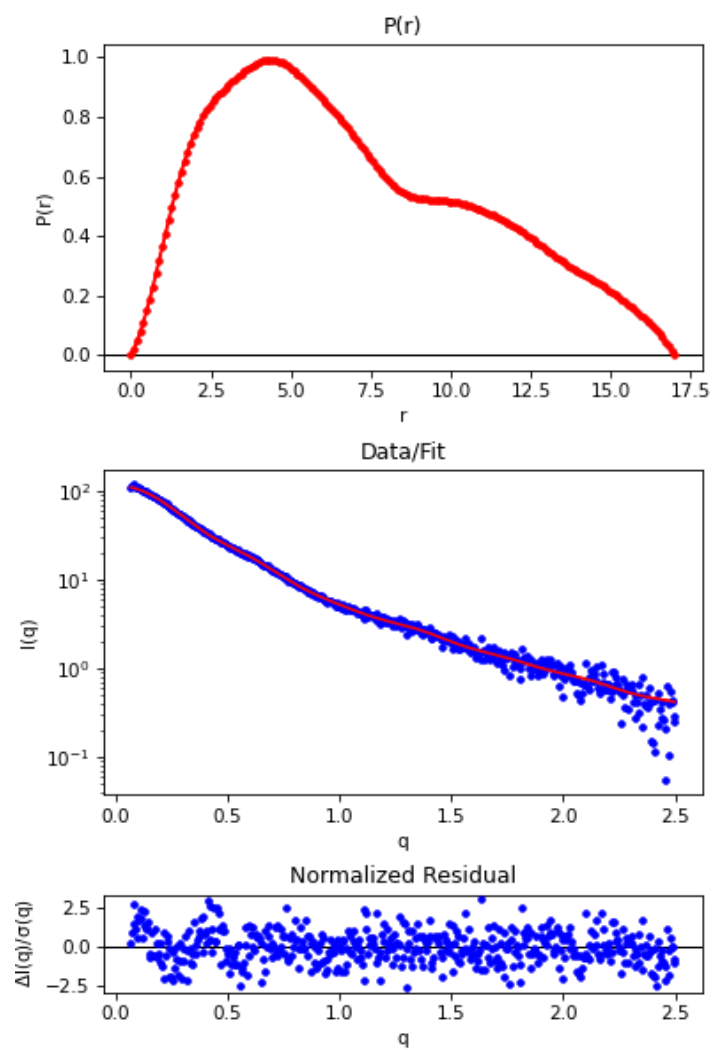


Fig S9: Pairwise distribution function used to calculate the SAXS envelope, with comparison (middle) with the experimental curve and associated residuals.

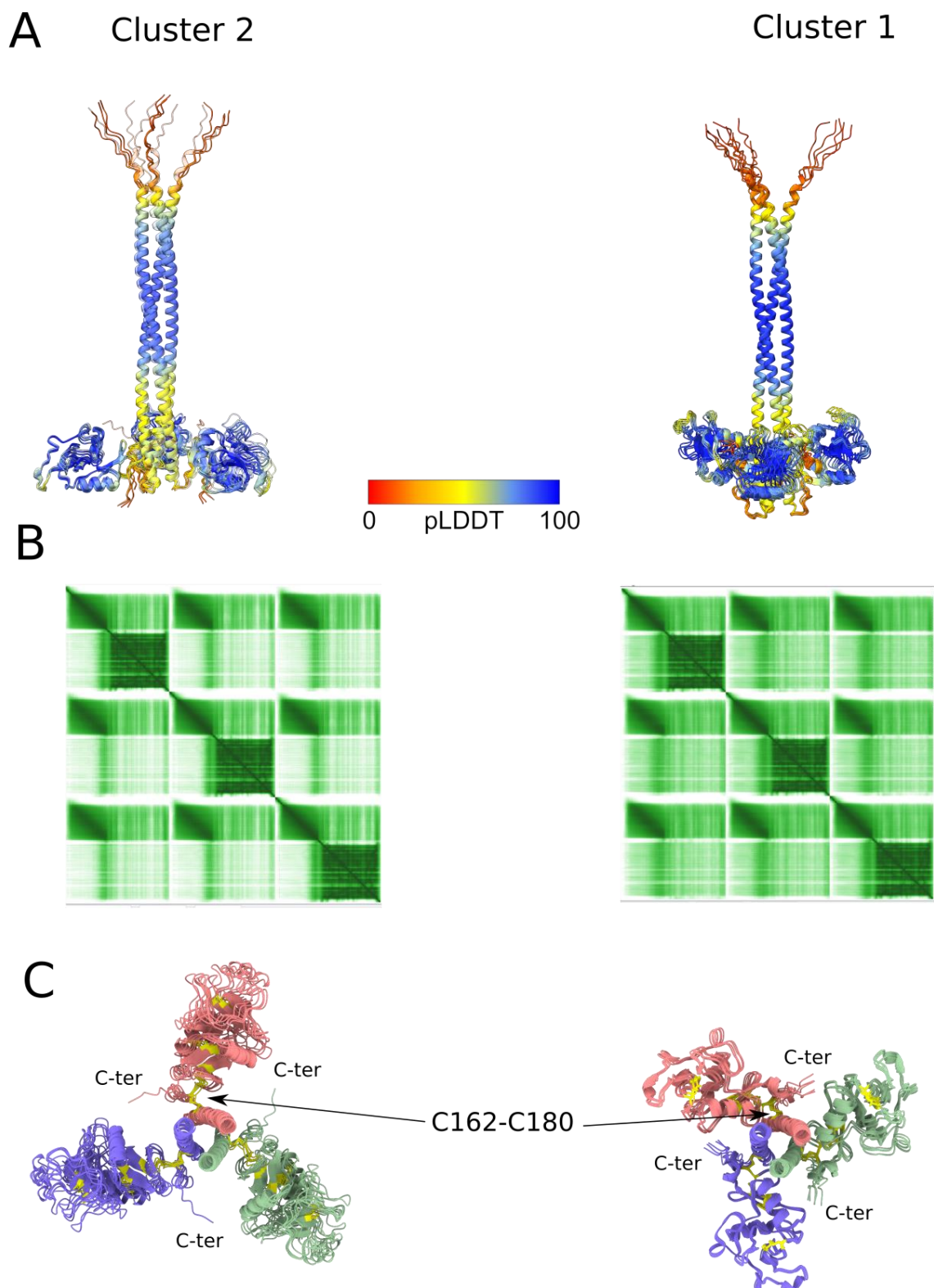


Fig S10 Alphafold models can be subdivided into two clusters. A) superimposed structures of both clusters with their PLDDT scores. The CRD structure is predicted with a 0.5 Å rmsd (across 128 residues 0.46 Å rmsd for cluster 1 and 0.52 Å for cluster 2) with respect to X-ray structure of MGL CRD (PDB:6PY1) B) The Prediction Alignment Error for representative structures of both clusters are showed and show little prediction of interactions between coiled-coil and CRD domains. C) structures of both clusters Superimposed on the coiled-coil motif. Disulfide bonds are shown in yellow.

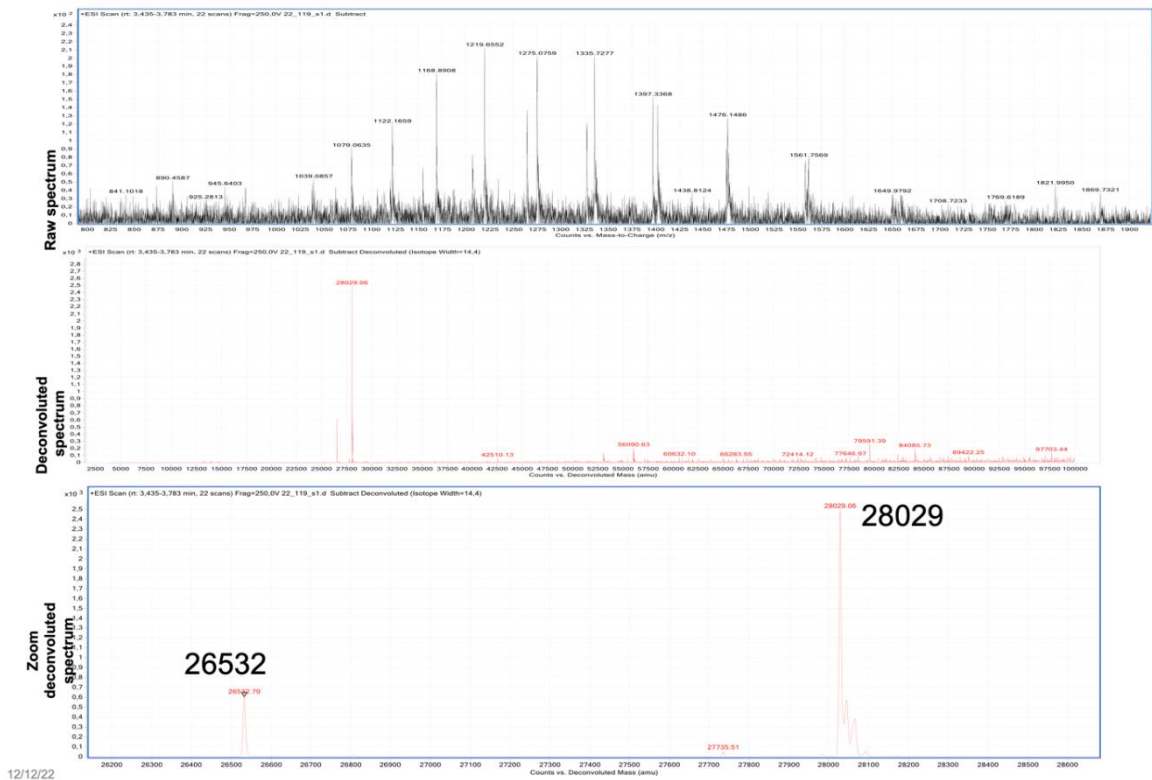


Fig S11. Electrospray Mass spectrometry of MGL-ECD. The expected mass of MGL-ECD is 28036.7 Da with loss of the first methionine. The main mass observed is 28029 Da corresponding to four oxidized cysteines. A minor peak at 26532 Da is probably a truncated form of MGL-ECD.

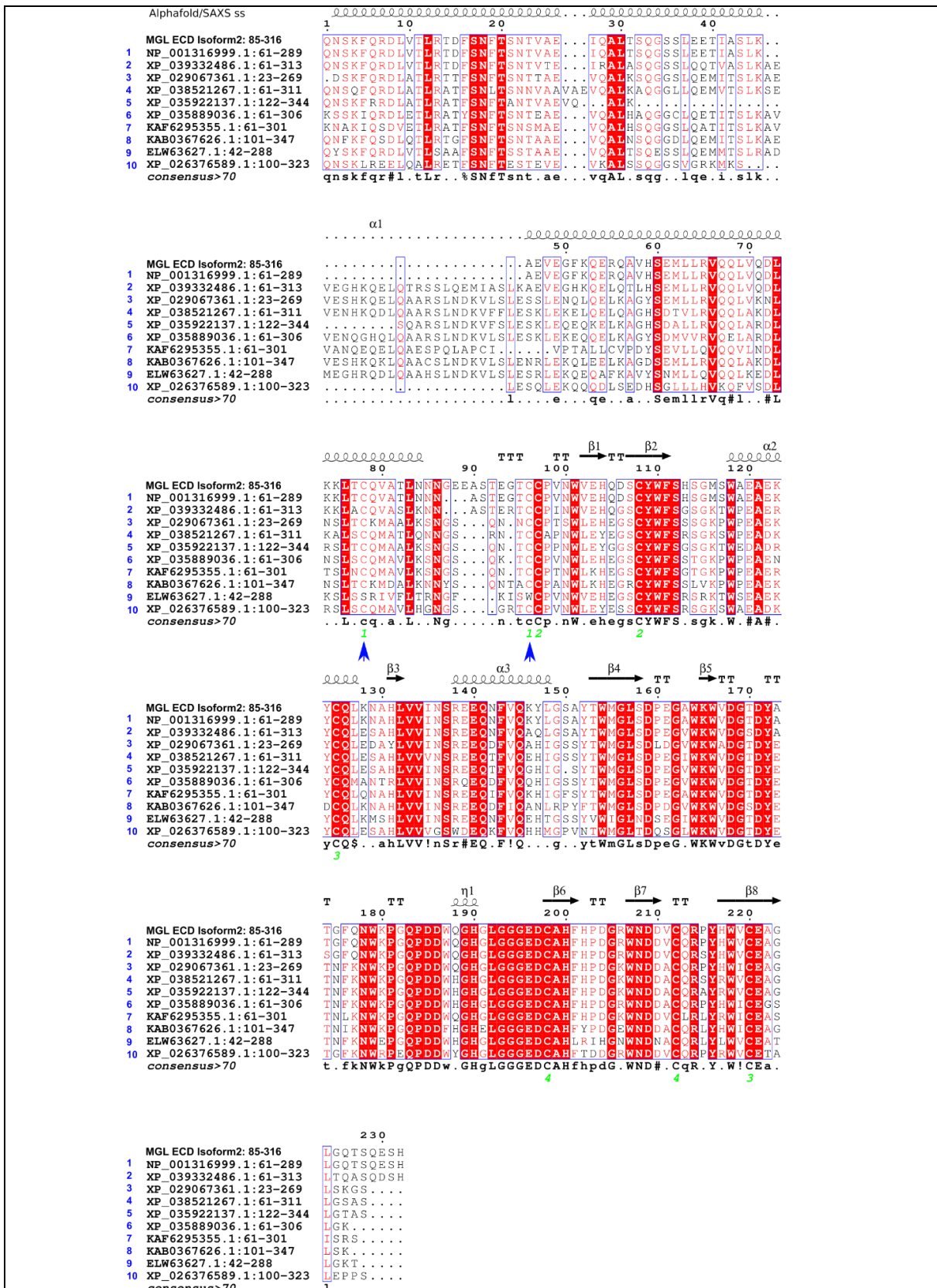


Figure S12: Sequence alignment of the MGL ECD sequence used in this study after a BLAST against clustered non-redundant database. Alignment was performed with [clustalO](#) and represented using Esript3 (Robert and Gouet, 2014). The secondary structure was extracted from the best alphafold/SAXS structure (ranked 5). Cysteines involved in disulfide bonds are shown with a number in green. Cys 162 and 180 are highlighted by a blue arrow.

	Cluster Composition	Cluster Ancestor	Representative sequence	Query Cover	E value	% ident	Acc. Len
1	14 member(s) 7 organism(s)	apes	CLEC 10 member A isoform 3 [Homo sapiens]	100%	1E-168	98.71	289
2	2 member(s) 2 organism(s)	primates	CLEC 10 member A isoform X1 [Saimiri boliviensis boliviensis]	100%	1E-140	77.34	313
3	2 member(s) 1 organism(s)	narwhal	CLEC 10 member A-like isoform X4 [Monodon monoceros]	97%	6E-119	65.74	269
4	6 member(s) 4 organism(s)	dog coyote wolf fox	CLEC 10 member A isoform X2 [Canis lupus familiaris]	98%	4e-113	63.92	311
5	2 member(s) 1 organism(s)	gray seal	ASGPR 1-like isoform X15 [Halichoerus grypus]	98%	4E-113	68.42	344
6	1 member(s) 1 organism(s)	pale spear-nosed bat	CLEC 10 member A-like isoform X1 [Phyllostomus discolor]	97%	9E-111	61.60	308
7	1 member(s) 1 organism(s)	bats	C-type lectin domain containing 10A [Myotis myotis]	97%	4E-108	62.45	302
8	1 member(s) 1 organism(s)	Reeves' muntjac	hypothetical protein FD755_020950 [Muntiacus reevesi]	97%	4E-103	61.20	349
9	1 member(s) 1 organism(s)	Chinese tree shrew	CLEC 10 member A [Tupaia chinensis]	97%	4E-99	60.56	509
10	1 member(s) 1 organism(s)	brown bear	ASGPR 1 isoform X1 [Ursus arctos]	98%	9E-97	60.09	323

Supplementary table 1: Cluster summary of sequences shown in the alignment on figure S14. For clarity, 1 cluster every 10 cluster was retained for the alignment, except for sequence 5 which represents the closest ASGPR sequence. Clustered non-redundant database is the ncbi NR database clustered with each sequence within 90% identity and 90% length to other members of the cluster.

Alphafold	chi2 SAXS	dockQ
0	14,35	0,58
1	13,91	0,57
2	4,45	0,55
3	4,38	0,55
4	14,06	0,55
5	3,34	0,54
6	4,27	0,54
7	13,62	0,53
8	4,04	0,51
9	13,75	0,51

Supplementary Table 2: Alphafold models with their DockQ score and the χ^2 value with respect to the SAXS curve evaluated by crysol. Cluster 2 structures are colored in grey. The best matching structure is structure 5 with a χ^2 of 3.3 and the best from cluster 1, structure 7 (χ^2 13.6)

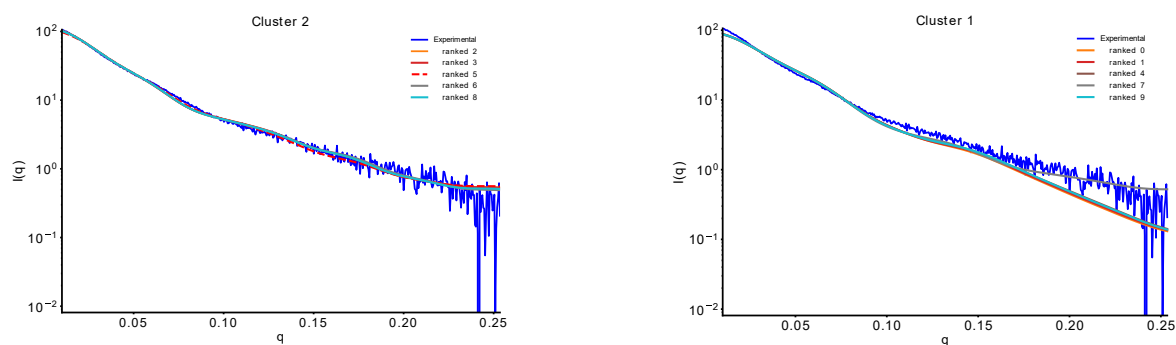


Fig S13 : Prediction of SAXS curves of the alphafold models and comparison with the experimental curve. χ^2 values are in SI table 1.

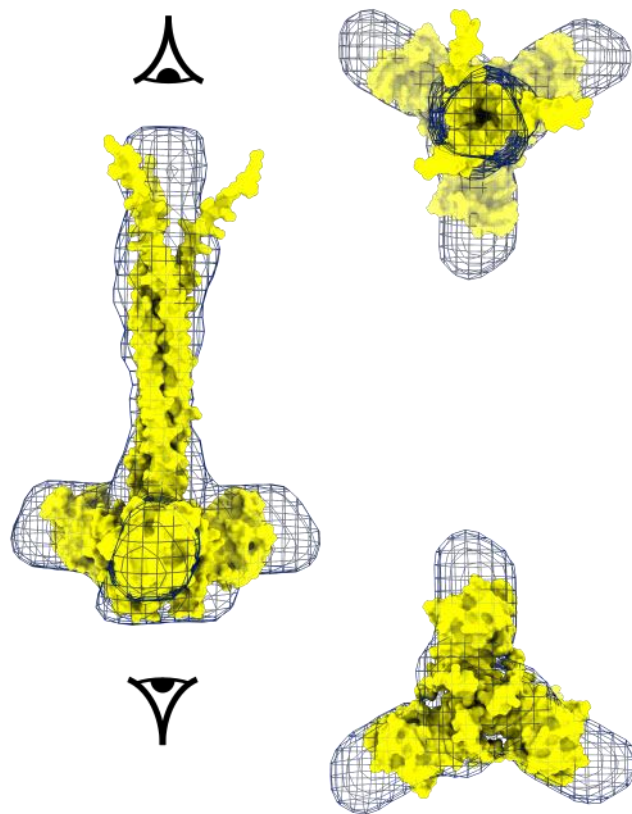


Fig S14 Comparison of cluster 1 structure with the best chi2 value with respect to the SAXS curve (ranked 7) is adjusted into SAXS calculated envelope.

References

- Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**:W320–W324. doi:10.1093/NAR/GKU316
- Shen Y, Bax A. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* **56**:227. doi:10.1007/S10858-013-9741-Y

V.7. Conclusion and discussion

Given the key role CLRs play in immune response modulation and homeostasis, they have been thoroughly studied and considered as potential therapeutic targets. Herein, we demonstrated the ability of MGL to bind *E. coli* R1 and R3 core OSs by an integrated approach using fluorescence microscopy, SPR, NMR, and a combination of SAXS and AlphaFold modelling. Fluorescence microscopy and flow cytometry primarily revealed the strong ability of MGL to bind *E. coli* R1 and R3 type core surfaces; SPR provided an estimation of the interaction affinity; ^1H - ^{15}N NMR was also used to identify a novel carbohydrate binding site distal to the canonical QPD motif.

Soluble versions of the R1 and R3 OSs induced backbone chemical shift perturbations on opposite surfaces of the MGL CRD and ECD canonical interaction site. This site is believed to be responsible for binding *E. coli* cores R1 and R3, and was confirmed using mutagenesis. A recombinant MGL containing a point mutation (D269) in the QPD canonical binding site in the CRD was generated. Remarkably, the D269H mutant protein also bound to *E. coli* cells and purified LOS R1 regardless of the QPD binding site⁸². Altogether, our findings revealed that MGL CRDs harbour two binding sites: a QPD Ca^{2+} -dependent binding site and a secondary glycan binding site.

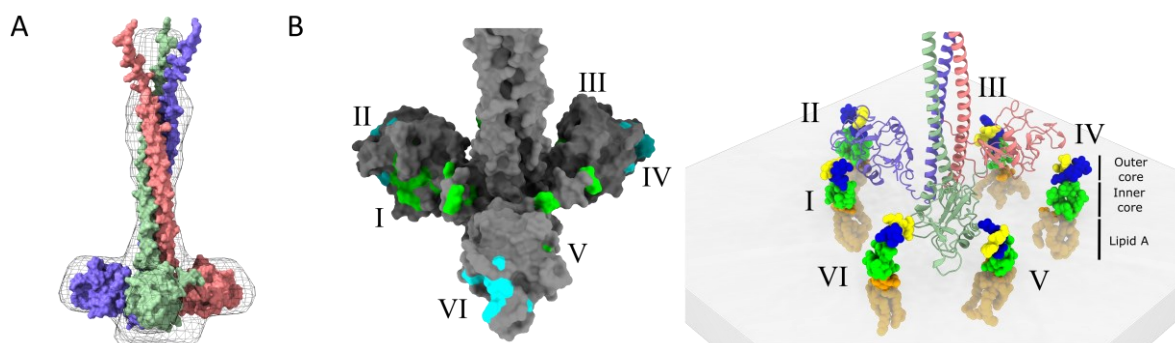


Figure 5.16: Combined SAXS-AlphaFold model of MGL-ECD. (A) Best AlphaFold model of MGL-ECD adjusted into the SAXS envelope. (B) CRD arrangement with up to six accessible glycan binding sites (I to VI). The NMR CSPs for the glycan binding sites are represented in cyan for GalNAc and green for OS R1. (B) Schematic view of 6 R1 LOS molecules on a surface facing the 6 glycan binding sites of MGL-ECD. Adapted from⁸².

The structure of MGL-ECD trimer is unknown. Combined data obtained from small-angle X-ray scattering (SAXS) experiments on MGL ECD with structures predicted by AlphaFold were used to build a model of the MGL ECD trimer. CRDs are positioned perpendicular to the coiled-coil domain (Figure 5.16A). Their orientation makes both the canonical QPD- and the secondary OS-binding sites available for binding glycans on a surface (Figure 5.16B) ⁸².

The presence of a secondary binding site in C-type lectins has already been reported for some CLR, including Langerin ²³⁵, Mincle ²³⁶, and DC-SIGN ²³⁷. Nevertheless, to our knowledge, the binding properties highlighted here for MGL are unprecedented for a C-type lectin. Secondary binding sites are thought to expand the lectin's ligand repertoire to ligands with different physicochemical features compared to the canonical binding site. That is the case for Mincle which binds the mycobacterial glycolipid Trehalose DiMycolate (TDM) resulting in immune response activation against mycobacteria. Binding at the secondary binding sites can induce allosteric changes within the canonical binding site. This was reported for Langerin for which the binding to heparin oligomers involves a secondary Ca²⁺-independent binding site and induces conformational changes of the primary site through the intradomain allosteric network ²³⁸.

The role of this new binding site in MGL glycan recognition cannot be determined from our data. However, its 3D arrangement suggests the accessibility of up to six glycan sites (2 sites per CRD), thus contributing to an increased glycan binding affinity. MGL CRDs orientation upon lateral rotation along the disulfide bridge, positioning CRDs perpendicular to the coiled-coil domain, can play a pivotal role in both the presentation of one or both glycan sites, and their subsequent plasticity to bind a variety of ligands. Thereby, this new binding site could contribute to the broad ligand spectrum MGL presents ranging from both Gram-positive and -negative bacteria including *S. aureus*, *C. jejuni*, *N. gonorrhoeae*, and *E. coli* glycans ^{81,100,101}.

The glycan preferences of the new binding site have to be characterized in order to conclude on its functional properties. This could be done by a combination of ligand screening using the MGL mutant lacking the canonical binding site, NMR monitoring of glycan binding, and finally mutagenesis of residues identified within this new binding interface. In addition, NMR-STD experiments would also shed light on the interaction as seen from the ligand side. The binding epitope of the ligand could thus be identified and it would be interesting to determine whether the binding is only limited to the outer core region of OS or whether residues from the inner core are also engaged in this interaction. Altogether, the presence of secondary binding

sites contributes to the regulation of lectins' functional diversity such as pathogen recognition, and immune response modulation. Characterizing these secondary binding sites and the specific binding events they are involved in represents promising opportunities for the development of selective targets towards a given lectin of interest. For instance, this second binding site could allow to selectively target MGL over the ASialoGlycoProtein Receptor-1 (ASGPR1) which shares the same affinity for GalNAc than MGL²³⁹.

The spatial configuration of MGL elucidates its ability to recognize various glycans on cell surfaces. We observed an increased affinity when studying MGL interaction with glycans presented on the cell surface compared to the interaction with isolated oligosaccharides or detergent solubilized LOS molecules. The binding strength remarkably increases from mM (when utilizing OS derivatives) and μM (when detergent solubilized LOS molecules), to nM when presented on the cell surface. Little is known about the arrangement of LPS at the cell surface and its influence on the binding. An important factor could be the dense network that LOS molecules present at the cell surface which is stabilized by divalent cations. These macroscopic structures are probably not conserved in detergent-solubilized micelles where detergent molecules could be inserted between LOS molecules, leading to more flexible structures in a less dense environment. This intriguing finding highlights the importance of ligand presentation and multivalence, hence demonstrates the need for innovative systems representative of the cell surface for LPS studies able to provide native-like information.

Furthermore, MGL binding to O-antigen carrying versions of R1 and R3 *E. coli* strains, O113:H21 and O157:H7 strains, respectively, is strongly influenced by this elongated sugar portion. The role of the O-antigen in pathogenic bacteria host defence evasion has already been described and reported for some strains including *Helicobacter pylori*, *Salmonella typhimurium* and *E. coli*^{232–234}. More investigations need to be carried out regarding MGL-pathogenic bacterial glycans interactions and consider to incorporate immunological assessments to the established *in-vitro* studies. Such framework would shed light on the role of MGL in the regulation of the immune response and subsequent pathogen evasion mechanisms. This would improve our understanding of infectious processes, and would also facilitate the development of new therapeutic strategies to counter bacterial infections and improve public health.

VI. Biomimetic LPS nanodiscs for interaction studies at the surface of Gram-negative bacteria

The outer leaflet of Gram-negative bacteria outer membranes is decorated with a crucial glycoconjugate component, LPS²¹ (Figure 6.1). LPSs are amphiphilic molecules composed of lipid A, bearing several acyl chains, and a polar moiety of glycans of variable lengths. They play a key role in bacterial virulence and antimicrobial resistance by forming an impermeable barrier. Due to their molecular structure and extension onto the cell surface, LPSs are deemed potent activators of the immune system²⁴.

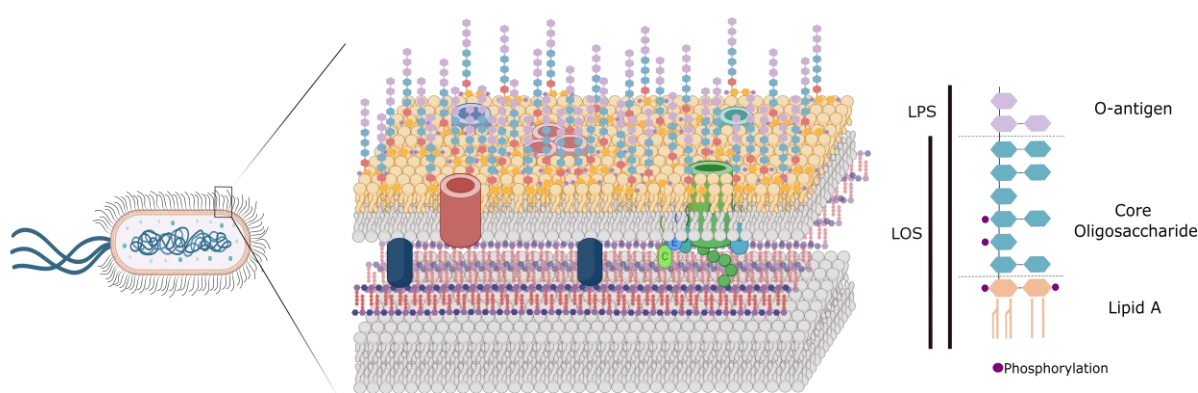


Figure 6.1: Overview of Gram-negative bacteria cell envelope. Overall organization of Gram-negative bacteria cell wall with a schematic representation of LPS chemical structure.

While LPSs are key element in bacterial integrity and recognition by the host, their study remains challenging. Purified LOS/LPS by extraction methods¹⁶⁴ assemble differently and adopt variable lipidic assemblies in solution. LOS assembles as vesicles of various sizes, while purified LPSs harbouring the long O-antigen form elongated micelles²⁴⁰. Given their amphiphilic nature, LPSs purification and fractionation have been a prerequisite for LPS studies. MS together with NMR spectroscopy played a central role in LPS structure determination, composition analysis, and deciphering of linkages between monosaccharides^{241,242}. Furthermore, LPS has been thoroughly studied, notably in model membrane systems, using a variety of biophysical approaches^{243–245} which significantly improved our understanding of the bacterial envelope. Despite major advances in LPSs research, numerous questions regarding their structure and mechanisms of action are still elusive. In this context, an increased emphasis has been placed on the development of innovative methodologies able

to provide native-like information on LPSs in particular, and membrane-related research in general.

Development of lipid nanoparticles, known as nanodiscs, has been a breakthrough in the study of membrane proteins (MPs) mainly through the use of amphipathic molecules such as helical proteins (MSP), peptides, or nucleic acids, generating nanometer-scale particles²⁴⁶. Although these approaches have contributed much in MPs research, they all require prior protein solubilization in detergent. An alternative detergent-free method has been developed. MP-containing nanodiscs are directly obtained from cellular membranes thus conserving a patch of native lipids around the protein. Amphiphilic styrene maleic acid copolymers (SMA) insert spontaneously into native or reconstituted membranes and form discoidal nanoparticles of high thermal stability²¹⁸. This novel system has become increasingly popular. Its ability to conserve the native lipid environment of membranes opened numerous perspectives for the use of SMA nanoparticles in a wide range of applications, notably in developing cell membrane-based vaccine formulations²⁴⁷.

Herein, we have used the SMA copolymers to form nanodiscs from purified LPSs or form LPS nanodiscs by direct extraction from bacterial outer membranes of laboratory and pathogenic *E. coli* strains. Among which we selected two *E. coli* strains lacking or not the O-antigen moiety F470 and O157:H7, respectively, found in verotoxigenic *E. coli* clinical isolates known as enterohemorrhagic *E. coli*²⁶. LPS-SMA nanodiscs were validated as cell surface mimetic models by several biophysical methods, and were also used to monitor interactions with MGL receptor and Polymyxin B antibiotic. Overall, this work shows that LPS nanodiscs are a promising approach for the study of structure and interactions of LPSs.

VI.1. Isolation and characterization of SMA LPS nanodiscs from *E. coli* strains

To assess the feasibility of LPS nanodiscs formation with SMA, it was first attempted on LOS and LPS purified from *E. coli* F470 and O157:H7 strains, respectively. LOS are extracted using the Phenol/Chloroform/light Petroleum (PCP) whereas the phenol/hot water method is used for the extraction of LPS¹⁶⁴. Purified LOS/LPS contained in the phenol fractions, assembled as large heterogeneous membrane structures in water (Figure 6.2B), were incubated with SMA(2:1) (referred to as SMALP 200 on Cube Biotech) (Figure 6.2A). Remaining insoluble material was removed by ultracentrifugation, and the supernatant was subjected to a sucrose density gradient with 5% / 25% / 45% (w/v) to remove residual polymers. The majority of LPS was found in the sucrose gradient fractions corresponding to 25-45%. These fractions were recovered, dialysed and analysed. Negative stain electron microscopy confirmed nanodiscs formation and showed circular particles presenting a larger size inhomogeneity than observed for protein based nanodiscs systems consistent with previous reports on SMA based nanodiscs²⁴⁸ (Figure 6.2B).

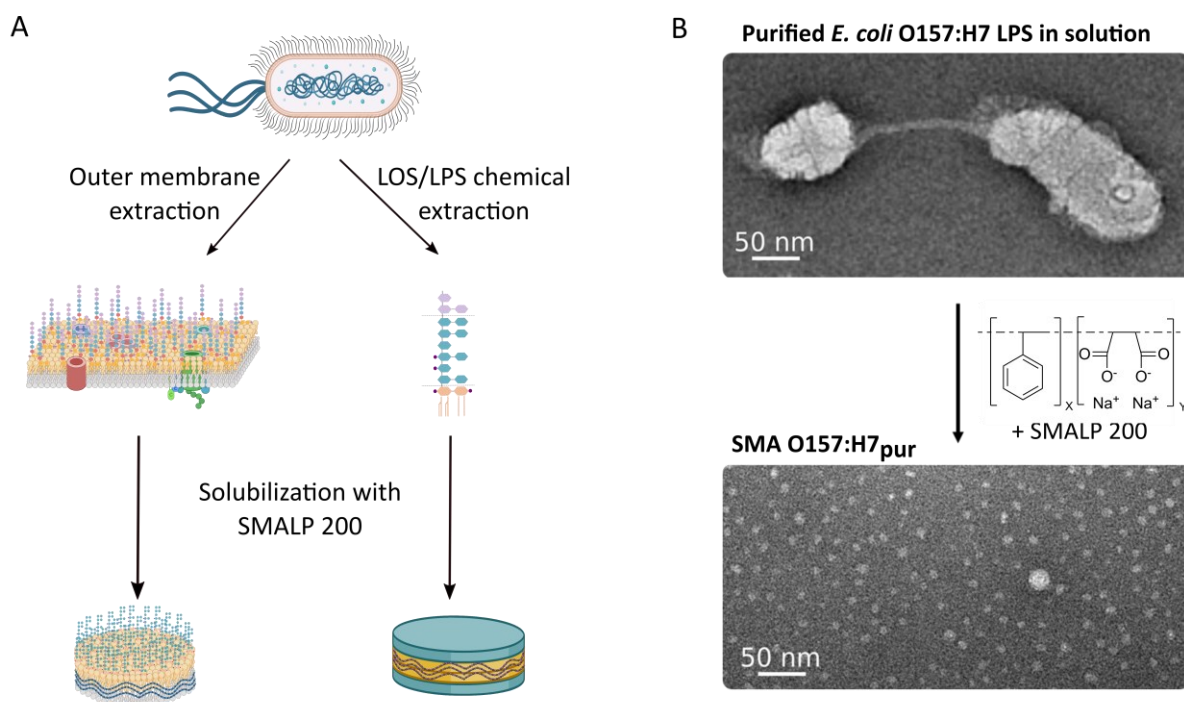


Figure 6.2: Preparation and characterization of SMA nanodiscs. A) Simplified protocol for nanodiscs preparation from both outer membranes and purified LPS. B) Negative staining EM images of *E. coli* O157:H7 purified LPS before and after solubilization with SMALP 200.

Formation of LOS/LPS nanodiscs from purified samples was successful. They provide valuable LPS membrane models, but they are mostly constituted of LPS and are then not fully representative of the asymmetric bacterial outer membrane. It was thus attempted to form LPS nanodiscs directly from purified bacterial outer membranes, with their native arrangement (Figure 6.2A). After bacterial disruption, outer membranes were separated from inner membranes *via* a sucrose gradient²⁴⁹, and treated with SMA polymer followed by a 2nd sucrose gradient for nanodiscs isolation similarly as with purified LOS/LPS.

In parallel to R1 and O157:H7 *E. coli* strains, we prepared SMA nanodiscs from outer membranes of other strains that are studied in our laboratory. This includes *E. coli* R3 (presenting the same core OS as O157:H7 *E. coli*), as well as two laboratory *E. coli* K12 strains. One being the wild type strain and an *E. coli* K12 strain harbouring an O16 O-antigen moiety. The latter was built by our collaborators in Milano (Università degli Studi di Milano, Department of Pharmacological and Biomolecular Sciences) following transformation of the wild type strain with PMF19 plasmid. This plasmid harbours a streptomycin resistance cassette and allows constitutive expression of WbbL, which restores O16 antigen production in wild type strains and generated an *E. coli* K12 O16 stain.

The preparation protocols were adapted for each strain regarding the optimal sucrose gradient for outer membranes isolation. Indeed, *E. coli* R1 and R3 types membranes were better separated using sucrose gradients of 20% / 45% / 73% (w/v). The LOS molecules residing in the outer membrane induce envelope separation into an upper low-density inner membrane, and a high-density outer membrane. Thus, the inner membrane is found at the 20-45% interface whereas the outer membrane is separated at the 45-73% interface (Figure 6.3). On the other hand, membranes of the other strains (*E. coli* K12 WT, *E. coli* K12 O16, and *E. coli* O157:H7) were found to be properly separated using rather the 20% / 53% / 73% (w/v) sucrose gradient. In this case, the inner membrane is separated from the outer membrane and localizes to the 20-53% sucrose interface while the outer membrane layer is found at the 53-73% sucrose interface. (Figure 6.3).

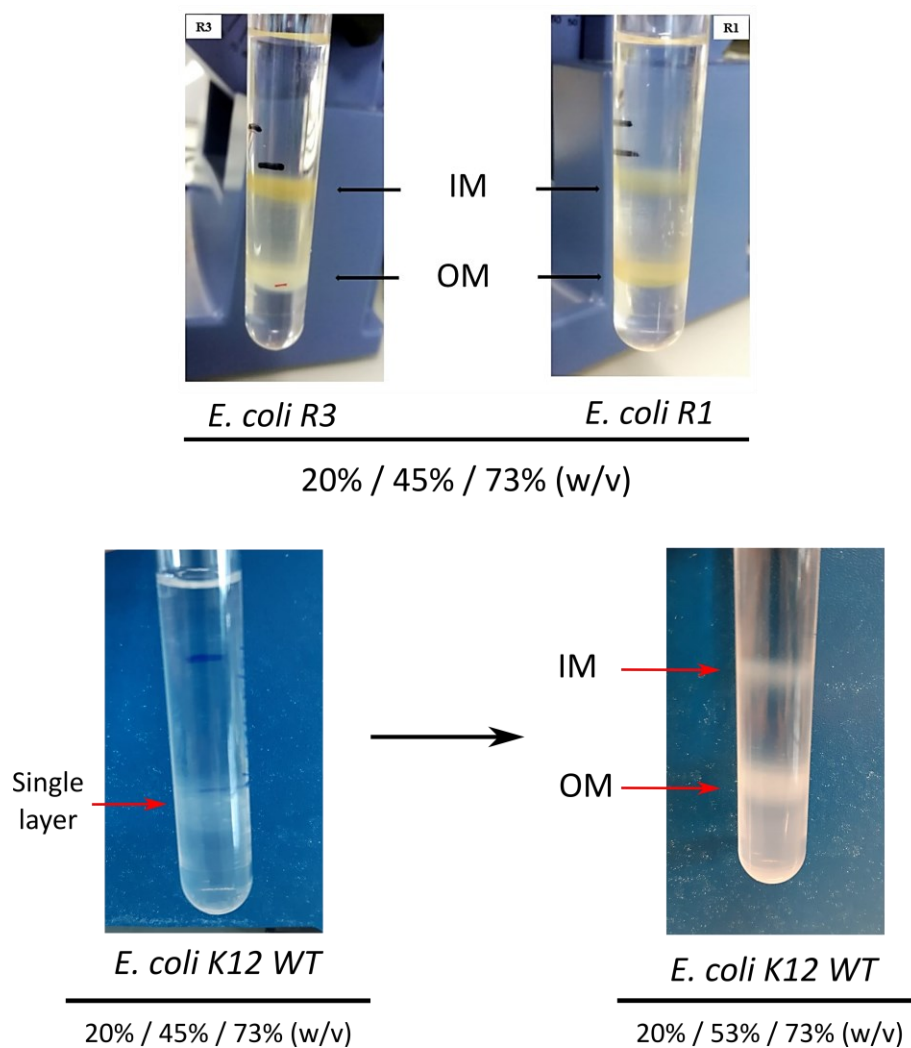


Figure 6.3: Representative images of different Gram-negative membranes separation sucrose density gradients. Top: *E. coli* R1 and R3 membranes separation using the 20%/45%/73% (w/v) sucrose gradient. The inner membrane localizes to the 20-53% sucrose interface while the outer membrane localizes to the 53-73% sucrose interface. Bottom: *E. coli* K12 membranes didn't separate using the 20%/45%/73% (w/v) sucrose gradient, but did separate using the 20%/53%/73% (w/v) sucrose gradient.

The size distribution of the prepared nanodiscs was assessed by Dynamic Light Scattering (DLS) and negative staining EM. DLS relies on the Brownian motion exhibited by dispersed particles in solution in all directions. A laser beam illuminates the sample, and the fluctuations in the scattered light are detected at a scattering angle θ . Fluctuations rate correlates with the diffusion rate, which in turn depends on the hydrodynamic radii (R_h) of the molecules. Consequently, smaller particles move at higher speeds and exhibit faster diffusion and fluctuation rates compared to their larger counterparts. Thereby, by extracting the diffusion

coefficient, particles size (R_h) determination can be done through application of the Stokes-Einstein equation. For all the prepared nanodiscs, this method revealed the presence of monodisperse nano-objects ranging from 15-30 nm diameter (Figure 6.4). Their morphology was characterized by negative staining EM revealing a population of circular particles consistent with previous reports on SMA-stabilized nanodiscs²⁵⁰ (Figure 6.4).

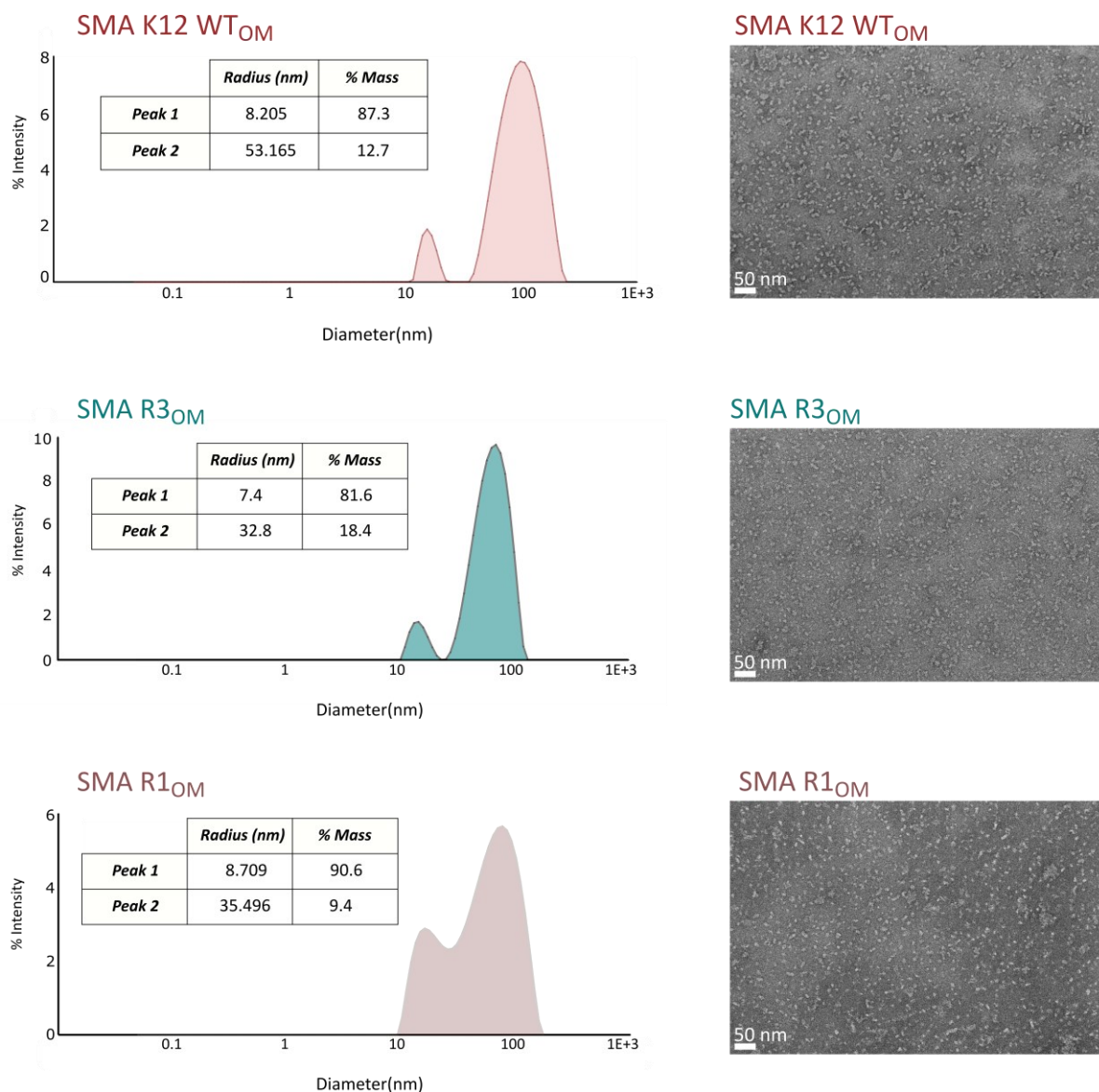


Figure 6.4: outer membrane nanodiscs characterization using DLS and negative staining EM. (Left) Particle size distribution curves of SMA K12 WT_{OM}, SMA R3_{OM}, and SMA R1_{OM} nanodiscs measured by DLS. (Right) Negative staining EM images of SMA K12 WT_{OM}, SMA R3_{OM}, and SMA R1_{OM} nanodiscs. Negative-stain grids were prepared using the mica-carbon flotation technique and stained using 2% sodium silicotungstate (SST). Scale bars: 50 nm.

Among the nanodisc types produced, SMA R1_{pur}, and SMA R1_{OM}, SMA O157:H7_{pur} and SMA O157:H7_{OM} were later characterised and compared in terms of size distribution and thickness by AFM. The AFM imaging was performed by our collaborators Dr Jean-Luc Pellequer and Dr. Jean-Marie Teulon, AFM imaging group within the IBS. Given the negative charge of SMA polymers, nanodiscs were adsorbed on a mica surface coated with Ni²⁺ ions. The samples were imaged in ambient air, their topographies were recorded and revealed images similar to what was obtained by negative staining EM (Figure 6.5A). Surfaces and heights of the four samples were measured (Figure 6.5B, Annex 7). Surface measurements of the 4 nanodiscs preparations allowed to estimate an average surface of around 940 nm² and an average diameter of about 30 nm ± 5nm. In the case of SMA nanodiscs of purified LPSs, as we consider that they are only constituted of LPS molecules, this diameter corresponds to an estimation of around 500 LPS per disk when considering an average area of each LPS molecule of 190 Å²²⁵¹. On the other hand, SMA-OM nanodiscs are composed of a mixture of membrane components including LPS, phospholipids, and membrane proteins. Therefore, although we cannot estimate the LPS content in outer membrane nanodiscs, we expect it to be lower compared to nanodiscs of purified LPSs. Reported diameters of SMA nanodiscs are mostly around 10 nm knowing that the corresponding solubilization experiments have been carried out with an excess of SMA. The low SMA-to-lipid ratio may result in an increase of nanodiscs size²⁵², and thus can explain the size differences compared to our preparations.

The height of the nanodiscs determined by AFM was found to be ~17 nm (Figure 6.5B). Reported thicknesses of LOS liposomes estimated using SANS measurements are of about 4 nm. However, the actual bilayer thickness could be underestimated in the SANS experiment due to the high hydration of the glycan moieties which would generate similar scattering profile as the aqueous bulk. Thus, these estimations more likely represent the thickness of the hydrophobic layer rather than the whole thickness of the LOS bilayer^{253,254}.

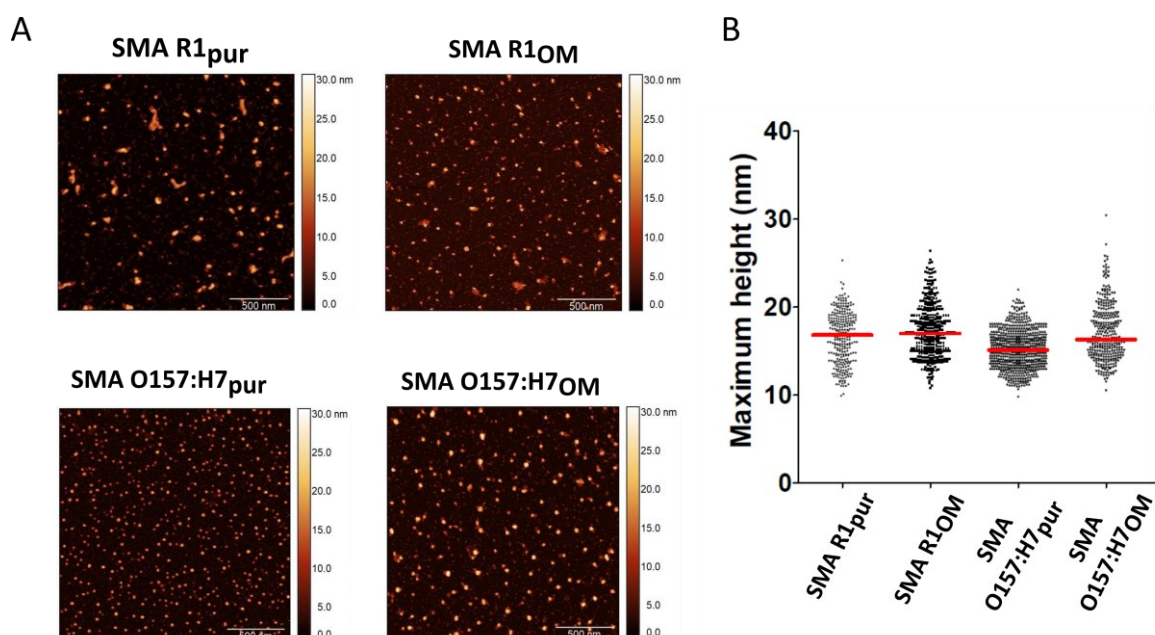


Figure 6.5: AFM characterization of SMA nanodiscs. AFM images of SMA nanodiscs (scale bars: 500nm) (A) and heights comparison of the different nanodiscs (B). Samples were loaded on a freshly cleaved mica surface coated with Ni^{2+} ion. Sample solution was applied and deposited for 2 min on the Ni-coated mica surface and dried under a nitrogen flow. Imaging was performed in air using the PeakForce Tapping mode. Height and surface area were obtained from two different images.

VI.2. Structural analysis of LPS nanodiscs by solid-state NMR

NMR spectroscopy has been extensively used for the structure elucidation of LPS and for epitope mapping of LPS-protein interactions. Here we dispose of LPS nanodiscs of different chemical nature, composition of the lipid bilayer but with the same overall morphology. These objects are too large for solution NMR as their slow tumbling rate leads to significant signal broadening. However, they can be studied by solid-state Magic Angle Spinning (MAS) NMR. Spinning the sample at high frequency around an axis tilted at the magic angle (54.74°) with respect to the static magnetic field B_0 averages out anisotropic interactions, thus enabling atomic resolution analysis of large systems.

ssNMR experiments under MAS allow the acquisition of a different set of 2D ^1H - ^{13}C correlation experiments. ssNMR employs INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) and CP (Cross-Polarization) based experiments to enhance the signal of insensitive nuclei (in this case ^{13}C). In INEPT-based experiments, originally developed for solution NMR, the magnetization is transferred through scalar (J) couplings, and is effective for the detection

of the most flexible parts within the studied system. In contrast, CP-based experiments rely on the use of dipolar couplings for magnetization transfer. Thereby, it is particularly effective in the detection of rigid parts of the studied system. In the context of LPS glycoconjugates, each constituting moiety will be effectively detected using either INEPT- or CP-based correlation experiments (Figure 6.6). For instance, the flexible O-antigen moiety with its extension from the membrane is detected using scalar (J) coupling methods. It thus can be seen in both INEPT-based ssNMR and solution NMR experiments where CP transfer is less efficient (Figure 6.6). On the other hand, the more rigid parts proximal to the core of the bilayer including the core OS, particularly the inner core, and lipid A are mostly detected in CP-based experiments (Figure 6.6). Together, by combining the complementary data provided by CP and INEPT experiments, we can get insight into the overall structure and dynamics of LPS molecules.

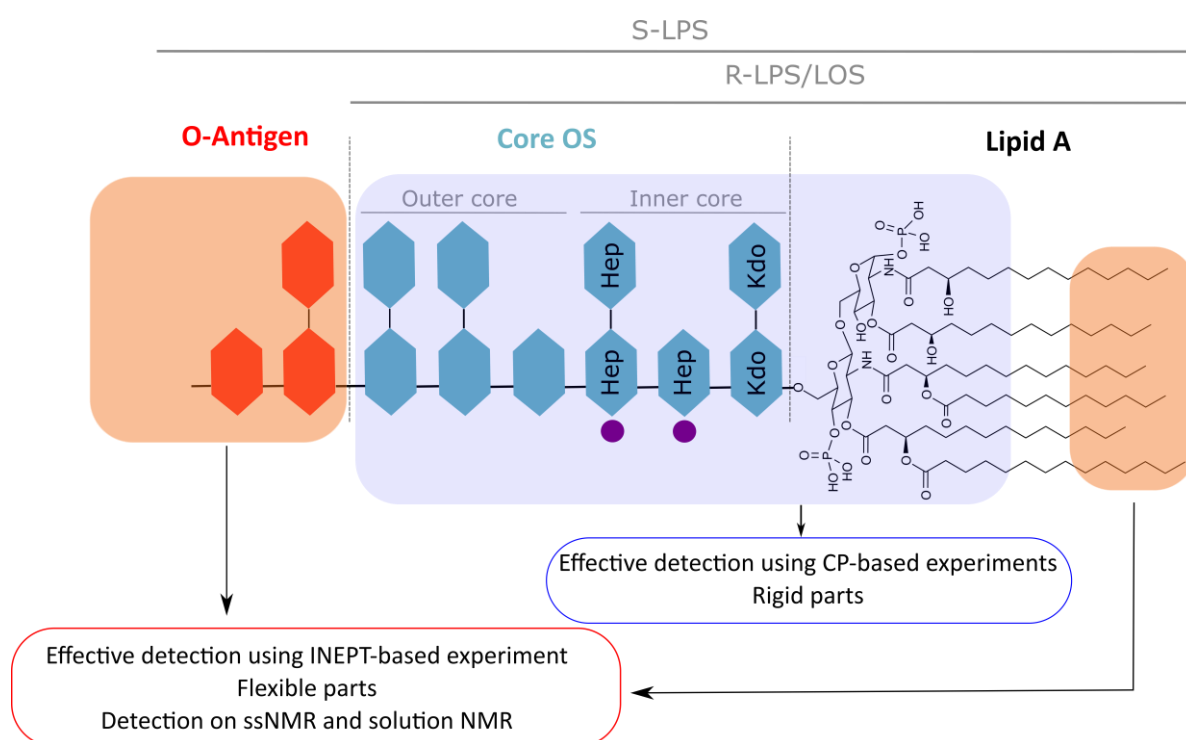


Figure 6.6: NMR detection of different LPS regions. LPSs flexible regions highlighted in orange are effectively detected using scalar (J) couplings methods in both INEPT-based ssNMR and solution NMR experiments. Whereas, LPSs rigid parts highlighted in light purple are detected using CP- based experiments.

Beside MAS-NMR we also used solution NMR as a first and quick approach for sample quality assessment. Recording a 1D ^1H spectrum in solution shows the overall components of the studied system could be observed, including lipid signals, sugar signals, and also SMA polymer resonances (Figure 6.7).

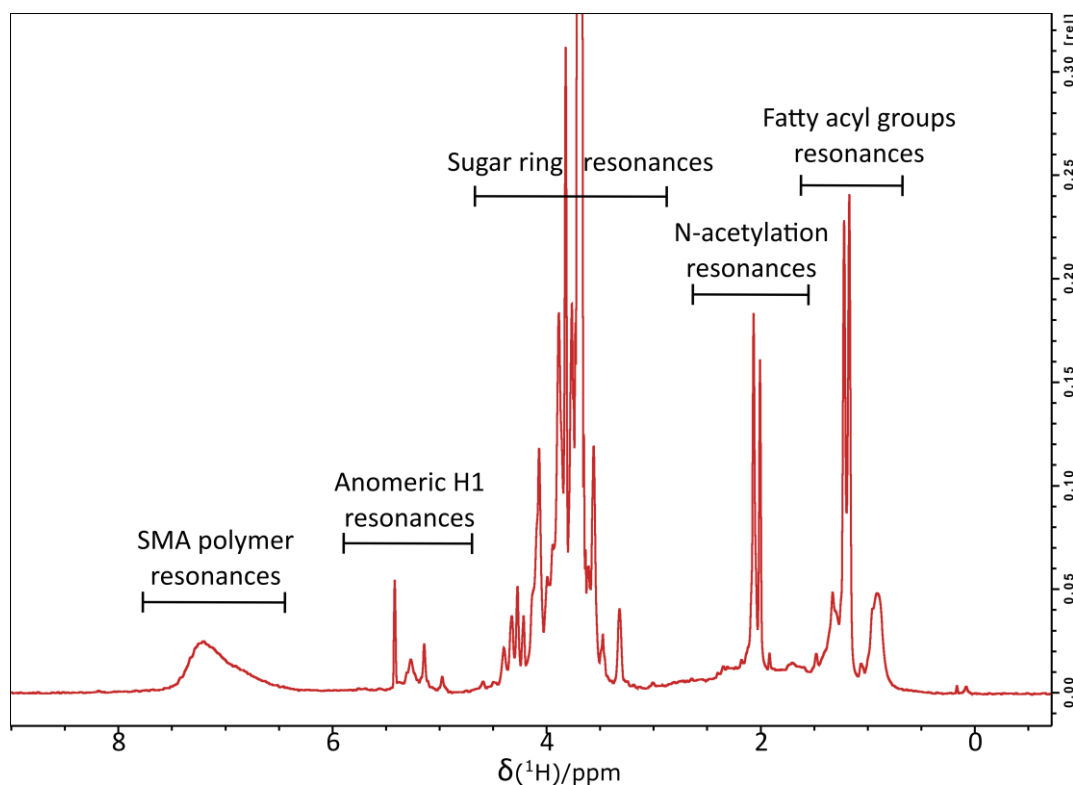


Figure 6.7: Solution NMR 1D ^1H spectrum of SMA O157_{pur}. The sample was prepared in 25 mM Tris pH 7.4 100 mM, 100% D₂O, and NMR experiments were recorded at 323 K. The chemical shift regions corresponding to the components of the studied system are indicated. SMA: Styrene Maleic Acid.

A 2D ^1H - ^{13}C correlation spectrum of LPS glycoconjugates consists of three main distinguishable regions. An anomeric region (H₁, C₁) found at 4.5-5.5 ppm for ^1H and at 90-110 ppm for ^{13}C chemical shifts. The sugar ring resonances (H₂→H₆ / C₂→C₆) are found at 3.2-4.4 ppm for ^1H and at around 50-75 ppm for ^{13}C chemical shifts. Lastly, aliphatic groups, corresponding to acyl chains and/or deoxy-sugar signals appear at around 1-2.2 ppm for ^1H chemical shifts and ^{13}C chemical shifts at around 15-30 ppm.

Experimentally, when recording 2D ^1H - ^{13}C solution NMR experiments on O-antigen carrying samples, the obtained spectra are dominated by O-antigen resonances, which is not surprising given the high number of glycan residues in this portion and its high flexibility (Figure 6.8). Regarding SMA O157_{pur} nanodiscs spectrum (Figure 6.8A), four signals

corresponding to the anomeric position of the sugar residues constituting the O-Antigen are observed (Glc, PerNAc, GalNAc, and Fuc). However, no signal could be observed for the core OS residues presumably due their slower dynamics and a lack of sensitivity compared with the pronounced signals originating from the flexible O-Antigen portion.

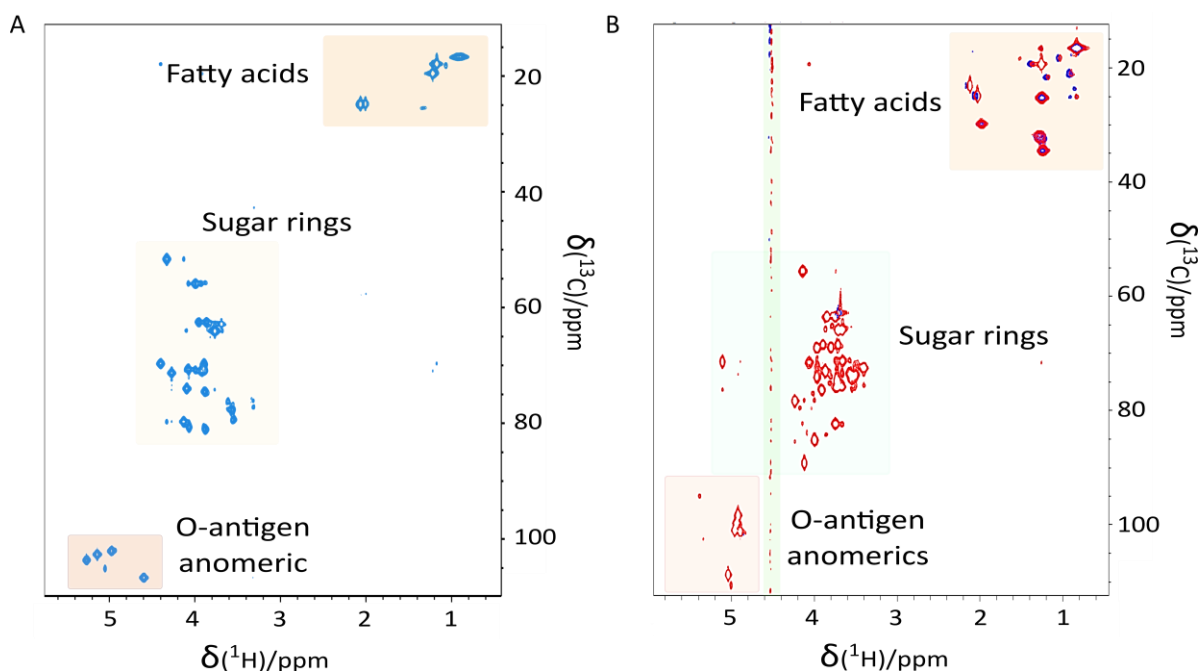


Figure 6.8: Solution NMR ^1H - ^{13}C correlation spectra of the O-antigen carrying nanodiscs. (A) Correlation spectra of SMA 0157_{pur}. (B) Overlay of correlation spectra of SMA K12-O16_{OM} (red) and SMA K12wt_{OM} (blue). Resulting spectra shows only O-antigen resonances, whereas no core OS signals were detected. The samples were prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were recorded on a 600 MHz spectrometer at 323 K. The chemical shift corresponding to the different detected regions are highlighted.

On the other hand, core OS signals have not been observed for O-antigen free samples neither (Figure 6.8B, 6.9). This can be explained by their lower molecular tumbling and proximal distance of the core OS to the membrane core part of the system. Surprisingly, when treating these O-antigen-free nanodiscs with EDTA, the obtained spectrum contained all the expected resonances of the LOS, including the previously not detected signals from the core OS (Figure 6.9). EDTA is known to chelate divalent ions that stabilize the membrane, which results in LPS disaggregation. However, when LOS liposomes are treated with EDTA, this does not improve the NMR spectrum and no core OS resonances are observed. Therefore, we assume that EDTA disrupts the nanodiscs while the remaining SMA polymer acts as a detergent to keep

LPS soluble. Disrupting the LPS nanodiscs overall organization leads to an increase in dynamics, rendering these components detectable in solution state NMR.

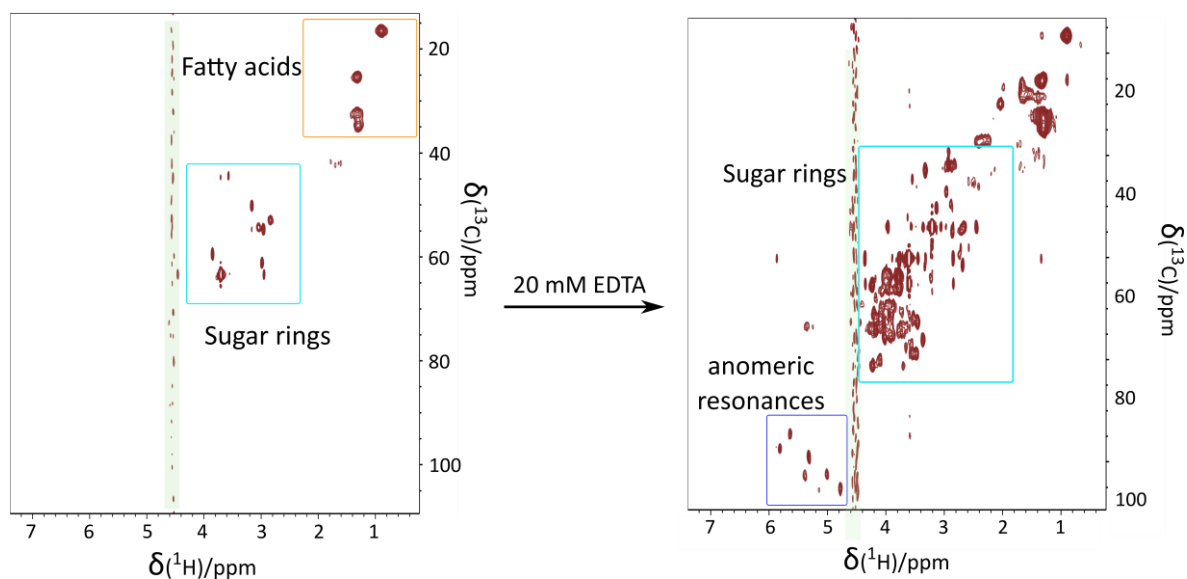


Figure 6.9: Solution NMR ^1H - ^{13}C correlation spectra of SMA R1_{pur} nanodiscs. The resulting spectrum did not reveal any core OS resonances in the anomeric region (Left). Upon addition of 20 mM EDTA, the obtained spectrum showed all the distinctive resonances of the LOS molecules, including anomeric core OS resonances. Samples were prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were recorded on a 600 MHz spectrometer at 323 K. The chemical shift corresponding to the different detected regions are highlighted

We next considered analysing SMA-LPS/LOS nanodiscs using solid-state NMR. In our case, around 3 mg of ^{13}C -labeled LPS/LOS in SMA nanodiscs were sedimented inside a 1.3 mm rotor, spun at 55 kHz MAS frequency, and analysed using ssNMR experiments relying on CP and INEPT. In order to find the optimal condition for spectra acquisition, different temperatures ranging from 25°C to 50°C have been investigated. For that, ^1H - ^{13}C correlation spectra were recorded at each temperature, and the spectral quality was evaluated (Figure 6.10A). Given the complexity of nanodiscs spectrum, the sugar anomeric resonances ($\text{H}_1\text{-C}_1$) being well resolved in the 90-105 ppm region in the ^{13}C dimension were selected to monitor the spectral quality in terms of signal to noise ratio. NMR experimental parameters, including hard pulses and CP conditions essential for efficient magnetization transfer resulting in maximum signal intensity with improved signal-to-noise ratio, were optimized. Following temperature monitoring, it turned out that the most informative spectra where most of the sugar anomeric signals could be detected were the ones recorded at higher temperature (Figure 6.10B). Thus, as SMA-based nanodiscs stand out for their thermal stability^{218,219}, we selected 50°C as the sample temperature for spectra acquisition.

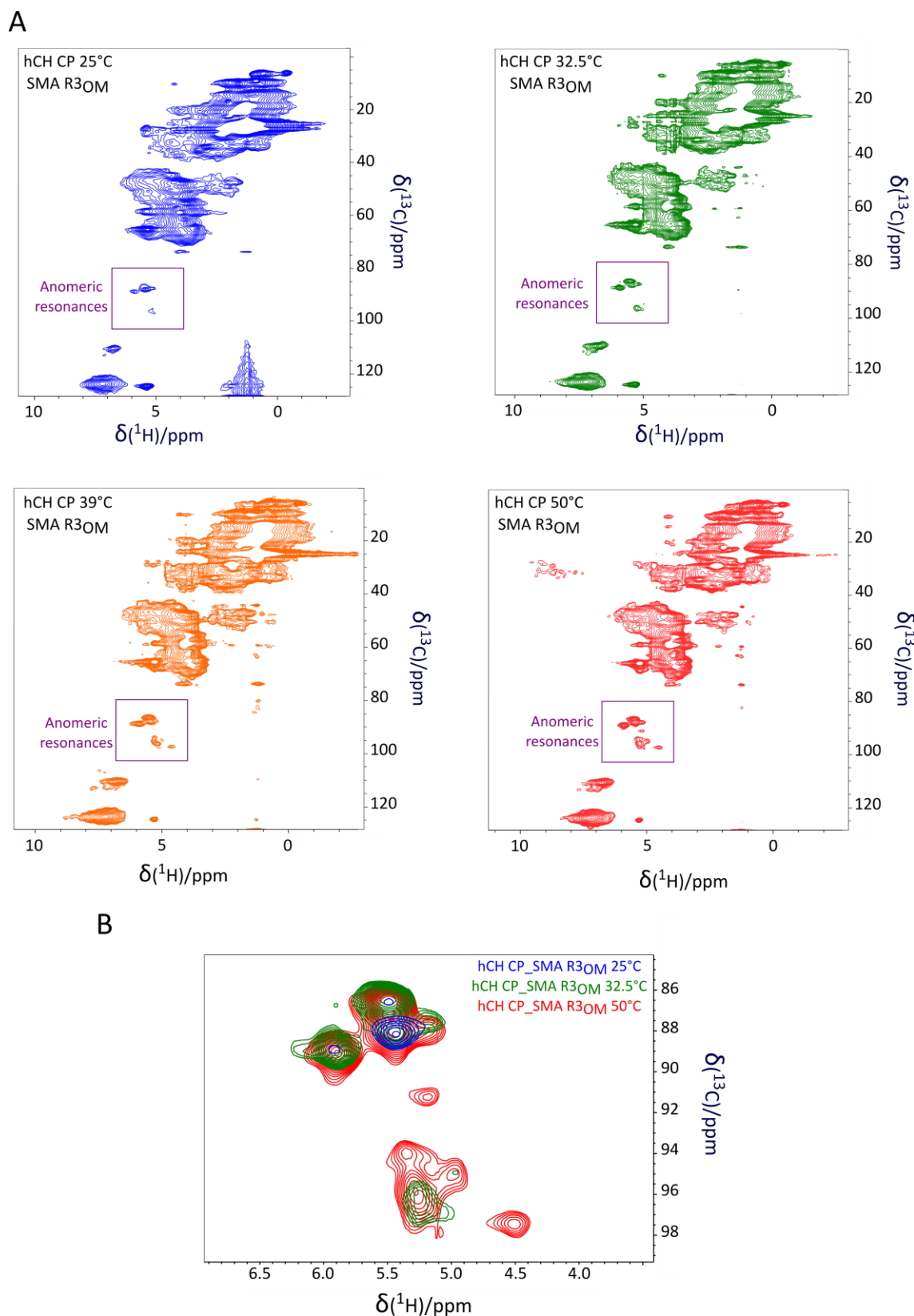


Figure 6.10: Spectra quality monitoring following temperature variation. (A) ^1H - ^{13}C hCH CP correlation spectra of SMA R_{3OM} nanodiscs at 25°C (blue), 32.5°C (green), 39°C (Orange), and 50°C (red). (B) Zoom on the anomeric region of the overlaid ^1H - ^{13}C hCH CP correlation spectra recorded at 25°C (blue), 32.5°C (green), and 50°C (red). The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were performed on a 600 MHz spectrometer at 55 kHz MAS and a sample temperature of 50°C.

In parallel to monitoring temperature effects on spectra quality, we considered getting insights into the phase transition temperature of the nanodiscs system. For that perspective, the variation of intensity of some fatty-acid methylene group signals, which are temperature-sensitive and thus indicative of the phase transition, is monitored. That is performed by recording ^1H - ^{13}C CP-based spectra of the sample at different temperatures (25, 32.5, 42 and 50°C). By plotting these resonance intensities as a function of temperature, a curve is traced showing gradual intensity shifts as a function of the temperature (Figure 6.11).

As only four data points were monitored, the phase transition temperature was approximately determined using the linear interpolation method given by $y=y_1+((x-x_1)\times(y_2-y_1)/(x_2-x_1))$ which can be written in our case as $S_{tr}=T_1+((T_{tr}-S_1)\times(T_2-T_1)/(S_2-S_1))$ where (T_1, S_1) and (T_2, S_2) are known data points, S_{tr} is the signal intensity at the phase transition temperature T_{tr} . With the set of data points we dispose, the phase transition seems to occur between $T_1=25^\circ\text{C}$ and $T_2=39^\circ\text{C}$, T_{tr} was determined using the formula: $T_{tr}=T_1+((S_{tr}-S_1)/(S_2-S_1)\times(T_2-T_1))$. T_1 and T_2 correspond to 25 and 39, respectively, while S_1 and S_2 are their respective signal intensities. S_{tr} is calculated as $(S_1+S_2)/2$. The phase transition from the gel to the fluid phase is estimated to occur around a temperature of 32°C . Reported phase transition temperature of LPS in model membranes measured by Attenuated reflectance-Fourier transform InfraRed spectroscopy (ATR-FTIR) was found to occur at $36.2 \pm 1.6^\circ\text{C}$ ²⁴⁵. The evaluation of nanodiscs phase transition temperature value is still within the reported temperature range although it was approximately determined from a limited number of data points. Furthermore, slight differences can also arise from compositional differences between the studied systems and their analysis by two distinct experimental methods presenting different sensitivities and experimental conditions.

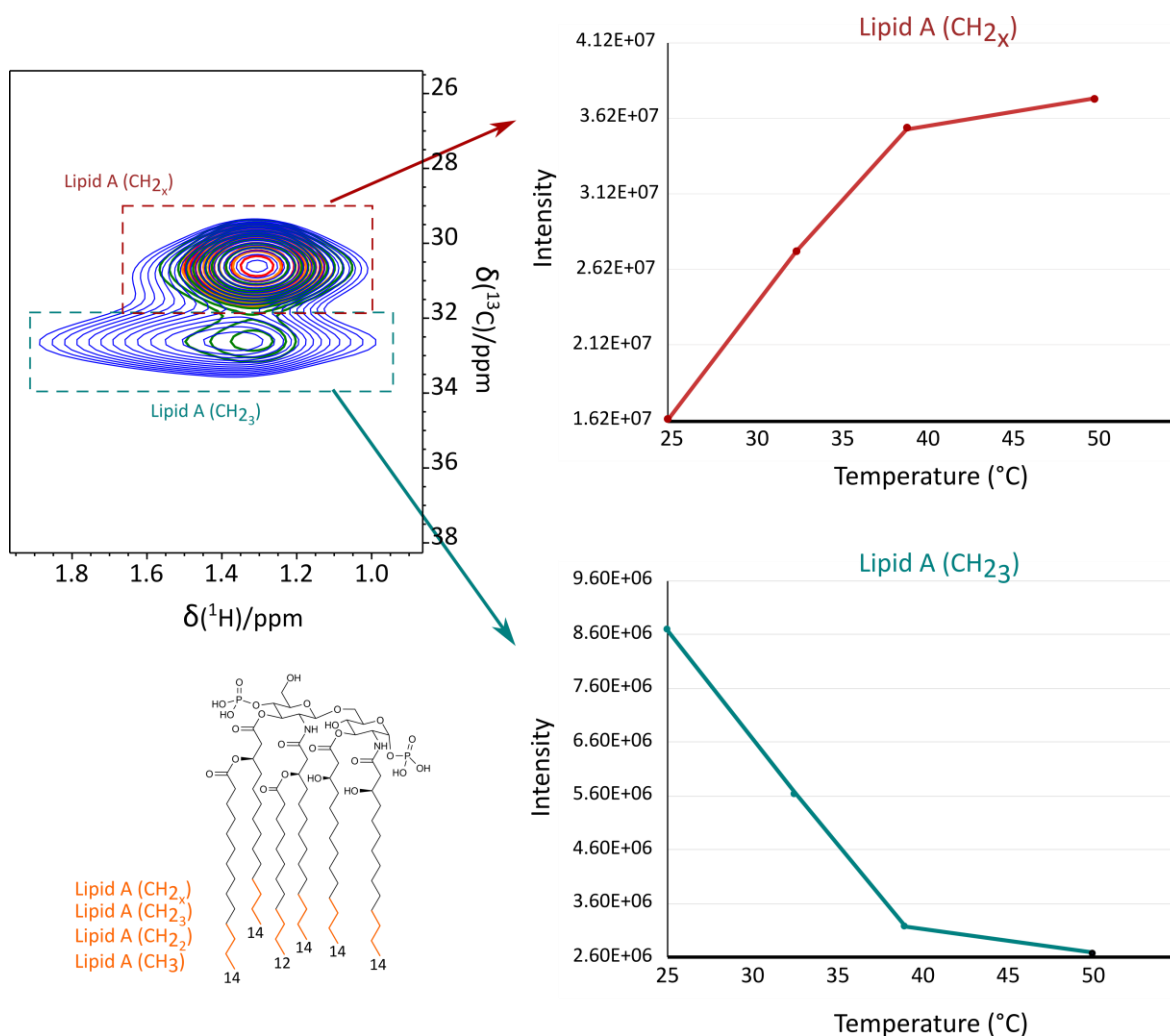


Figure 6.11: Monitoring LPS nanodiscs phase transition temperature using solid state NMR spectroscopy. Left: Zoom on the fatty-acid methyls groups region of the overlaid ^1H - ^{13}C hCH CP correlation spectra recorded at 25°C (blue), 32.5°C (green), and 50°C (red). Right: Plotting curve of the variation of the monitored groups' intensity as a function of temperature. The monitored lipid A groups are represented. CH_{2x} stands for CH_2 groups at the middle of lipid A acyl chains that are all chemically equivalent.

Following the required parameter optimizations made so far, SMA R1_{pur} nanodiscs were the first to be studied by ssNMR. The recorded ¹H–¹³C correlation experiments showed most of the expected signals corresponding to the distinct LOS structure regions (Figure 6.12). OS signals including anomeric signals superimposed perfectly with a reference spectrum of the purified delipidated oligosaccharide (OS) recorded in solution NMR. Thereby, the assignment was achieved by resonance transfer from the delipidated OS sample to the nanodisc sample (Figure 6.12, Annex 8).

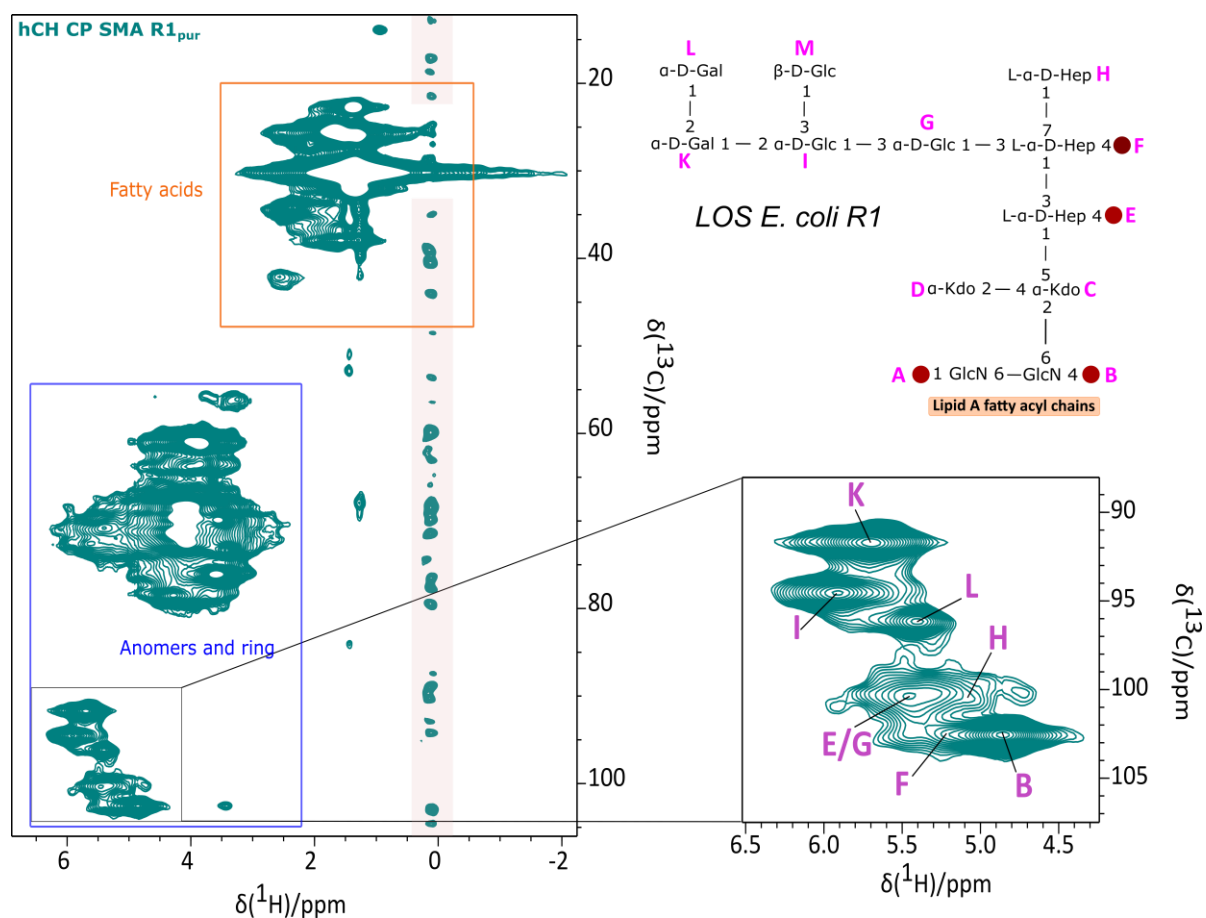


Figure 6.12: Solid state NMR ¹H–¹³C hCH CP correlation spectrum of SMA R1_{pur} nanodiscs. Left: Full ¹H–¹³C correlation spectrum on which chemical shifts corresponding to the different detected LOS regions are highlighted. Right: Zoom on the anomeric sugar region resonances with their corresponding assignments. The chemical structure of LOS R1 is indicated with the annotations used for the assigned residues. The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were performed on a 600 MHz spectrometer at 55 kHz MAS and a sample temperature of 50°C.

For nanodiscs systems harbouring O-antigen moieties, solid-state NMR experiments could only be acquired on O157:H7 LPS, given that the K12 O16 rotor got unfortunately crashed during analysis. O157:H7 LPS was thus investigated in pure or outer membrane derived nanodiscs. The ^{13}C - ^1H INEPT spectrum of SMA-O157:H7_{pur} is predominantly showcasing the O-antigen portion for which resonances could be assigned under MAS (Figure 6.13A, Annexes 9 and 10). Conversely, the SMAO157:H7_{OM} ^{13}C - ^1H INEPT spectrum shows many signals characteristic for the outer membrane, in addition to LPS signals. These include phospholipid signals such as phosphatidylethanolamine (PE), and phosphatidylglycerol (PG) (Figure 6.13B).

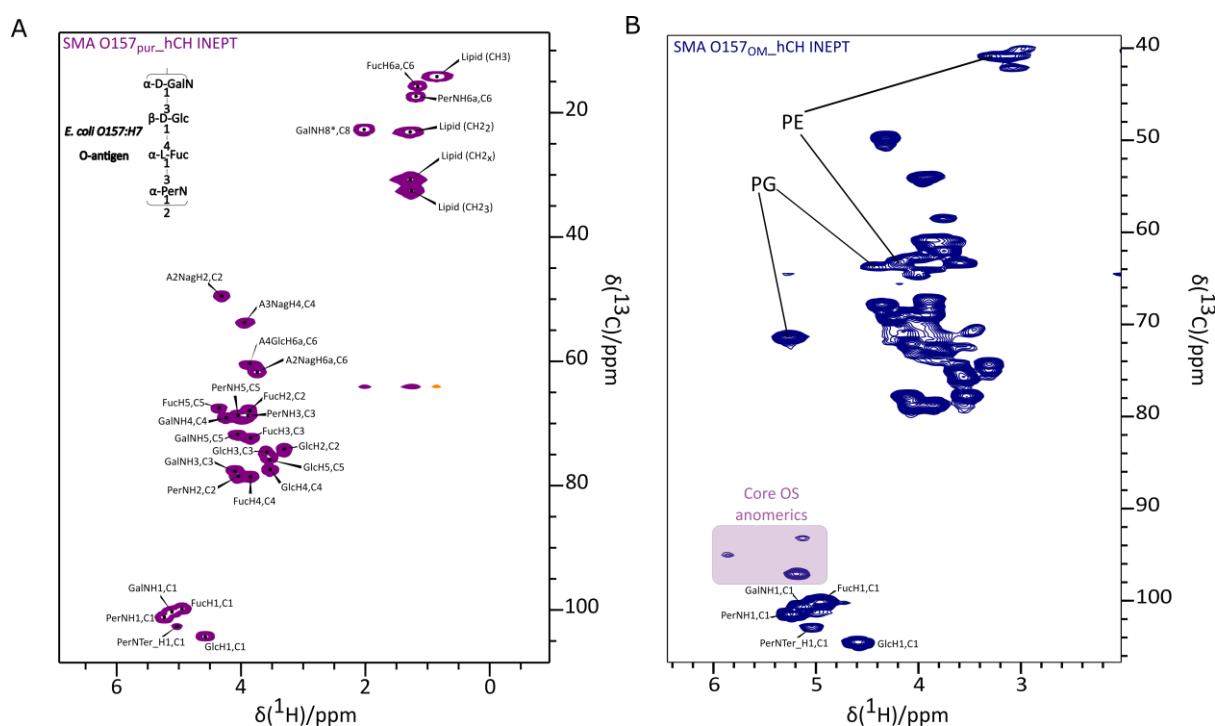
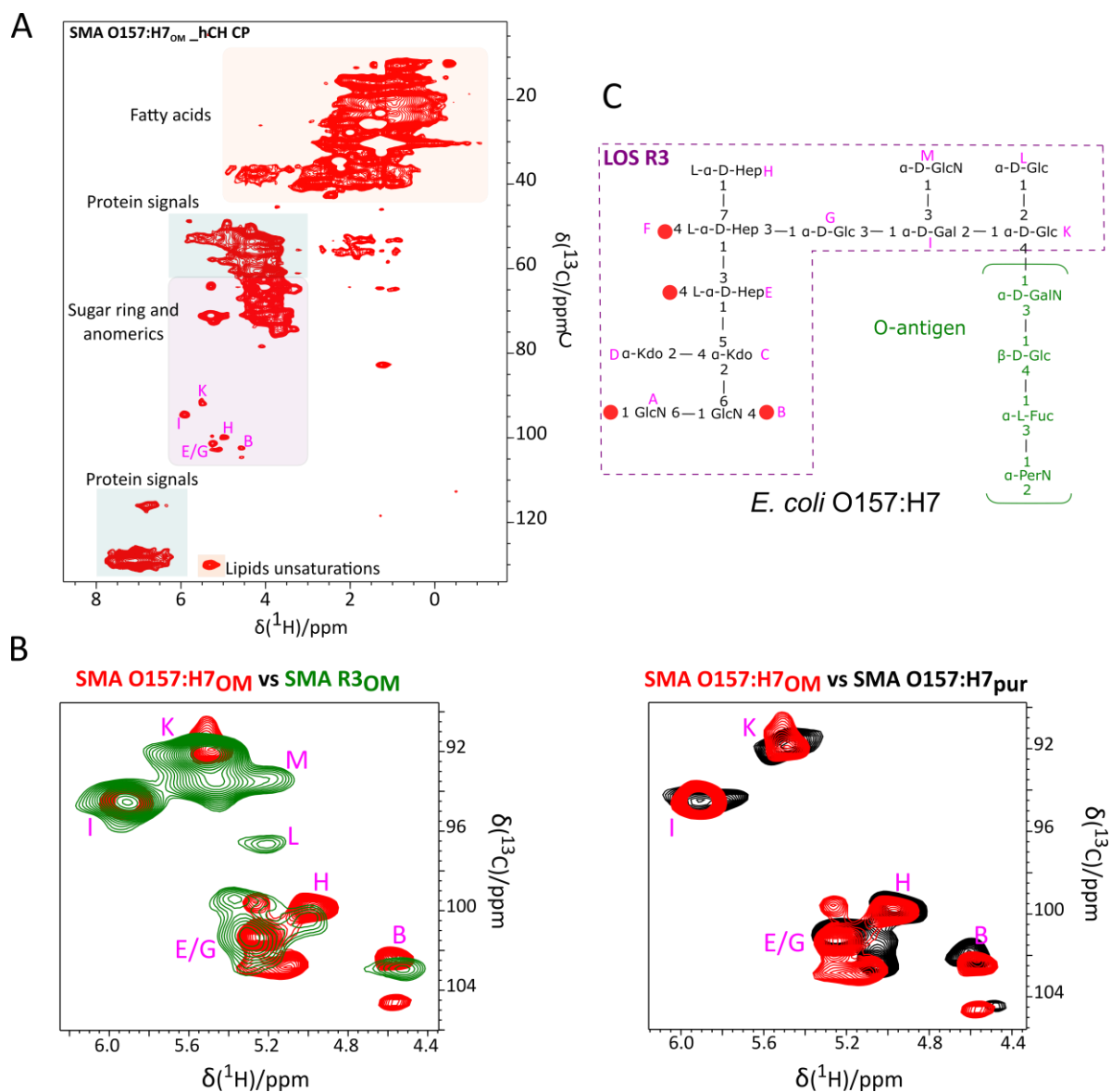


Figure 6.13: Solid state NMR characterization of LPS nanodiscs from *E. coli* O157:H7. (A) MAS INEPT spectra of SMA-O157_{pur} and SMA-O157_{OM} (B). The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were performed on a 600 MHz spectrometer at 55 kHz MAS and a sample temperature of 50°C.

The ^1H - ^{13}C CP spectra recorded on SMAO157:H7_{OM} mainly show signals from the rigid portions including protein-characteristic signals (Figure 6.14A, Annex 11). All the resonances of the O-antigen could be assigned (Annexes 9 and 10). Nevertheless, a full *de novo* assignment of *E. coli* O157:H7 core OS could not be achieved due to a lack of correlations in assignment experiments. The sequential resonance assignment of LPS signals was thus performed aided by assignments of purified LOS of *E. coli* 653 (R3 core type presenting the same core OS as *E. coli* O157:H7) reconstituted in DHPC micelles in solution. 3D TOCSY

experiments were recorded on the latter giving access to intra-residue connectivities (Annex 12). These assignments were then transferred to ssNMR spectra of O157:H7 nanodiscs of purified LPS and to OM nanodiscs since these spectra overlaid nicely (Figure 6.14B). Only subtle differences in the LPS glycan resonances are observed between OM or purified LPS nanodiscs and could be ascribed to slightly different arrangements of the LPS in native-like outer membranes (Figure 6.14B).



At this stage, most of the NMR resonances are assigned and their corresponding positions identified. The NMR data highlights the different compositions of the LPS nanodiscs, whether they originate from pure LPS or outer membranes. These chemical groups could now be used as probes for the measurement of local dynamics. These analyses will thus give insights into native membranes flexibility and highlight how the presence of O-antigens influences (or not) their dynamics. Furthermore, this data can also be used and be compared during antibiotic interactions and would highlight the membrane's behaviour upon ligand interaction.

Analysis of dynamics has already been employed to SMA R3_{OM} nanodiscs. Resonances corresponding to different regions of the outer membrane system were selected. This included a core OS anomeric resonance, lipid and proteins signals. Rotational Echo Double Resonance (REDOR) experiments ²⁵⁵ were thus recorded by Alicia Vallet (NMR platform engineer, Biomolecular NMR/IBS) and order parameters were determined. These latter reflect amplitude of motions and range from 1, for rigid motions, to 0 for high flexibility. The analysis outcome revealed a difference in the dynamics throughout the membrane (Figure 6.15A). The order parameters for lipids and OS residues range from 0.1 to 0.2, indicating a relatively significant flexibility, whereas the proteins aromatic resonances presented a higher order parameter of 0.45 (Figure 6.15B). We thus conclude that the membrane components are highly dynamic and this flexibility seems to increase as we move towards its outmost parts, possibly due to the greater solvent accessibility these parts likely present which can be correlated to their motion's amplitude.

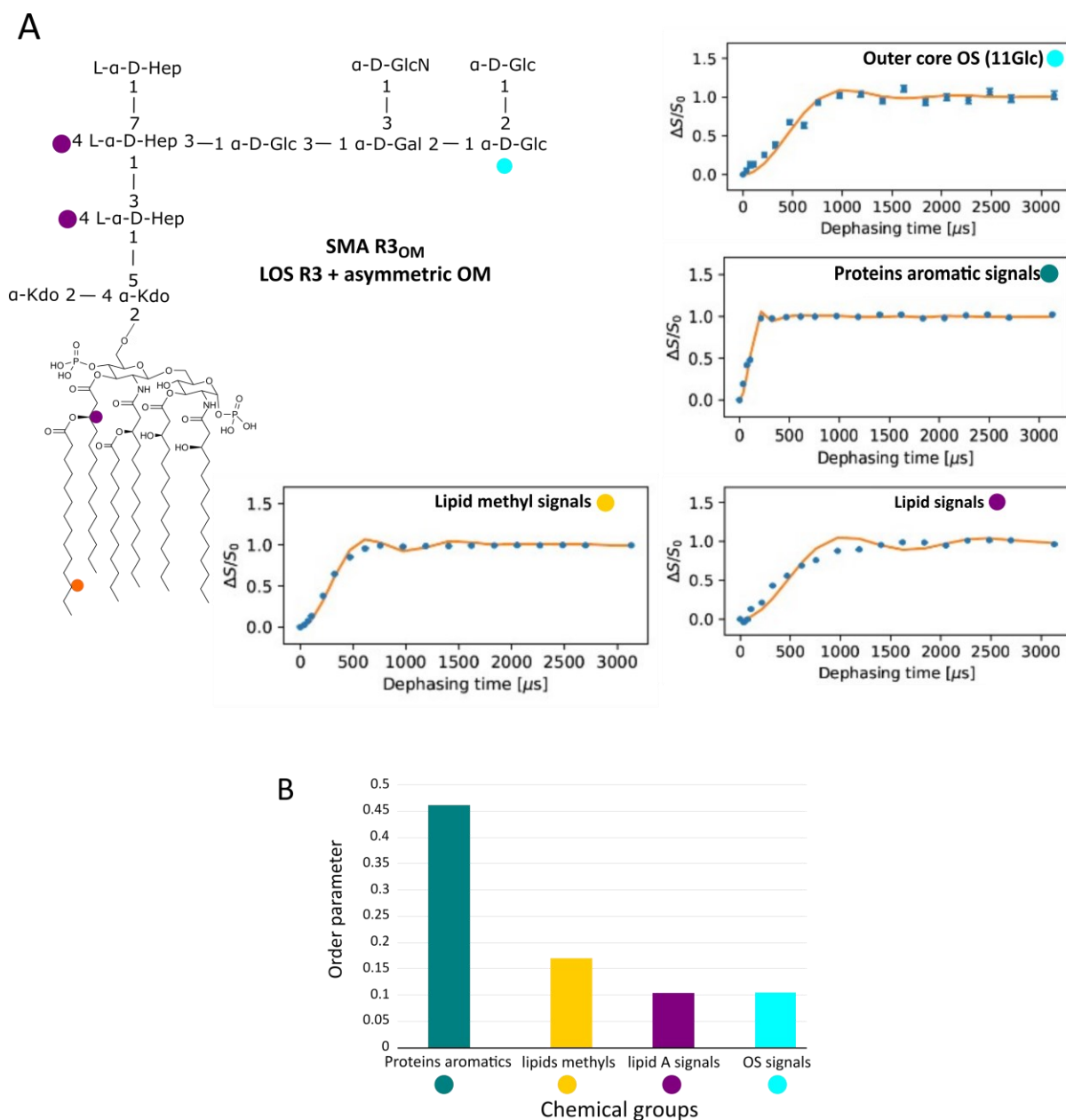


Figure 6.15: ssNMR spectroscopy dynamic analysis of SMA R3_{OM}. (A) Simulated recoupling curves $\Delta S/S_0$ as a function of the dephasing time for the different chemical groups withing SMA R3_{OM} nanodiscs. (B) Histogram plotting of the determined order parameter of each of the analysed group. The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O. NMR experiments were performed on a 600 MHz spectrometer at 55555 Hz MAS and a sample temperature of 50°C.

VI.3. SMA LPS Nanodiscs are a good model to explore interactions with antibiotics and immune receptor proteins

LPS SMA nanodiscs have been successfully prepared either from isolated LPS or directly from bacterial outer membranes. To determine the capacity of these membrane mimics to be used in interaction studies, we chose to evaluate the binding using BLI with polymyxin B (PmB), a well described polypeptide antibiotic known to interact with the Lipid A of LPS causing OM disruption and permeabilization^{256,257}. For that specific purpose, SMA-R1_{pur} was chosen for interaction studies as the simplest membrane mimic system. SMA-R1_{pur} was biotinylated using protocols already described²⁵⁸ and adapted to obtain a sparsely biotinylated sample (Figure 6.16A).

Biotinylated SMA-R1_{pur} was immobilised on BLI biosensors coated with streptavidin. The tips were then immersed in wells containing increasing concentrations of PmB. The specificity of the interaction was assessed using reference tips not loaded with SMA R1_{pur} (Annex 13A). The interaction was characterized by a very fast association step (Figure 6.16B). The binding response did not completely come back to zero upon the dissociation step. This is indicative that only a part of the PmB ligand dissociates, while some PmB remained bound. A local fit involving the analysis of each sensorgram taking into account both association and dissociation steps was performed. Association and dissociation rates were thus estimated for most of the titration points (Annexe 13B). The local fit of the data is of poor quality indicating an inadequacy with the chosen model (Annex 13B). Therefore, determination of the dissociation constant K_D was done by steady-state analysis at the end of the association step (Figure 6.16B. Annex 13C). This analysis produced a dissociation constant of $2.1 \pm 0.25 \mu\text{M}$, comparable with the affinity of PmB with different LPS molecules reported in the literature^{245,259}, and comparable to the minimal inhibitory concentration (MIC) range (1-8 mg/L)²⁶⁰.

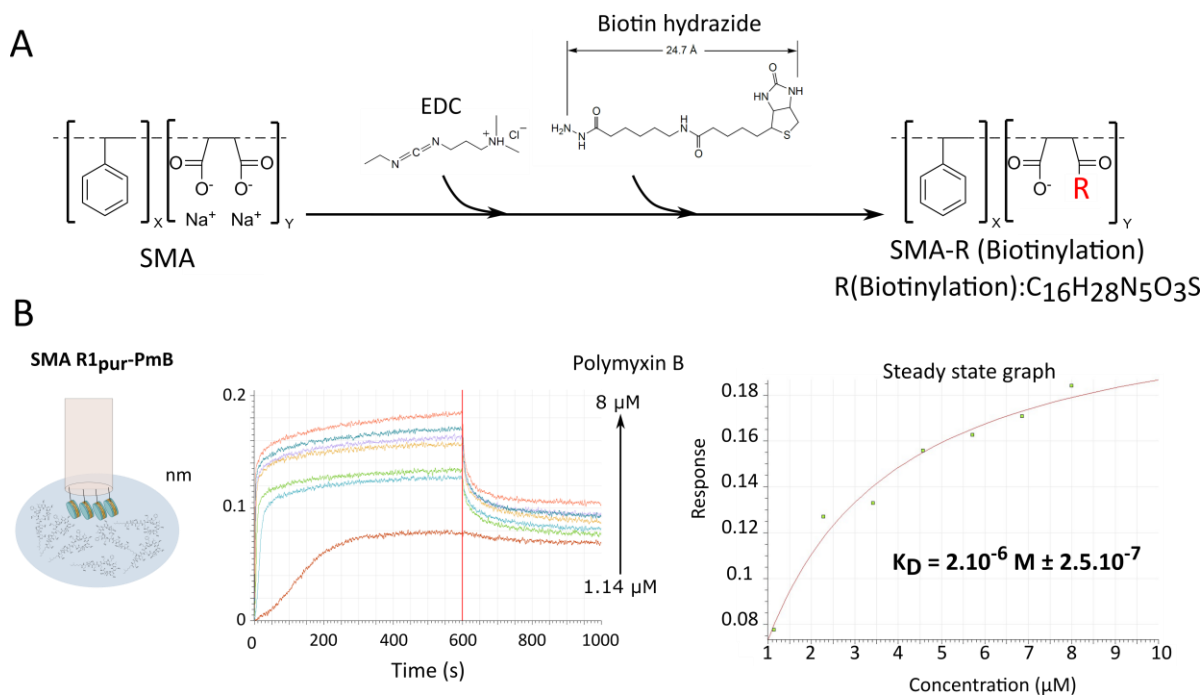


Figure 6.16: Interactions studies of immobilized SMA R1_{pur} with PmB by BLI. (A) Reaction scheme for the preparation of biotinylated nanodiscs. (B) SMA R1_{pur} tips are incubated with increasing concentrations of PmB (left). The BLI binding sensorgram at different concentrations at 25°C after subtraction of non-specific interaction contribution is shown (middle). The binding was followed for 600 s (association) then the tips were immersed in buffer for 400 s (dissociation). Steady state graph representation of binding responses as a function of polymyxin concentration with the determined dissociation constant K_D are presented (right).

Similarly, the interaction of SMA-R1_{pur} nanodiscs with MGL was next evaluated. As we have not evaluated the affinity of MGL- SMA-R1_{pur} nanodiscs so far, which gives an indication of the analyte (here MGL) concentration to use, prior optimizations for the determination of the appropriate MGL concentration for this interaction study were undertaken. Different MGL concentrations were tested, and the specificity of the interaction was assessed using reference tips on which no SMA-R1_{pur} nanodiscs have been loaded. The concentration around which MGL- SMA-R1_{pur} nanodiscs interaction were not dominated by non-specific responses and at which association and dissociation phases could be observed was determined. BLI experiments were carried out using a MGL concentrations ranging from 10 to 75 nM. At these conditions, non-specific responses of MGL ECD interaction with the naked-streptavidin

biosensor tips were around 40% (Annex 14A). Thereby, these responses together with the ones of the buffer were subtracted and the resulting sensorgram is presented at Figure 6.17.

The obtained sensorgram showed an interaction characterized by a slow association and an extremely slow dissociation (Figure 6.17). A local fit of both association and dissociation steps was performed. Association (k_a) and dissociation (k_d) rates were estimated for some points of the titration (Annex 14B). We have estimated an apparent dissociation constant K_{Dapp} value of around 26 ± 29 nM by averaging the determined dissociation constants (K_D) following the local fit and their standard deviation (Annex 14B). This interaction is similar (if not stronger) to what was obtained on BLI experiments when functionalized MGL ECD tips were immersed in wells containing increasing concentrations of LOS R1 liposomes (Annex 15). Altogether, these findings confirm once again MGL's high affinity for R1 LOS⁸². This affinity strength in the nanomolar range is a result of the avidity effect of the interaction of two multivalent partners, and is presumably facilitated by the favourable 3D arrangement of MGL allowing the presentation of all glycan binding sites at once.

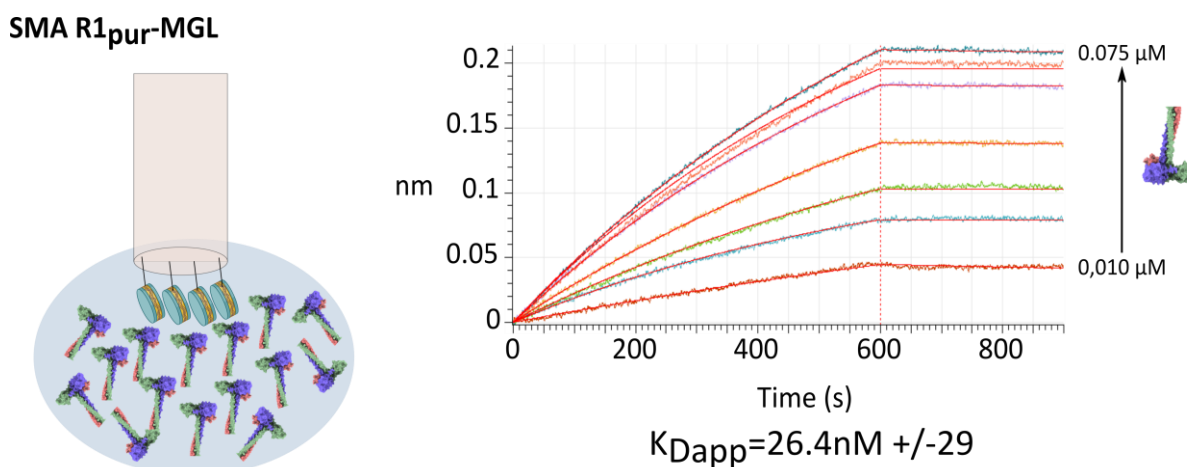


Figure 6.17: BLI interaction studies of immobilized SMA R1_{pur} with MGL ECD. Functionalized BLI biosensors are immersed in wells containing increasing concentrations of MGL-ECD with agitation for 600 s (association) then immersed in buffer for 300 s (dissociation) (left). The resulting BLI sensorgram from MGL ECD binding to SMA R1_{pur} tips at different concentrations at 25°C are presented (right) after subtraction of non-specific interaction contribution. An Apparent dissociation constant (K_{Dapp}) is calculated by averaging the determined dissociation constants (K_D) following the local fit and their standard deviation.

VI.4. Assessment of LOS nanodiscs interactions using QCM-D

In order to evaluate the behaviour of LPS nanodiscs further, we explored the interaction of LOS nanodiscs with MGL and PmB by quartz crystal microbalance with dissipation monitoring (QCM-D). QCM-D has been widely applied to assess real-time adsorption of supported lipid bilayers onto solid substrates. Measuring changes in the crystal resonance frequency (Δf) and energy dissipation (ΔD) at various overtones provides information on the mass change and the viscoelastic properties at different penetration depths of the deposited film

The fundamental frequency at which the crystal oscillates is 5 MHz, and overtones are multiples of this fundamental frequency. In this study, the frequency and energy dissipation changes were monitored at the 5th, 7th, and 9th overtones, corresponding to 25, 35, and 45 MHz, respectively. As higher overtones present lower penetration depths in the adsorbed layer, they provide information about changes occurring at or near the surface. Thereby, the 5th overtone is likely sensitive to changes happening near the crystal's surface, whereas the 7th and 9th overtones are more sensitive to changes occurring near the surface of the deposited bilayer. Analysing QCM-D data across these overtones provides valuable insights into the adsorbed bilayer characteristics, and helps discern its properties including viscoelasticity, density, and thickness.

QCM-D experiments were carried out by our collaborator Dr Samantha Micciulla at the Institut Laue-Langevin, ILL, and currently at the Interdisciplinary Laboratory of Physics, LIPhy.

The quartz surface covered with gold was prepared by deposition of SMA R1_{pur} nanodiscs. A fast adsorption kinetics was observed with a stable equilibration signal. The frequency decrease by ~ 43 Hz together with the dissipation increase to $\sim 2.10^{-6}$ in response to the attachment are consistent with the deposition of rigid bilayers (Figure 6.18A). The value of dissipation is higher compared to pure phospholipid bilayers, which can be explained by the presence of the hydrated sugar moieties of the LOS headgroups resulting in a more dissipative layer. Upon injection of MGL-ECD protein, a negative frequency shift at all overtones was observed (Δf of ~ 89 Hz) indicating an increase in mass upon MGL-ECD binding to the bilayer (Figure 6.18A). The increase in energy dissipation ($\Delta D \sim 3.10^{-6}$) on the other hand suggests that the bilayer became less rigid. The Δf and ΔD were monitored after buffer rinse ($250 < t < 400$ min) and show no significant mass loss or bilayer destabilization by MGL-ECD (Figure

6.18A). This once again confirms the strong affinity of MGL ECD for LOS nanodiscs and the high stability of the complex throughout the long rinsing steps it went through.

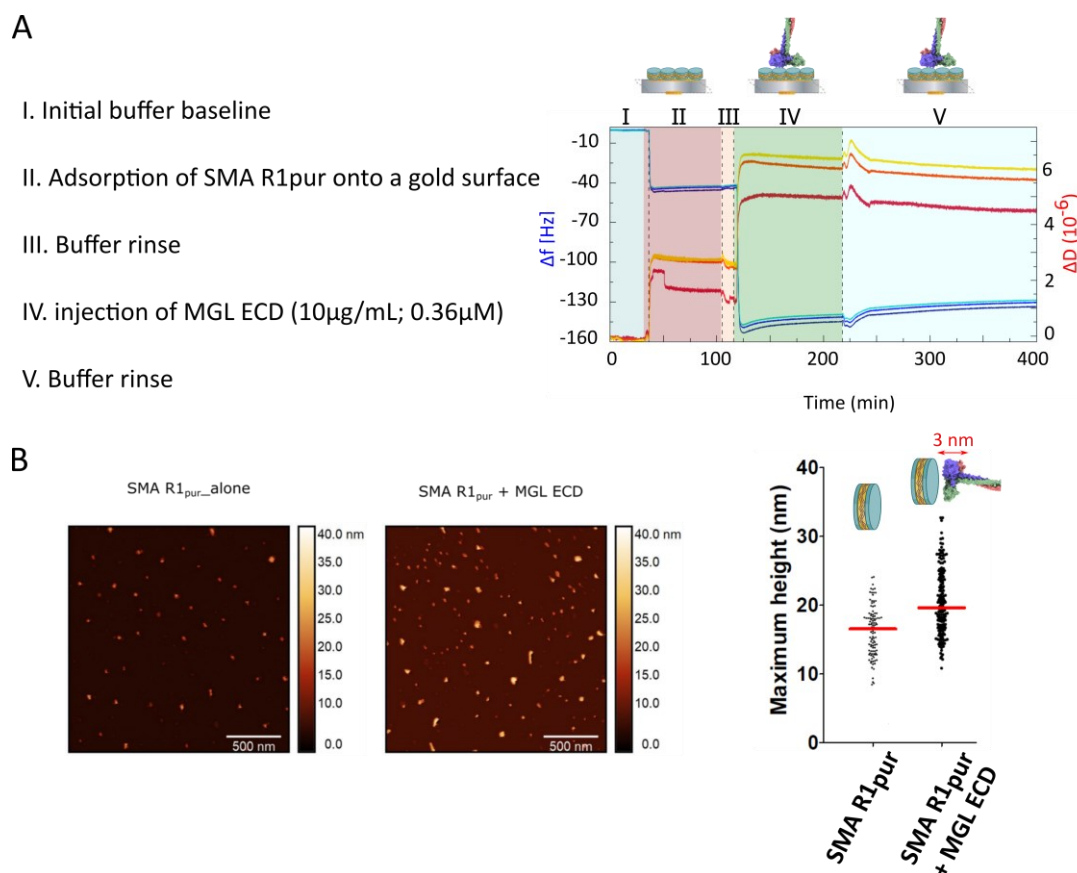


Figure 6.18: SMA R1_{pur} nanodiscs and MGL ECD interactions monitoring by QCM-D and AFM. (A) The different steps of the measurement are indicated at the left panel. Frequency (Δf) and dissipation energy (ΔD) monitoring for the 5th, 7th and 9th overtones during SMA-R1_{pur} adsorption and MGL binding. The 5th, 7th and 9th overtones measurements are represented with the red, orange, and yellow curves, respectively, on the ΔD axis and with dark blue for the 5th to light blue for the 9th overtones on the Δf axis (right). I, III and V steps are in buffer only. (B) AFM imaging of SMA R1_{pur} nanodiscs on a mica surface coated with Ni²⁺ ions in ambient air before and after incubation with MGL-ECD (Scale bars: 500 nm) (left) with their corresponding heights measurements (right).

MGL binding onto SMA R1_{pur} nanodiscs was further examined using AFM. Height measurements revealed an increase in nanodiscs thickness of about 3 nm (Figure 6.18B), presumably corresponding to the contribution of MGL-CRD to the total height. A total thickness of ~20 nm was therefore obtained but no obvious nanodiscs' aggregation or surface changes were observed (Figure 6.18B, Annex 7).

Similarly, interaction of SMA R1_{pur} nanodiscs with PmB was monitored by QCM-D. According to the literature, PmB binds onto the lipopolysaccharide layers and disrupts it when the system is heated up to 37°C²⁴⁵. Therefore, SMA R1_{pur} nanodiscs interactions with PmB was studied at 37°C and the experiment was performed using the same PmB concentration used by Paracini et al.²⁴⁵ (100 µg/mL).

To that end, bilayer preparation from SMA R1_{pur} was carried out at 37°C and provided a frequency decrease of ~20 Hz (Figure 6.19A). The frequency decrease with the observed mass stability upon rinsing is indicative of the formation of a stable bilayer. Upon PmB injection, an additional negative frequency shift of ~12 Hz was observed indicating an increase in mass following PmB binding to the bilayer (Figure 6.19A). Additionally, this binding is accompanied by a decrease in energy dissipation suggesting that the bilayer became more rigid. Furthermore, when the buffer was flushed onto the crystal for rinsing, which would remove the loosely bound molecules, a significant positive variation of Δf and a split of the overtones is observed. This is an indication of the film destabilization (Figure 6.19A).

The SMA R1_{pur}/PmB bilayer was left overnight in steady state with no liquid flow (Figure 6.19B). Both frequency shift and dissipation energy remain constant, without any variation. However, when the buffer is flown again onto the system (Figure 6.19C), additional mass loss is registered, with the Δf signal going almost back to zero. Altogether, the effects observed here upon PmB interaction including bilayer disruption and LPS mass loss correlate well with what was reported on PmB binding to LPS model membranes using neutron reflectometry²⁴⁵. This study revealed that PmB interaction caused a loss of the membrane asymmetry translated by leaflets components mixing and increased water penetration. These observations are in good agreement with the described destabilization and permeabilization of outer membranes by PmB.

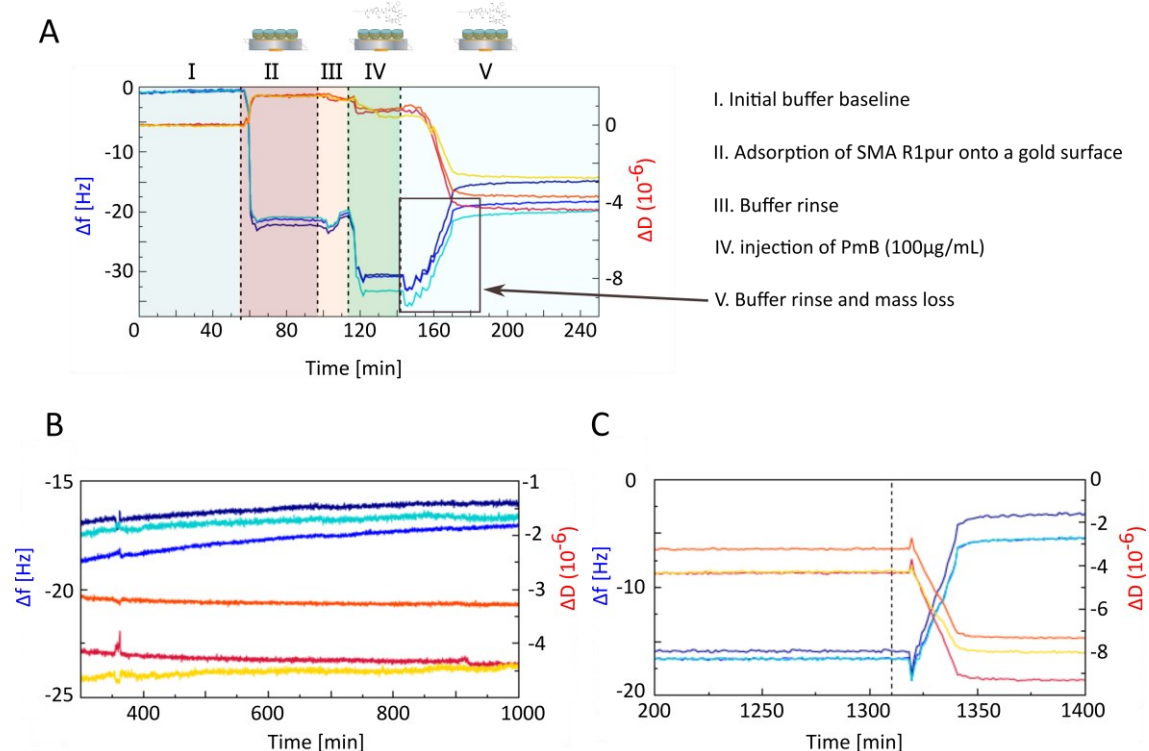


Figure 6.19: SMA R1_{pur} nanodiscs and PmB interactions monitoring by QCM-D. Frequency (Δf) and dissipation energy (ΔD) monitoring for the 5th, 7th and 9th overtones during SMA-R1_{pur} adsorption followed by PmB injection (A). The 5th, 7th and 9th overtones measurements are represented with the red, orange, and yellow curves, respectively, on the ΔD axis and with dark blue for the 5th to light blue for the 9th overtones on the Δf axis. The last buffer rinse resulted in the bilayer destabilization and mass loss. The chemical structure of PmB is indicated (top right). (B) Equilibration overnight of the SMA R1_{pur}-PmB system at 37°C in steady state. (C) Final rinsing of the SMA R1_{pur}-PmB by buffer solution at 37°C. PmB: Polymyxin B.

VI.5. Conclusion and discussion

A comprehensive understanding of LPS biology is crucial. However, challenges associated to LPS studies, including structural complexity and heterogeneity, low solubility, and contamination concerns, require multidisciplinary skills for LPS isolation, characterization, and manipulation techniques. Herein, we established a protocol for LPS nanodiscs preparation using SMA polymers. This was applied on different bacterial strains and different LOS compositions (purified and LOS from outer membranes). Their composition, shape, size distribution and thickness are thoughtfully evaluated. Structurally, most of the NMR anomeric resonances are assigned and their corresponding positions are identified. These well-resolved resonances can be used as probes for LPS structural and dynamical studies. Furthermore, SMA nanodiscs were successfully used to monitor interactions with different ligands and have been shown to be suitable for different biophysical methods.

Despite their great potential, SMA nanodiscs have their limitations. This includes the difficulties associated with the lack of SMALPs size control when compared to membrane scaffold proteins-based nanodiscs. For instance, MSP-bound nanodiscs allow fine tuning of the nanodiscs size (between 7 and 20 nm) by selecting an MSP belt with a specific length²²¹. On the other hand, SMALPs exhibit a heterogeneous size distribution²⁶². This can pose difficulties for applications that require uniform and monodisperse samples, including Cryo-EM studies for high resolution structure determination. Fortunately, these limitations are being overcome with the development of other SMA-based polymers. An ethanolamine modification of a commercially available low-molecular weight SMA polymer resulted in a derivative called SMA-EA²⁶². This latter has shown its ability to form SMALPs with improved size control by changing the lipid: polymer ratio. SMA-EA nanodiscs at a polymer:lipid greater than 2:1 resulted in isotropic small objects useful for solution NMR structural studies. Whereas, nanodiscs formed at a polymer:lipid below 2:1 resulted in larger macro-nanodiscs (>20 nm) which are used in solid state NMR experiments²⁶³. Additionally, another modification through the introduction of a non-chelating quaternary ammonium group was performed and resulted in a derivative called styrene maleimide – quaternary ammonium (SMA-QA) polymer²⁶². This latter not only improved the nanodiscs monodispersity, but also their stability in pH conditions (pH range of 2.5 to 10) and divalent cations concentrations up to 200 mM²⁶⁴.

While these limitations have been addressed, none of these polymers have been reported for solubilization of lipids or membrane protein directly from cell membranes. More importantly, none of these polymers are commercially available. Therefore, this limited

availability has hindered our ability to test these polymers on our LPS systems in order to improve the homogeneity and overall size of the SMA LPS nanodiscs for high-resolution structural studies, including Cry-EM and solution-state NMR.

Another important point is the charge of the polymer. The resulting polymer-protein interaction and their effect on the solubilization of membrane proteins have been documented²⁶². In our case, what can be pointed out is the possible electrostatic interactions that can occur between the SMA polymer (net negative charge of 2) and the interaction partner, for instance the MGL protein. Under the pH conditions our experiments were carried out, MGL is negatively charged (pI of 5), and so is the SMA polymer. In this case, one can exclude electrostatic interactions between MGL and SMA. In addition, the presence of ionic strength in the solution used can shield these charges on the polymer and potentially reduce these electrostatic effects.

These drawbacks could be avoided by using neutral polymers. Nevertheless, once again none are available. Another alternative would be the use of polymers with a reduced negative charge. That includes the use of non-SMA polymers such as amphipols (APols). Different derivatives of these polymers have been designed, including Poly(acrylic acid) (PAA)-based APols harbouring cycloalkane side chains (CyclAPols). These polymers have been validated for their efficient ability to extract membrane proteins from lipid membranes^{265,266}. With their reduced negative charge (net charge of 1) and effective solubilization capacity, they could be used as alternatives to SMAs. Two polymers are now commercially available, and one of them is already being tested for LPS nanodiscs preparations. So far, the resulting nanodiscs are of a better homogeneous size distribution compared to SMA-based nanodiscs. They will soon be further characterized, for instance by solution NMR analysis and interaction studies with MGL will be performed using other biophysical methods.

Although most of the presented interaction studies in this chapter were carried out on SMA-nanodiscs of purified LOS (SMA-R1_{pur}), outer membrane nanodiscs also hold promise to give a more realistic picture on interactions at the cell surface. Extensive efforts are devoted to a detailed characterization of this system. The presumed asymmetry of the nanodiscs, conserving the different layers of the native outer membrane is under investigation, notably using AFM imaging in solution. If we could confirm the maintenance of the membrane asymmetry in these objects, we believe that they represent an ideal model for bacterial cell surfaces, for example for interaction studies. Furthermore, interaction studies of MGL with SMA-R1_{OM} have been investigated. Given the complexity of the membrane system (mixture

of LPSs, lipids, and membrane proteins), the data analysis and interpretation require meticulous care. With implementing the right controls, we will depict LPS-MGL specific binding parameters.

So far, we have only employed this nanodiscs technology on some *E. coli* strains, but this can be extended to other public health threatening species, notably multidrug resistance strains. Polymer-based nanodiscs technology presents a valuable approach for constructing lipid bilayer models composed of molecular components derived from native cell membranes. Their chemical versatility allows straightforward modification for different biophysical methods, including SMA labelling with fluorescent probes for screening purposes. Their ability to mimic bacterial surfaces makes them valuable platforms in various research areas, including antimicrobial screening, study of host-pathogen interactions and immune responses, thus contributing to the progress in infectious disease research and therapeutic development.

VII. Structural studies of LPS-MGL interactions in membrane mimetics

Structural studies of LPS interactions with MGL lectin receptor in membrane mimetics are of paramount importance. Unravelling the molecular details of how MGL interacts with LPS provides invaluable insights into host-pathogen interactions and immune responses. Such structural studies are crucial for deciphering the specificity, affinity, and dynamics of LPS-MGL binding. Thereby, contributing to the development of novel therapeutic agents targeting bacterial infections and immune-related conditions.

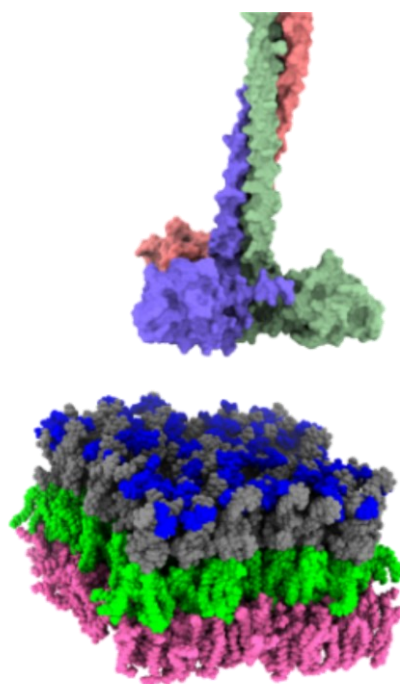


Figure 7.1: Schematic representation of MGL ECD on bacterial membranes.

This chapter focuses on the advancements made so far with the aim of constructing a model of MGL ECD arrangement on LPS membrane nanodiscs. To that end, click chemistry is on the top of the approaches we employed for Electron Microscopy (EM) studies of the system. That consists of MGL ECD labelling with gold nanoparticle to render it visible by negative staining EM. Furthermore, different MGL ECD constructions are designed. Either for further MGL glycan secondary binding site investigation, or for the design of shorter MGL ECD versions for structural studies purposes, notably for solution NMR spectroscopy.

VII.1. Electron microscopy for MGL-LPS nanodiscs structural studies

Before attempting any MGL ECD-LPS nanodiscs complex negative staining EM grids, we first managed to image both components separately. These initial studies would allow determination of the experimental conditions (protein concentrations and ratios) to be used for the complexes analysis. The grids preparation together with the images acquisition was performed with Dr Michel Thepaut.

Negative staining EM of SMA R1_{pur} nanodiscs provided nice images with objects of good monodispersity (Figure 7.2). Larger size objects are also observed (~10%) which may presumably correspond to aggregates. The applied dilution factor seems correct, even if it would be possible to work slightly with more concentrated samples. Conversely, MGL ECD imaging was not as promising as nanodiscs. Images show a very cluttered background and no isolated protein was observed. Although different grids were prepared with different dilutions, the background seemed slightly lighter but still did not allow distinct protein observation (Figure 7.2).

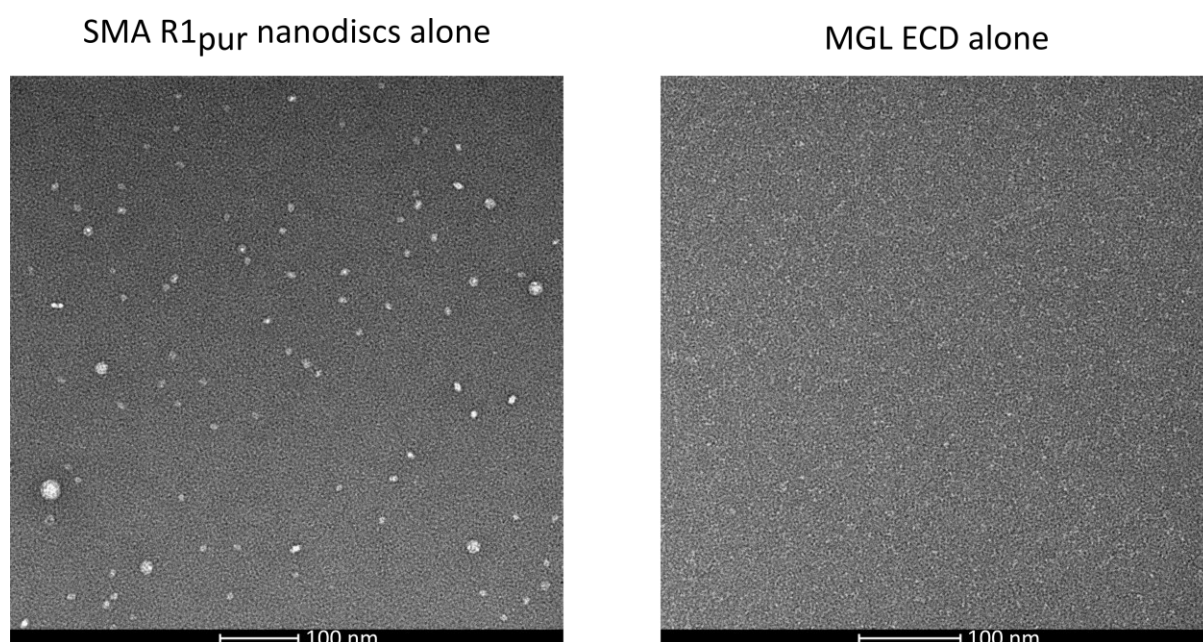


Figure 7.2: Negative staining Electron Microscopy characterization. EM images of SMA R1_{pur} nanodiscs (left) and MGL ECD (right). Negative-stain grids were prepared using the mica-carbon flotation technique and stained using 2% sodium silicotungstate (SST). Scale bars: 100nm.

Of course, when considering the large macromolecules size suitable for EM, MGL ECD with a molecular weight of 84 kDa is at the size limit for EM. However, it is worth noting that negative staining EM images of the DC-SIGN lectin revealed good quality images where isolated particles were observed²⁶⁷. Compared to DC-SIGN, MGL ECD presents a neck domain approximately twice shorter than DC-SIGN. The neck comprises ~200 amino acids for DC-SIGN that form a neck of ~20-25 nm long compared to ~100 amino acids for the one of MGL with a length of only ~17 nm. Furthermore, DC-SIGN is tetrameric while MGL ECD is only trimeric, thus the thickness of the neck is too low to result in the same contrast as for DC-SIGN. Altogether, despite the good results obtained with DC-SIGN, this explains the expected shorter size of MGL and the lower contrast for MGL and the difficulties to clearly identify it. In conclusion, the MGL is apparently too small to be clearly visible in these conditions. Hence the need to label it with gold nanoparticles to render it visible for EM.

VII.1.1. MGL ECD chemo-selective labelling approach for negative-staining EM

Chemo-selective N-terminal labelling of MGL ECD relies on sortagging reaction using depsipeptide as a substrate for the transpeptidase sortase A (SrtA) from *Staphylococcus aureus*. SrtA mediates the covalent anchoring of virulence factors to Gram-positive bacteria cell wall. It catalyses the ligation reaction of proteins carrying an LPXTG recognition motif (where X represents any amino acid) to an N-terminal oligoglycine sequence in the peptidoglycan²⁶⁸. Proteins N-terminal labelling using SrtA requires only a single N-terminal glycine on the protein. Fortunately, given the MGL ECD construct harbouring Factor Xa protease cleavage site and 3 glycine residues at the N-terminus of the protein (pET-30-StreptagII-Xa-3G-ECD MGL), a simple digestion with Factor Xa protease yields an N-terminal glycine on the protein (3G-MGL ECD) needed for the reaction.

While a variety of synthetic peptide substrates could be used for protein labelling, depsipeptide substrates stand out for their high efficiency in SrtA-mediated ligation reactions. The replacement of the peptide bond between threonine and glycine residues with an ester linkage results in a more efficient ligation reaction compared to regular peptide substrates. When using depsipeptides, the released alcohol group following SrtA-mediated ligation is a poorer nucleophile compared to the amine group released from regular peptides. This drives the equilibrium towards the ligation product and renders the reaction irreversible. Subsequently, depsipeptide substrates effectively enable irreversible full protein labelling with small amounts of substrate and SrtA, where only 1.5–3 molar equivalents of the depsipeptide

substrate and 0.1–0.2 molar equivalents of SrtA are needed. Conversely, with unmodified peptides, 20–50 molar equivalents of peptides and 2–3 equivalents of SrtA are required ²⁶⁹.

This sorting reactions have already been used in the group for the preparation of lectin-oriented SPR surfaces *via* lectins biotinylation. In our case, a dibenzocyclooctyne (DBCO) group is grafted on the depsipeptide forming DepsiDBCO and ligated to MGL ECD by SrtA. Following this step, gold nanoparticles (Azido-5 nm Nanogold) will be added to the labelled protein *via* interaction with DBCO to render lectins visible in negative-staining EM. The employed approach is summarised below in Figure 7.3.

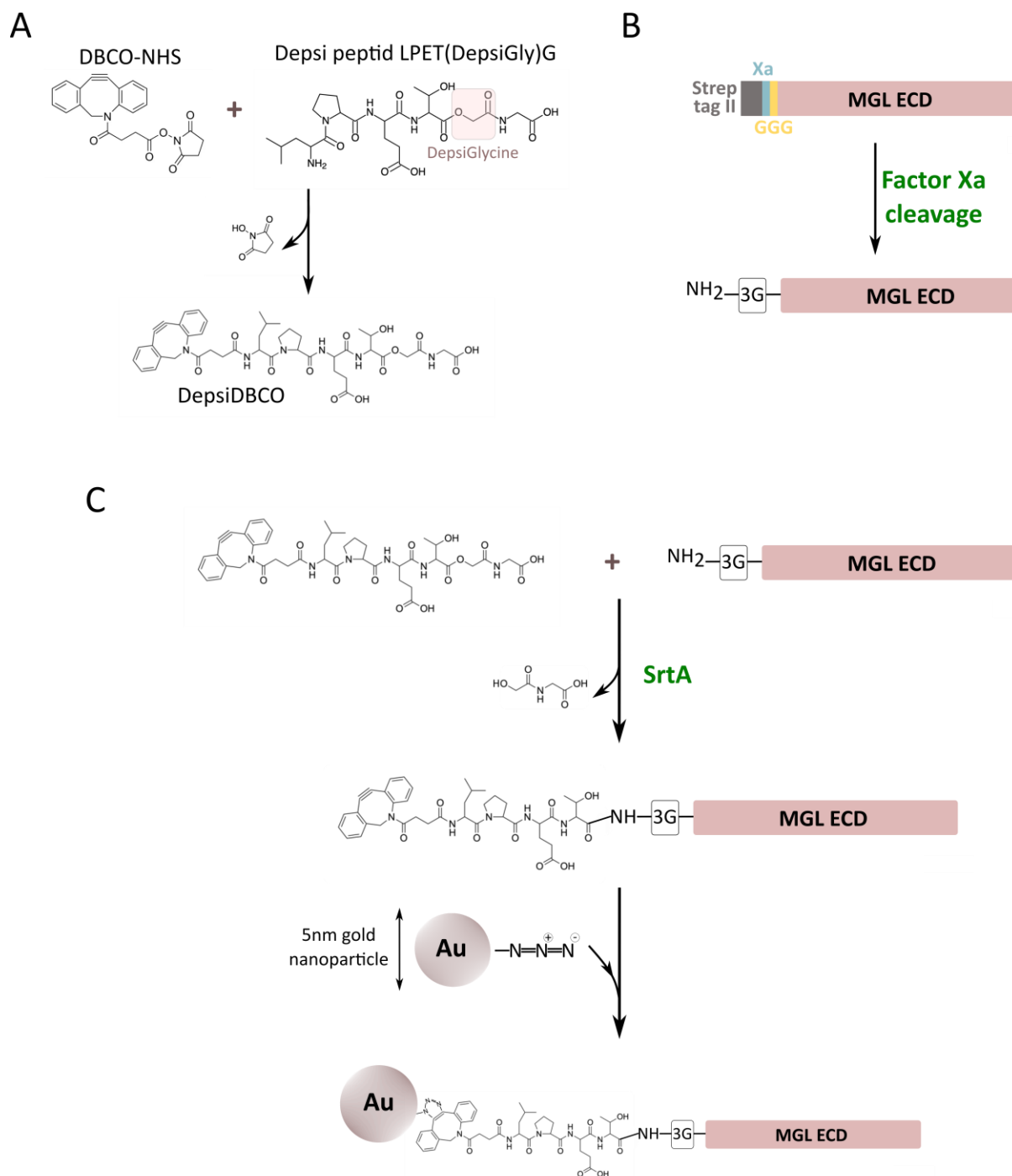


Figure 7.3: Summary of the different reactions carried out for MGL ECD chemo-selective labelling. (A) Formation of DepsiDBCO substrate following the coupling of Depsi LPTE(Depsiglycine)G peptide with DBCO-NHS. (B) MGL ECD Strep tag cleavage using factor Xa protease and generation of 3G-MGL ECD. (C) SrtA-mediated ligation of depsiDBCO substrate onto 3G-MGL ECD. Post-sorting click reaction, gold nanoparticles (Azido-5 nm Nanogold) are added to the labelled protein via interaction with DBCO. DBCO-NHS: Dibenzocyclooctyne N hydroxysuccinimidyl ester. SrtA: Sortase A.

Following the presented framework above, MGL ECD was first subjected to a Strep-tag II cleavage using Factor Xa protease followed by a SEC on a Toyopearl HW-50S. Afterwards, the sortase A ligation reaction is carried out using cleaved MGL ECD and DepsiDBCO as a substrate. This latter reaction is carried out at 37°C for 6h and followed by a purification on a GalNAc affinity chromatography followed by SEC on a Toyopearl HW-50S. The different steps are monitored by 12% SDS PAGE and MS analysis.

The MS analysis of MGL ECD protein before tag cleavage revealed the presence of a contaminant, referred to as R-3G-MGL ECD, resulting from an N-terminal degradation of the protein (Figure 7.4). This contaminating species is not able to react with Srt A given the absence of the N-terminal Glycine residue and will remain in the final sample since it cannot be separated by SEC nor GalNAc affinity.

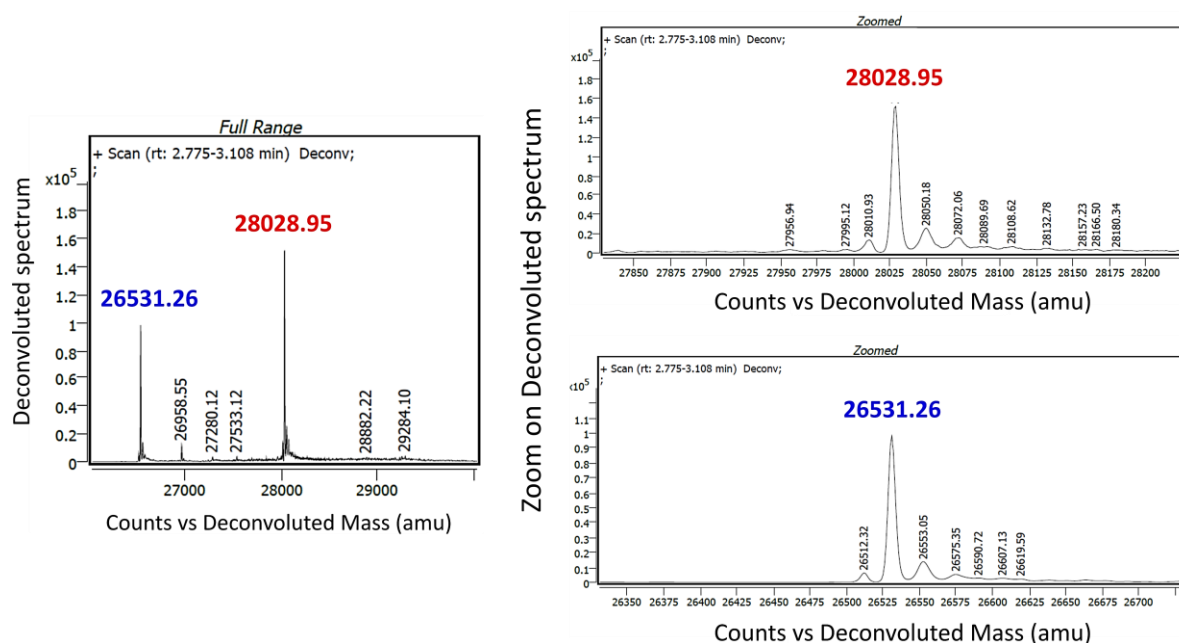


Figure 7.4: Mass spectrometry analysis of MGL ECD stock solution before Strep tag cleavage. The observed average mass of 28028.95 Da is in agreement with the oxidized protein form (StrepTag-Xa-3G-MGL ECD). The 26531.26 Da mass on the other hand corresponds to a degradation product of the protein at its N-terminal (R-3G-MGL ECD).

Post-factor Xa cleavage, a SEC was performed and the pooled fractions were also analysed by MS. Expectedly, two main populations were detected: the 3G-MGL ECD resulting from the factor Xa cleavage and the contaminant product already present in the sample (R-3G-MGL ECD) discussed above. No mass corresponding to StrepTag-Xa-3G-MGL ECD was found which is indicative that the cleavage reaction was complete. The expected and observed masses are given in table 4 and the MS spectrum in figure 7.5.

Table 4: List of the expected on observed masses following Factor Xa cleavage reaction.

	<i>Masses (Da)</i>	<i>Correspondence of each mass</i>
<i>Expected masses</i>	28028,73 Da	Non-digested protein (StrepTag-Xa-3G-MGL ECD)
	26374,71 Da	Product of the Factor Xa digestion (3G- MGL ECD)
	26531.26 Da	Contaminant species (R-3G-MGL ECD)
<i>Observed masses</i>	26375,20 Da	Product of the Factor Xa digestion (3G- MGL ECD)
	26531,56 Da	Contaminant species (R-3G-MGL ECD)

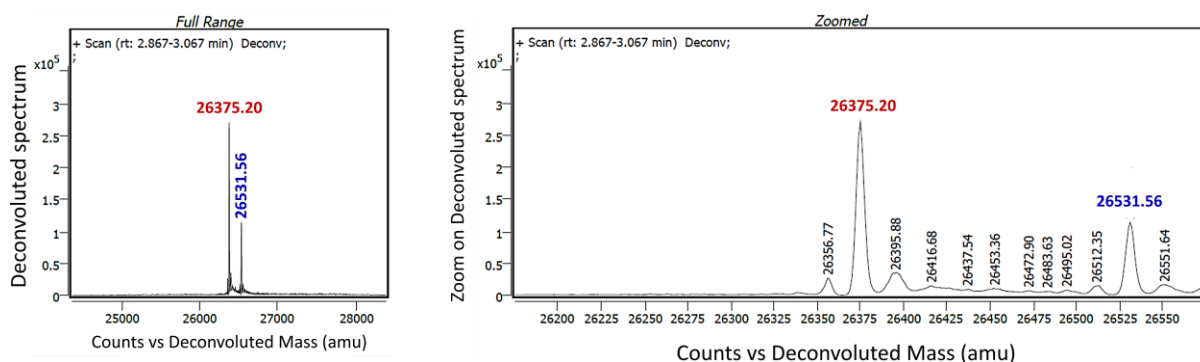


Figure 7.5: Mass spectrometry analysis of MGL ECD after factor Xa cleavage. The observed average mass of 26375.20 Da is in agreement with the cleaved form of the protein (3G-MGL ECD). Whereas, the 26531.56 Da mass corresponds to the degradation product of the protein at its N-terminal (R-3G-MGL ECD) already present in the protein's stock solution.

The last step consists of SrtA ligation reaction (see methods section) and purification of the resulting product. Whereas protein did elute from the GalNAc affinity column, no protein was eluted from the SEC purification step (Figure 7.6A). An elution step was then attempted using EDTA to recover the product (Figure 7.6B). The resulting elution profile and the corresponding analysis on 12% SDS PAGE are shown in Figure 7.6B. The migration profile was composed of several different bands. Thus, no conclusion could be drawn regarding MGL ECD labelling by SrtA at this stage.

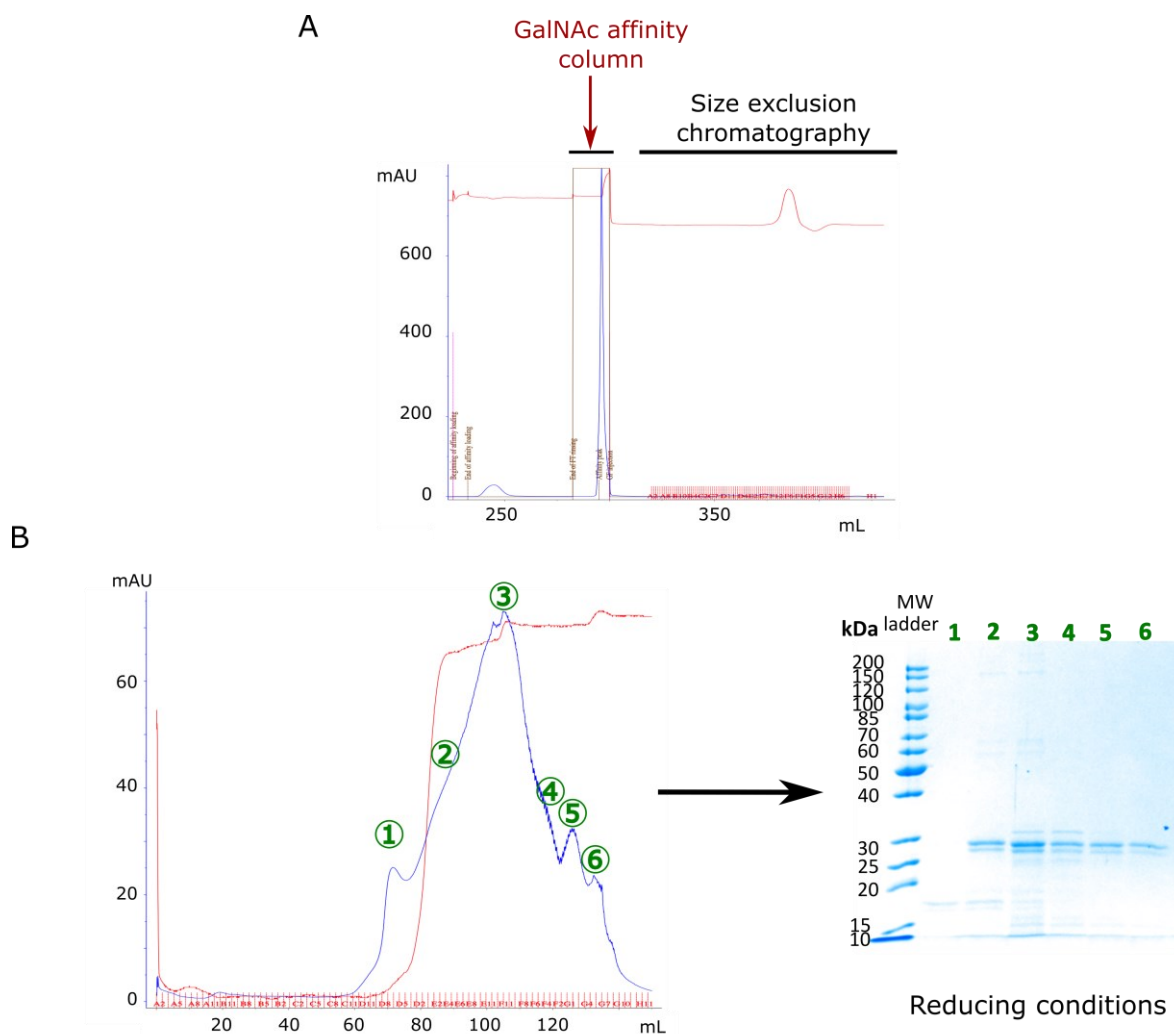


Figure 7.6: Purification of MGL ECD following SrtA reaction. (A) Full chromatogram of MGL ECD two-steps purification consisting of a first GalNAc affinity chromatography followed with a size-exclusion chromatography on a Toyopearl HW-50S. An elution peak was observed on the GalNAc affinity column whereas no peak came out after the SEC. (B) Elution of the SEC column using 10 mM EDTA with the representative 12% SDS-PAGE of the indicated fractions on the SEC chromatogram in reduced conditions.

MS analysis of some of the fractions was conducted. The significant expected and observed masses are indicated below in table 5.

Table 5: Expected and observed masses following SrtA labelling reaction of 3G-MGL ECD.

	<i>Masses (Da)</i>	<i>Correspondence of each mass</i>
<i>Expected masses</i>	28028,73 Da	Intact protein (StrepTag-Xa-3G-MGL ECD)
	26374,71 Da	Product of the Factor Xa digestion (3G- MGL ECD)
	26531,26 Da	Contaminant species (R-3G-MGL ECD)
	27102,54 Da	DBCO-3G-MGL ECD
	22797,53 Da	SrtA present in the reaction mixture
<i>Observed masses</i>	26375,18 Da	Product of the Factor Xa digestion (3G- MGL ECD)
	26531,61 Da	Contaminant species (R-3G-MGL ECD)
	24617,44 Da	Protein degradation products

According to the observed masses, no mass corresponding to DBCO-3G-MGL ECD was detected. Only masses corresponding to unreacted protein (3G-MGL ECD) and MGL's degradation products, R-3G-MGL ECD and other degradation products, were seen (Figure 7.7). At this stage, it was clear that the reaction did not work. Given the presence of the MGL 3G-ECD population, we hypothesized that either the SrtA was no longer functional or the DepsiDBCO was not good.

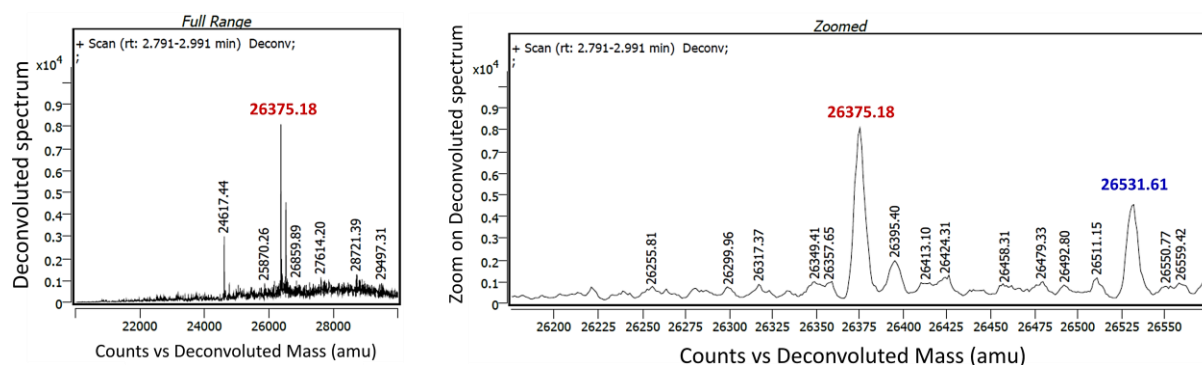


Figure 7.7: Mass spectrometry analysis of MGL ECD Following SrtA ligation reaction. The observed average mass of 26375.20 Da is in agreement with the cleaved form of the protein (3G-MGL ECD). The 26531.56 Da mass corresponds to the degradation product of the protein at its N-terminal (R-3G-MGL ECD). Additional masses were also observed presumably corresponding to more degradation products.

Further control experiments were carried out by Dr Michel Thepaut which showed that both the SrtA and the DepsiDBCO substrate were good. Nevertheless, the reaction efficiency was not promising and the yields of labelled protein were considerably low, since only small amounts of the protein could be labelled while the contaminant R-3G-MGL ECD was dominating the preparation. Therefore, the idea was to design a new MGL ECD construct without the StrepTag and factor Xa cleavage site. Once that done, a SrtA ligation reaction will be carried out on the fresh protein preparation, which hopefully will be more efficient in absence of degradation contaminants, thus opening the way to structural characterization of gold nanoparticles labelled MGL ECD in interaction with previously described LPS nanodiscs by negative staining EM.

VII.2. Perspectives for the MGL ECD project

VII.2.1. Mutants design for further MGL glycan secondary binding site investigation

Considerable efforts have been made with the aim of characterizing and dissecting the molecular recognition of MGL with *E. coli* bacterial surface glycans. So far, we could localise their binding sites either on the protein side or the ligand part using NMR spectroscopy. Although the presence of the secondary binding site was verified using mutagenesis, more investigations are still needed for a complete characterization. To this end, additional mutants of this new binding interface need to be designed and subjected to investigation.

Previous studies revealed that the main interaction between MGL and R1 OS involved glycans from R1 outer core¹⁰¹. Our collaborators in Naples (Department of Chemical Sciences, University of Naples Federico II) performed a docking study of R1 OS outer core onto the new MGL binding site identified by NMR followed by a molecular dynamic simulation (Figure 7.8). E206 and E245 were found to interact with 4- and 6-OH, and 2- and 3-OH moieties of glucose G and M, respectively. Furthermore, Q210 was involved in the interaction with 4-OH moiety of galactose L, and galactose K interacts with W250 (Figure 7.8). The residues found to be involved in ligand binding were in agreement with the observed CSPs in our previous NMR titration experiments, especially Q210 and E206 residues.

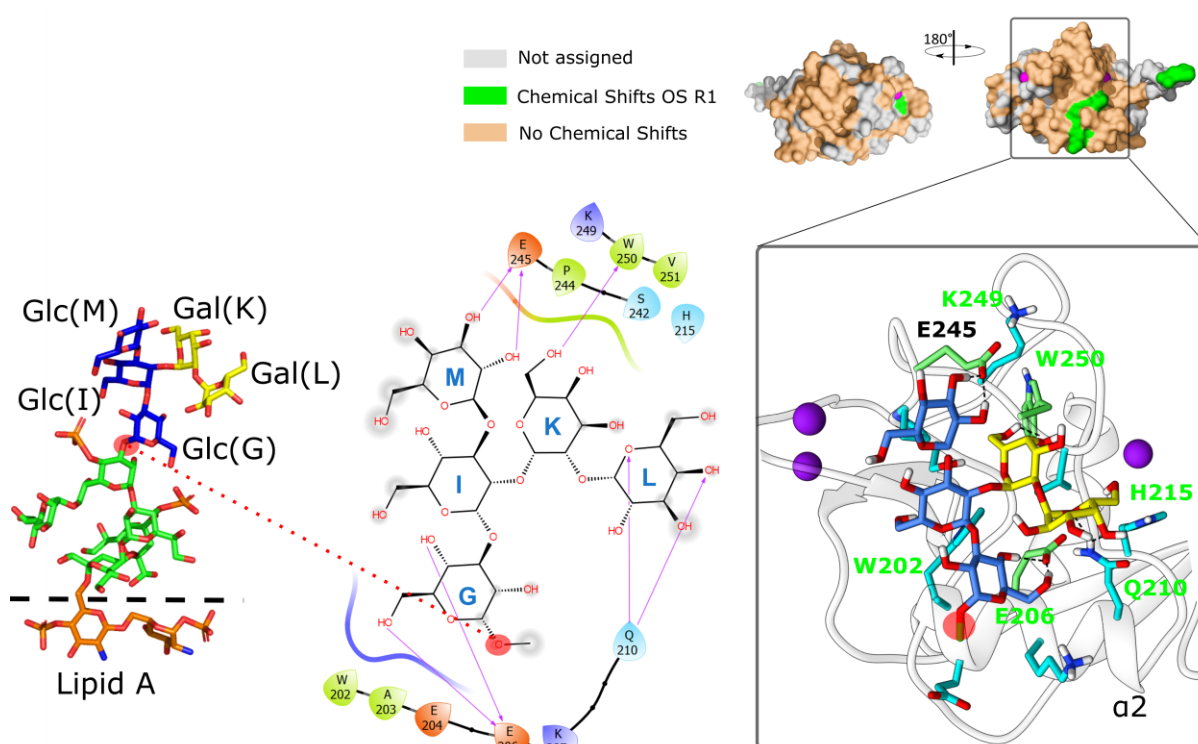


Figure 7.8: Docking and molecular dynamics of R1 OS outer core binding onto MGL new OS binding site. (Left) Stick representation of OS R1. (Middle) Schematic representation of MGL and *E. coli* R1 OS outer core interaction. Pink arrows represent hydrogen bonds. (Right) 3D representation of the MGL- R1 OS outer core complex. Residues of MGL experiencing CSPs are labelled in green. Calcium ions are represented in purple.

Following these findings, Q210 and E206 were selected for the design of a MGL ECD double mutant. Two constructs were designed: one only mutated in the secondary binding site, MGL ECD^{Q210A, E206A} referred to here as construct 1, and another mutated on both the canonical and the secondary binding sites, MGL ECD^{D269H, Q210A, E206A}, referred to here as construct 2. Interaction of these mutants with LOSs will be undertaken using different biophysical

techniques, as previously done with wild type MGL ECD for detailed characterization of the secondary binding site and its implication in LOS recognition.

VII.2.2. Design of shorter MGL ECD versions for structural studies

In addition to the characterization of MGL's secondary binding site, the design of shorter MGL ECD constructs was attempted. The long flexible coiled-coil of MGL-ECD hinders some structure-based approaches including solution NMR and crystallogenes. We wanted to obtain shorter constructs that maintain the trimeric assembly of MGL. Therefore, shorter constructions with different coiled-coil lengths have been designed.

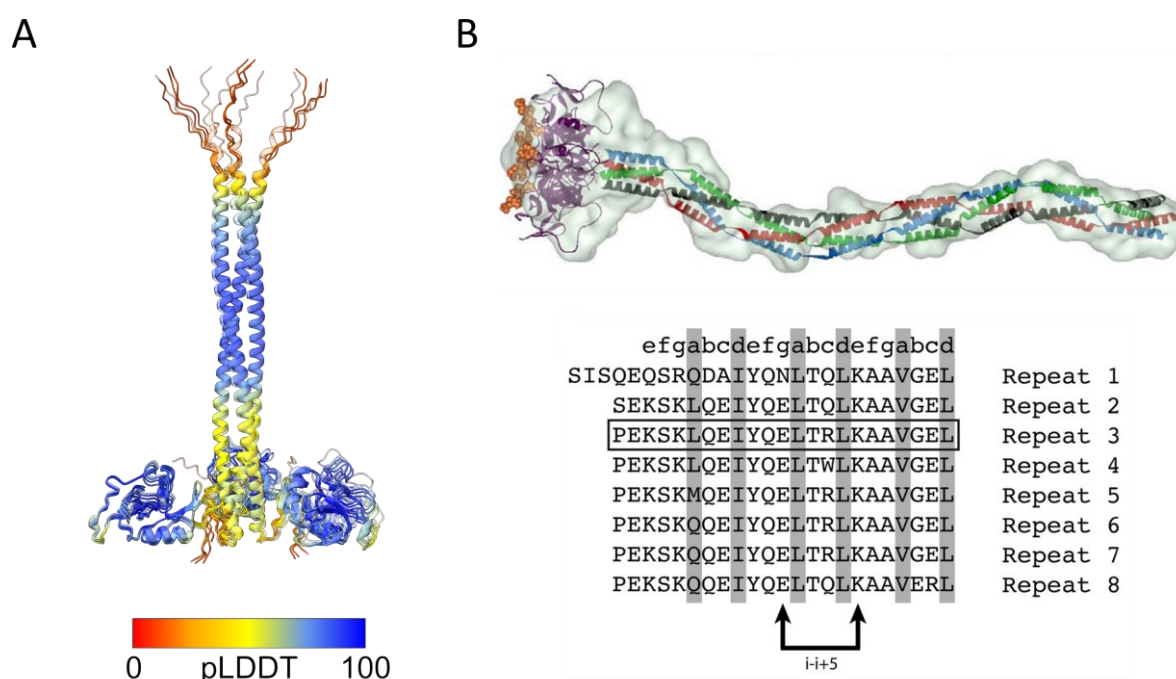


Figure 7.9: MGL ECD and DC-SIGN ECD models. (A) MGL ECD AlphaFold model with its PLDDT scores. (B) Top: DC-SIGN *ab initio* constructed model following manual positioning of CRD tetramers (PDB ID 1k9l) and the neck modules arranged in coiled-coil structures. Bottom: The repetitive segments composing DC-SIGN's neck domain with the repeat's positions (a to d) formed by hydrophobic residues are highlighted in grey. Adapted from ^{82,270}.

The MGL ECD coiled-coil arrangement differs from that of DC-SIGN (Figure 7.9), which bears repetitive segments. Unlike MGL ECD, DC-SIGN's neck region consists of seven complete repeats, each 23 amino acids long, along with one incomplete repeat. These repeats feature a unique arrangement of hydrophobic residues, spaced evenly, which potentially favours coiled-coil interaction among the helices within the tetramer's four chains (Figure 7.9B) ²⁷⁰. Consequently, in order to define the positions where the coiled-coil cuts will take

place in the context of MGL ECD, the generated AlphaFold model was taken into account (Figure 7.9A).

Two positions were chosen: in the first construct, referred to here as construct 3, the coiled-coil domain was cut before residue Arginine 149 (Figure 7.10). A second construct, referred to here as construct 4, was designed following the removal of the coiled-coil segment before residue Threonine 124 (7.10).

The corresponding DNA fragments were cloned into the pET-30 expression vector. The resulting expressed protein contains a StrepTagII, a Factor Xa protease cleavage site, and 3 glycine residues at the N-terminus of the protein (pET-30-StreptagII-Xa-3G- MGL constructs).

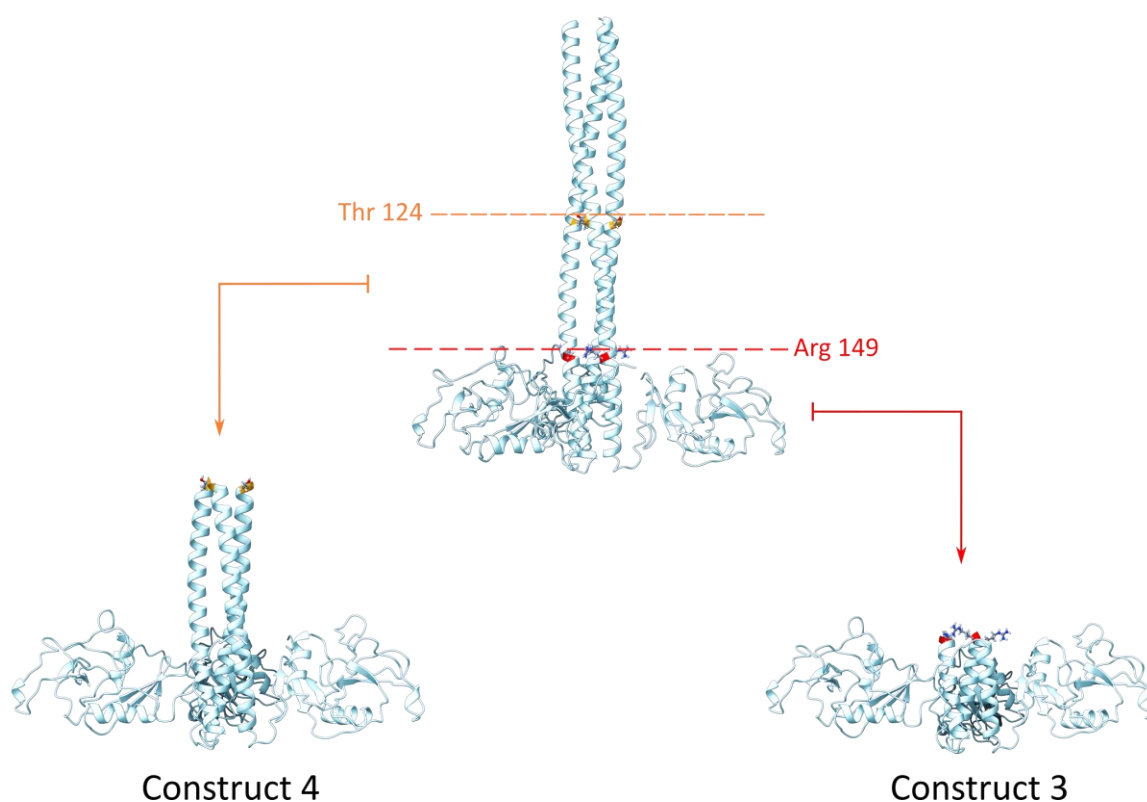


Figure 7.10: Design of shorter MGL ECD constructions of different coiled-coil lengths. Two constructs have been designed following the sequence cut at the N-terminal extremity. A first version following the removal of the structure segment downstream residue Threonine 124. An even shorter construct downstream Arginine 149 residue is also designed. The positions corresponding to the sequence cuts for both constructs are highlighted in orange and red. Both constructs are designed with a StrepTagII, a Factor Xa protease cleavage site, and 3 glycine residues at the N-terminus of the protein (StreptagII-Xa-3G-Construct 3/4).

VII.2.3. Expression trials of designed MGL ECD constructs

Regarding the expression conditions, they were all carried out similarly to the intact MGL ECD. The very first expression test was carried out in *E. coli* BL21(DE3) strains. The induced and non-induced fractions were analysed using 12% SDS PAGE. Constructs 1 and 2 are expected at 28 kDa, construct 3 at 21 kDa (186 residues), and construct 4 at 24 kDa (211 residues). Surprisingly, no intense band was observed for constructs 3 and 4 at the expected molecular weight. On the other hand, a band around 28 kDa was observed for constructs 1 and 2, but was observed on both the induced and non-induced fractions (Figure 7.11). This could be due to plasmid expression leakage, however that is unusual for the pET30 expression vector, which harbours a lac I gene to repress both T7 polymerase and protein of interest (here MGL constructs) expressions.

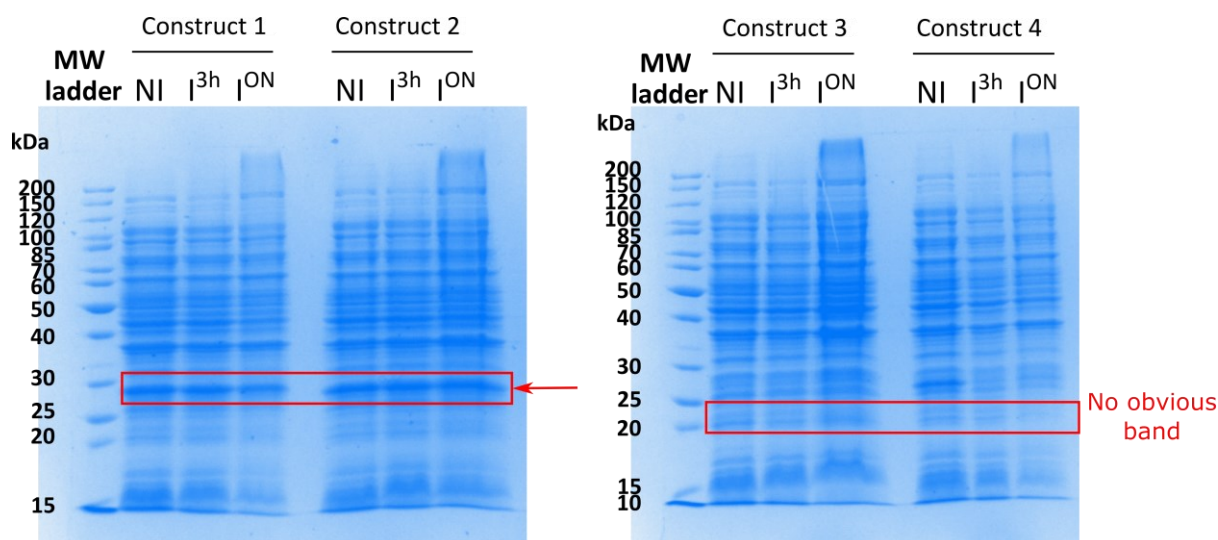


Figure 7.11: Expression test of MGL ECD constructs. Cells were cultured in LB medium. Two expression conditions were tested: induced at an OD_{600} of 2.5 with 1mM IPTG for 3h at 37°C, 180 rpm, and induction at an OD_{600} of 0.7 with 1mM IPTG Overnight at 20°C. NI: non-induced; I: induced.

Different strains as well as different culture conditions were then tested and summarized in table 6. However, no protein expression could be observed whatever the condition used. Even the “plasmid expression leakage” that was observed previously for constructs 1 and 2 was no longer observed.

Table 6: Summary of the different strains, culture conditions used for MGL constructs expression tests.

Bacterial strains	Culture media	Temperatures	Optic density 600nm of induction
BL21(DE3)	LB	Expression induction for 3h at 37°C	0.7
BL21(DE3) Star			
BL21(DE3) pLysS	TB	Expression induction overnight at 20°C	2.5
Rosetta (DE3)	M9		
Rosetta (DE3) pLysS			

At that point, we suspected the plasmid from which the constructions were designed (pET-30-StreptagII-Xa-3G-MGL ECD). We thus attempted to design the same constructions from a plasmid we already tested in the lab and confirmed its protein expression. Again, no protein expression was observed whatever the strain or culture condition we employed (Figure 7.12). Bacterial growth under kanamycin antibiotic was as usual, and the sequencing provided with the designed constructs showed the correct DNA sequence for the insert. This lack of protein expression can be explained either by instability of the resulting protein (degradation), or by inefficient transcription or translation, due to codon usage and/or secondary structures of the RNA transcript.

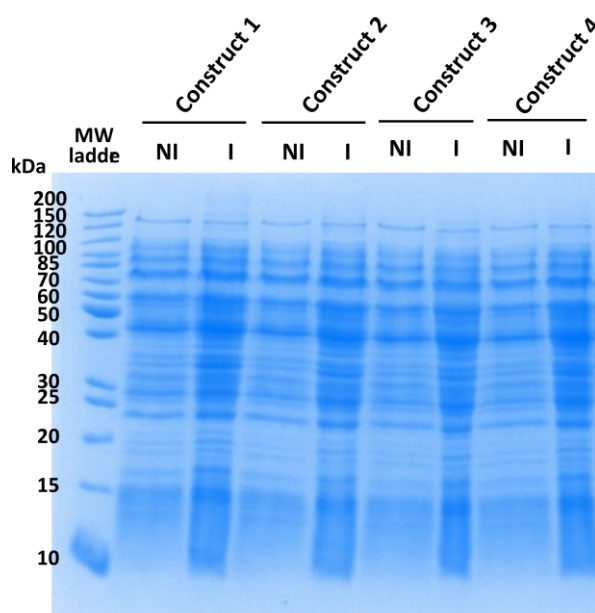


Figure 7.12: Expression test of MGL ECD constructs. Cells were cultured in LB medium. Protein expression was induced at an OD_{600} of 0.7 with 1mM IPTG for 3h at 37°C, 180 rpm. NI: non-induced; I: induced.

VII.3. Conclusion

At this point of the project, we have focused on adapting an integrative structural approach for the study of LPS interaction with MGL receptor in membrane mimetics. Combining NMR and other biophysical methods provided data on MGL-LOS interactions in the different system contexts studied so far will provide necessary data that highlight important determinants of the interactions. In the future, they will be used to build a model of the ECD arrangement on LPS surfaces. This chapter mainly focused on the advancements made so far for the future investigation of MGL ECD-LPS nanodiscs complexes by means of EM studies, and other structural biology methods, notably NMR or even X-ray crystallography.

Click chemistry is the key approach we relied on for EM studies. While the initial labelling attempts of MGL with DepsiDBCO have not yielded satisfactory results, it did confirm the functionality of required substrates and enzymes for the reaction. These findings offer valuable insights into potential challenges, and will be taken into account in the refinement of the protocols and troubleshooting methodologies for our upcoming experiments. The design of MGL ECD lacking the Streptag and use of a fresh protein preparation would probably contribute to limit the degradation contaminant, hence increase the yield of labelled protein necessary to launch MGL grafting with gold nanoparticles. This latter step may require considerable fine tuning of experimental conditions, including the reactant concentrations, reaction time, and temperature to maximize the yield of the protein-gold nanoparticle conjugate with the desired protein orientation and unattenuated functionality. This also applies to MGL ECD new construction design where further testing of experimental conditions and modification of gene sequences is necessary to improve protein expression.

General conclusion and future perspectives

VIII. Conclusion

As with other C-type lectin receptors, MGL plays an important role in cell adhesion and immune responses determined only by carbohydrates recognition. Cell surfaces are covered by dense and diversified glycan structures, including bacterial cell walls which are decorated by LPSs in Gram-negative bacteria. The immune system must distinguish pathogenic from commensal bacteria and efficiently eliminate the former. Although some CLR recognition patterns of defined glycan motifs are reported, a lot still needs to be understood regarding the molecular mechanisms behind this recognition and the immune response generated. These questions arise partly due to the intricate nature of potential bacterial ligands' structures, together with the absence of detailed structural data on lectin-bacterial glycan complexes.

In this PhD research project, we applied an integrative approach to the study of such a CLR-glycan interaction. From the lectin side, we focused on the ExtraCellular Domain (ECD) of MGL and complemented the data using the Carbohydrate Recognition Domain (CRD) of the same protein. From the bacterial glycan ligand counterpart, we employed different glycan ligands with increasing complexity ranging from oligosaccharides isolated from deacylated LPS to intact LPS molecules reconstituted in membrane mimetic systems. This approach enabled us to elucidate how both MGL constructs engaged with the different studied ligands and highlighted binding specificity and affinity of these protein domains towards the studied ligands in their different molecular contexts.

Initially, we demonstrated the ability of MGL to bind *E. coli* core OS by integrated approaches spanning from the cellular to the atomic level. We first expressed and purified MGL ECD as a perdeuterated version. To our knowledge, this represents the first instance of a full-length C-type lectin ECD investigated using NMR spectroscopy. By employing chemical shift perturbation, upon titration of delipidated LOSs, we identified a new carbohydrate binding site located on a surface opposite to the canonical QPD motif. This particular site is responsible for binding *E. coli* cores R1 and R3, and was validated through mutagenesis experiments of the canonical binding site. Altogether, these findings pointed out that the canonical binding site plays a contributory role in the interaction, but it does not serve as its primary determinant⁸². A property already reported for some CLRs including Langerin²³⁵, Mincle²³⁶, and DC-SIGN²³⁷, however, to our knowledge, a first of its kind for MGL. Although this secondary binding

site is believed to play a significant role in expanding the recognized ligand spectrum as discussed in chapter 5, its glycan preferences and functional implications are yet to be determined.

We then focused on the investigation of the arrangement of trimeric MGL ECD. For this, a model was constructed by combining data obtained from Small-Angle X-ray Scattering (SAXS) experiments with molecular structures predicted by AlphaFold. This model revealed that the CRDs are positioned perpendicular to the coiled-coil domain, thus facilitating simultaneous access to both the canonical QPD- and OS-binding sites for glycans located at a bacterial surface⁸². With six binding sites present, this arrangement significantly enhances the affinity of the interaction. This 3D arrangement distinguishes MGL from other described lectins, including Langerin²⁷¹. Structural analysis of a truncated version of trimeric Langerin revealed the presence of multiple interactions between the coiled-coil and the CRD domains. This results in a rigid structure of the trimeric protein with the three CRDs maintained in fixed positions, and their carbohydrate-binding sites located 42Å apart from one another²⁷¹. This arrangement can likely impose constraints on the type of ligands that can be accommodated by Langerin trimer, where only ligands with a given spacing can be bound²⁷¹. Conversely, MGL CRDs are located 80 Å apart with no extensive interactions with the coiled-coil domain⁸². This characteristic allows the enhancement of MGL plasticity to bind avidly to glycans with a broader spectrum of spacing. A unique 3D arrangement compared with other C-type lectins that we believe is correlated with the unusual positioning of the glycan binding sites, where the secondary binding site is on the opposite surface to the canonical binding site.

We then investigated MGL-*E. coli* glycans interactions in *ex-vivo*-like setups on whole *E. coli* cells. We could show that the presence of the O-antigen entity significantly influences the binding. More importantly, these experiments on live cells demonstrated the influence of the ligand arrangement and presentation on the binding strength. In other words, this latter exhibits a significant increase, ranging from millimolar (mM) when utilizing core oligosaccharide (OS) derivatives, to micromolar (µM) when using intact LOS molecules, and further to nanomolar (nM) levels when presented on the cell surface. This finding underscores the critical role of ligand presentation, emphasizing the necessity for innovative systems that accurately mimic the cell surface environment for LPS studies.

From that perspective, we developed a protocol for preparing LPS nanodiscs using SMA polymers. This approach was based on previous studies demonstrating the ability of these polymers to separate patches of native bilayers from biological membranes. This protocol was applied on LOS from various strains, using either purified LOS or direct LOS extraction from

outer membranes. The resulting nanodiscs were successfully used to investigate interactions with various ligands, including MGL lectin, and proved suitable for a range of biophysical methods. More importantly, these objects allowed the measurement of strong binding affinities in the nanomolar range, demonstrating them to be a much better model for the bacterial cell surface compared to other LPS model systems tested so far. We believe that this is attributed to the native-like LPS arrangement within the nanodiscs which ensures tight LPS packing and geometry closer to their native arrangement on the cell surface. The integration of outer membrane nanodiscs, which are more representative of the cell surface, will provide an even more realistic picture of processes occurring at the bacterial cell surface. Given their ability to mimic bacterial surfaces, they could be employed in studies ranging from antimicrobial screening to exploring host-pathogen interactions and immune responses, including structural studies. We are still working on leveraging this technology to build a detailed model depicting the arrangement of MGL ECD on LPS membrane nanodiscs. This would provide valuable insights into the molecular mechanisms underlying these interactions.

In summary, the exploration of MGL in particular, and lectins in general, binding sites and functional implications, coupled with the advancements in LPS nanodiscs technology, holds significant promise for understanding the mechanism underlying pathogen recognition by the host immune system. The on-going improvement of this technology will undoubtedly enhance our ability to study complex host-pathogen interactions valuable for potential therapeutic agents' development.

IX. Future perspectives

IX.1. SAXS/SANS for MGL-LPS nanodisc model construction

Determining the 3D structure of macromolecules is a prerequisite for understanding the molecular basis of cellular processes, their potential involvement in diseases, and also for therapeutic drug development purposes. The remarkable progress in major structural biology techniques has generated essential structural knowledge which contributed much in many areas including drug design ²⁷². Obviously, each technique holds pros as cons and each sample requires specific treatment to render it suitable to the method of choice. LPS-bound MGL is a complex system. Studies at atomic resolution would be challenging due to the complex stability, sample homogeneity, and molecular flexibility. Nevertheless, even if high resolution 3D structure determination of MGL-LPS nanodiscs complex seems impossible, construction of a model highlighting the complex component binding poses can be envisaged. Valuable information can be obtained by small-angle scattering-based techniques, notably SAXS together with Small-Angle Neutron Scattering (SANS). These methods have proven powerful and complementary in providing low resolution structural information of complex biomolecular systems in solution.

Sample preparation ensuring the homogeneity and stability of MGL-LPS complex together with subsequent analysis of the data including background subtraction will be performed. Deuteration will play an essential role in the refinement of MGL-LPS complex models in SANS. The advantage of this method is that the contrast of individual components of the complex can be modulated by gradual replacement of hydrogens by deuterium. Selective deuteration of specific parts of the complex at non-exchangeable hydrogen positions will provide enhanced contrast and resolution ²⁷³. For instance, deuteration of MGL or LPS can enhance the contrast between the MGL and LPS molecules. Consequently, combining both SAXS and SANS data outcomes will allow determination of the respective orientations and positions of the complex's components.

Although deuterated LPS molecules are not commercially available, they can be extracted from the producing bacterial strain grown in a deuterated medium with deuterated carbon sources ²⁷⁴. Additionally, if LPS could be deuterated in specific chemical groups (e.i outer core OS groups), insights into the binding sites could also be gained and contribute to the refinement of the obtained data on perdeuterated molecules. Le Brun *et al.* ²⁷⁵ have presented a procedure to partially deuterate LPS molecules using modified C1 (ModC1)

minimal medium. This was performed using *E. coli* J5 strain (ATCC 43745), which is a rough mutant of *E. coli* O111:B4. This mutant has an enzymatic defect and lacks uridine diphosphate galactose-4-epimerase, thus is unable to synthesize galactose²⁷⁶. Consequently, when grown in absence of galactose, it produces incomplete LPS lacking galactose residues. In contrast, when grown in galactose-supplemented medium, complete LPSs identical to the wild type strain are produced. These strains characteristics could be exploited to achieve controlled deuteration of specific LPS glycan groups. From that perspective, selective LPS labelling can be achieved using recombinant versions of our bacterial strains of interest exhibiting enzymatic defects in the metabolism of LPS sugar components. Thereby, the different specific labelling could be achieved by introducing deuterated substrates that would compensate these defects. Nevertheless, this procedure would require engineering of strains of interest and extensive condition optimizations. To our knowledge, no data have been reported regarding the use of such mutants for specific deuteration purposes.

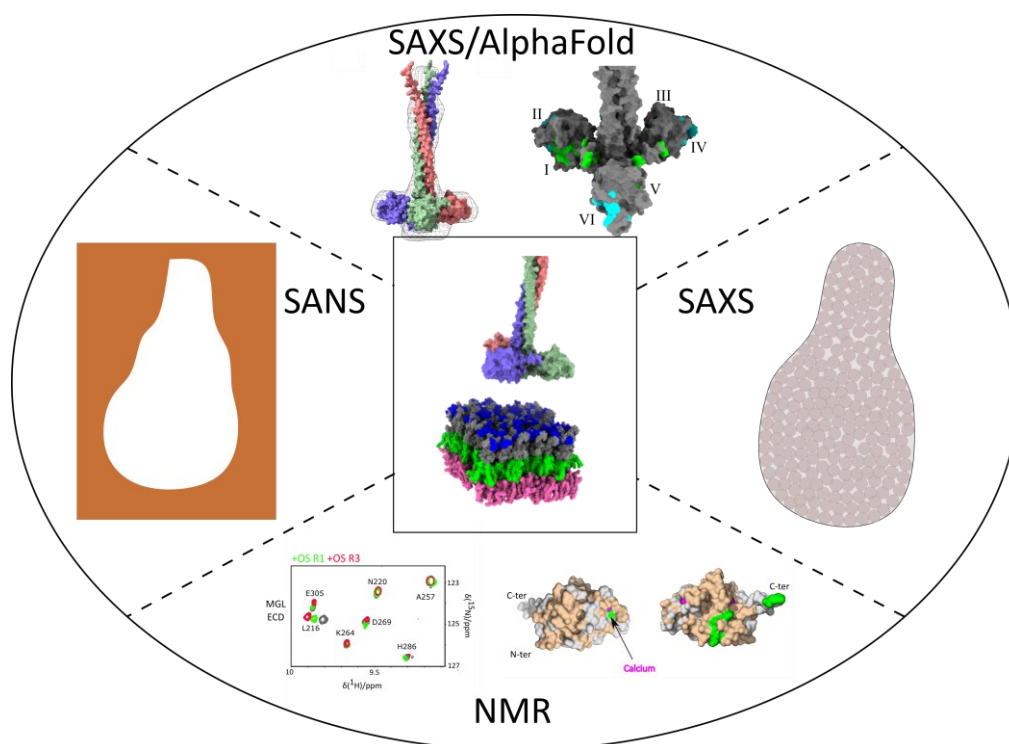


Figure 8.1: MGL-LPS nanodiscs model construction. Outline of the multidisciplinary approach for MGL-LPS nanodiscs complex model construction in solution by combining NMR and small-angle scattering.

Later on, constructed models could be refined using additional experimental data (e.g. data from ongoing EM studies, NMR CSPs, MGL ECD SAXS/AlphaFold model) (Figure 8.1) together with docking and molecular dynamic simulations. Negative-stain microscopy imaging of MGL-SMA LPS nanodiscs will be used to evaluate the homogeneity and integrity of the sample. Images obtained from a detailed negative staining study will be submitted to a 2D classification which would provide insights into the overall complex organization. Furthermore, high resolution structure determination of MGL in its trimeric form is also envisaged. More efforts will be devoted for the production of shorter coiled-coil versions of MGL-ECD for X-ray crystallography studies purposes, and also for solution NMR spectroscopy analysis in complex with LPS nanodiscs. Moreover, Cryo-EM may also be employed using antibodies directed against MGL CRDs. By combining the outcomes of each work package, the puzzle can be reconstituted and this will ultimately help improve the modelling of the molecular arrangement and interaction positions within the complex. The combination of these different integrative approaches has already been used and reported for structure elucidation spanning from small to large proteins, and to protein complexes in solution^{277–280}.

IX.2. Extension of nanodiscs technology to other bacterial strains with potential structural characterization approaches

While our current work has been mainly focused on Gram-negative strains, there exists promising potential to expend this SMA-LPS nanodiscs to a broader spectrum of bacterial species. Particularly those of major public health concern including strains involved in bloodstream infections in hospitals, such as *Klebsiella pneumoniae* and *Acinetobacter spp*, or even strains responsible for common bacterial infections like *Neisseria gonorrhoea* where 60% of the isolates have shown high ciprofloxacin resistance²⁸¹. This system could also be expanded to commensal species, including those found within human gut microbiota.

The ease of SMA-LPS preparation together with SMA's chemical versatility offers a valuable opportunity to render these objects suitable for different biophysical methods aiming to study structure and dynamics of intact LPS molecules. Such objects can be used as a platform to study molecular interactions of potential immunity receptors with LPS originating from various bacterial strains as they could serve to screen the efficiency of antimicrobial substances.

A variety of approaches can be envisaged. As SMA nanodiscs were proved promising for surface-sensitive techniques such as QCM-D, they could be extended to neutron and X-ray reflectometry. Both techniques have been extensively employed in the study of LPS containing

membrane models^{22,282,283}. These techniques would offer valuable information, especially for the LPS nanodiscs part, on which obviously more structural information would be appreciated. X-ray reflectometry can provide information on the thickness of LPS nanodiscs and will allow further structural characterization of LPS containing nanodiscs, including surface density and roughness through the generated electron density profiles. On the other hand, neutron reflectometry for its part can probe the detailed structure, orientation, and penetration depths of LPSs within the nanodiscs. Even more, dynamic and structural events occurring upon interaction could also be captured. By taking advantage of neutrons sensitivity to isotopes, deuteration of specific chemical groups within the nanodiscs allows to determine their thickness and position.

Subsequent combination of these approaches' outcomes with the ones generated by methods previously mentioned including SAXS, SANS, EM, and NMR would definitely provide essential structural information and would fill a considerable part of the puzzle regarding LPS nanodiscs molecular architecture and structural organization.

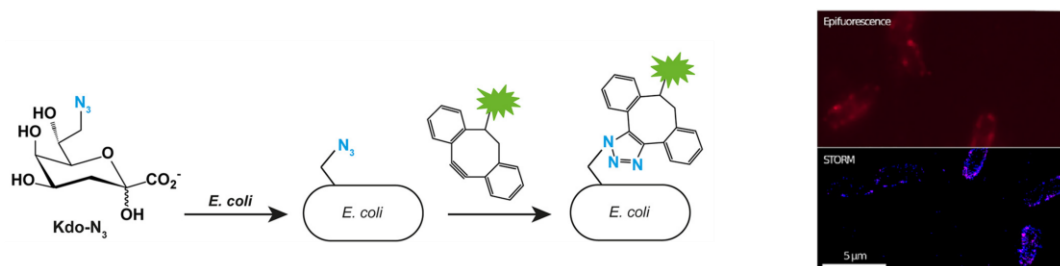
IX.3. Super-resolution microscopy and AFM imaging

Super-resolution microscopy and AFM imaging can offer valuable insights into the organization, and mechanical properties of MGL-LPS interactions. Super-resolution microscopy techniques, notably Single-Molecule Localization Microscopy (SMLM), including STORM, can provide insights into the spatial organization of MGL on LPS covered bacterial surfaces, their clustering and co-localization on the cell wall. In addition to the protein labelling we have used so far, other labelling schemes could be explored. LPS molecules can be metabolically labelled through incorporation of a sugar precursor, Kdo-N₃, in its inner core OS^{284,285}, followed by a coupling with a DBCO-conjugated fluorophore commercially available (Figure 8.2A). Alternatively, other azido-sugar precursors could also be considered for core OS and O-antigen labelling²⁸⁶.

Following up the previous investigations on MGL recognition of O-antigen carrying strains, this method can be considered to see whether MGL interaction with the O113:H21 strain takes place at the O-antigen portion level or if this moiety just restricts MGL access to the core OS. In the case of the former possibility, super resolution microscopy might be utilized and would reveal signals far from the membrane corresponding to MGL binding to the O-antigen forming long extension from the membrane. Whereas, if only limited signal would be

observed on the cell surface, it would confirm that the O-antigen portion indeed limits MGL access to core OS recognized residues.

A



B

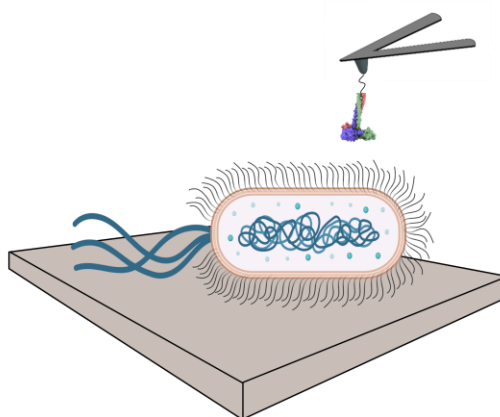


Figure 8.2: Super-resolution microscopy and AFM imaging for MGL-LPS nanoscale characterization. (A) Representative scheme of LPS metabolic labelling using Kdo-N₃ precursor followed by copper-free click chemistry labelling with a fluorophore compatible for both fluorescence microscopy and high-resolution microscopy. Image taken from ⁹⁸. (B) AFM-based Force Spectroscopy (AFM-FS) measurement of the mechanical properties of lectin-LPS interactions using MGL-specifically functionalized and bacterial cells chemically fixed on glass surfaces.

On the other hand, AFM, more precisely AFM-based force spectroscopy (AFM-FS) can be used for the measurement of the mechanical properties of lectin-LPS interactions, notably binding forces and rupture events. AFM-FS enables interaction analysis involving two biomolecules attached on the AFM tip and a substrate surface. Following-up with whole cells imaging, one way to perform AFM-FS would involve the immobilization of bacterial cells to glass substrates while the MGL counterpart can be grafted onto the cantilever tip by means of click chemistry-based approaches (e.g. MGL ECD-Biotin on streptavidin tips) ensuring convenient MGL orientation and specific attachment on their N-terminal extremity (Figure

8.2B). Such approach can provide a bunch of mechanical properties regarding MGL-LPS complex, including binding strength, lifetime of the complex, dissociation rate, and number of transitions in the dissociation process. Combining these analysis outcomes can give insights into the specificity of the interaction. A unique dissociation event presumably reflects the specificity of the binding with only a single orientation needed for the complex formation, whereas several dissociation events indicate a less-specific binding ²⁸⁷. Finally, the great advancements made so far in AFM and the development of High-Speed AFM (HS-AFM) have opened wide perspectives for molecular dynamics assessments of biomolecular complexes. Thanks to its elevated temporal resolution compared with conventional AFM, it is now possible to capture conformational changes of macromolecules upon interaction ²⁸⁸. Such technique could also be considered for the investigation of potential dynamic processes and conformational changes which may occur on the bacterial surface part in response to MGL binding, or potentially following antibiotic treatment.

IX.4. Link the in-vitro interaction to the immune system modulation

Following *in-vitro* and *ex-vivo*-like interaction studies we have already established, future investigations should incorporate immunological assessments. Previous studies have been carried out following *M. tuberculosis* infection and have reported the role of some CLRs in the immune response, including MMR, DC-SIGN and dectin. Moreover, the role of MGL in the recognition of *M. tuberculosis* has been described later by Naqvi and co-workers. They carried out *in vitro* and *in vivo* experiments which could demonstrate the antibacterial role of murine MGL-1 receptor during *M. tuberculosis* infection ²⁸⁹.

In our case, a similar approach could be explored but would require collaboration with research groups with a specific expertise. One way to proceed involves incubation of previously labelled bacterial cells (pathogens vs commensals) with macrophages. This set-up would allow the real-time tracking of potential phagocytosis events by fluorescence imaging (Figure 8.3A). Furthermore, in order to determine whether MGL is involved in the internalization process, MGL+ and MGL- macrophages could be generated and the immune response following infection with the studied bacterial strains could highlight the role and implication of MGL in modulating the immune responses ²⁸⁹. The evaluation of the expression level of the lectin of interest, for instance MGL, and its accumulation on the cell surface could also be monitored. Furthermore, administering LPS-SMA nanodiscs derived from the same

bacterial strains and monitoring the expression levels of cytokines and other effectors would provide valuable insights into the activated signalling pathway, and how different LPS variants modulate host immune responses (Figure 8.3B). By correlating the different outcomes observed across different bacterial strains, we can gain insight into the molecular mechanism underlying bacteria-host cell interactions, and their immune modulation properties.

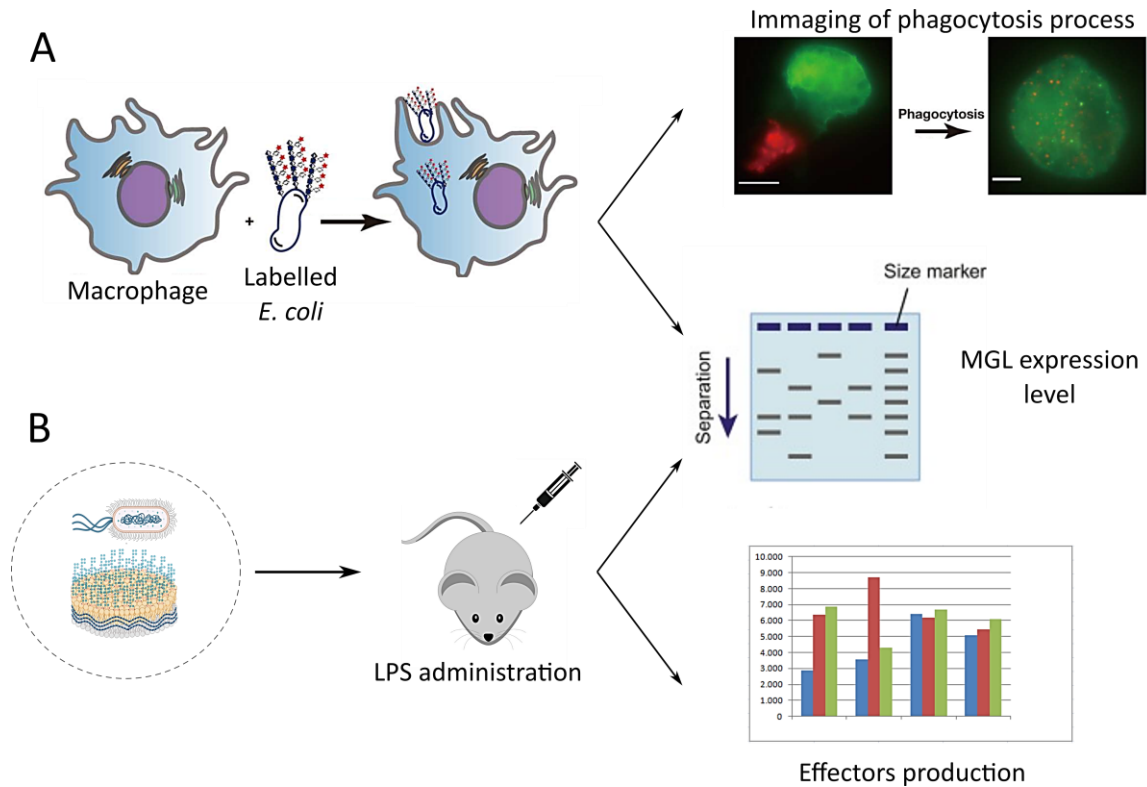


Figure 8.3: Immunological assessment of MGL-LPS interactions. (A) Fluorescence real-time monitoring of phagocytosis process of labelled bacterial cells by macrophages and potential protein MGL expression level examination. Adapted from ²⁸⁶. (B) LPS administration and effectors and MGL production evaluation.

Material and methods

- **MGL-ECD expression and purification**

MGL-ECD (UniProt entry number Q81UN9-2, residues Q61-H292) harbouring an N-terminal StrepTagII and a Factor Xa cleavage site was produced as reported¹⁰¹. Labelled MGL-ECD was over-expressed in *E. coli* BL21(DE3) cells in M9 minimal medium, enriched with ¹⁵N-NH₄Cl (1g/L) U- [¹³C]-glucose (2g/L), as inclusion bodies. The pellet was re-suspended in 30 mL **calcium buffer** (150 mM NaCl, 4 mM CaCl₂ and 25 mM TRIS pH 8) supplemented with one Complete EDTA-free anti-protease tablet. Cells were lysed by 2 sonication cycles (90% amplitude for 12 min in cycles of 2 s sonication & 10 s break). The pellet was recovered by ultracentrifugation at 100000 g for 30 min at 4°C, washed twice with 30 mL **urea buffer** (25 mM TRIS pH 8, 150 mM NaCl, 2 M urea and 1% Triton-X100), and with 30mL **calcium buffer** using a Potter-Elvehjem to resuspend it. An ultracentrifugation at 100000 g for 30 min at 4°C was performed after each washing step. Inclusion bodies were finally solubilized in 30 mL **guanidine buffer** (150 mM NaCl, 6 M Guanidine, 0.01% β-mercaptoethanol and 25 mM TRIS pH 8) and insoluble material was removed by ultracentrifugation for 30 min at 100000 g at 4°C. The protein concentration was adjusted to 2 mg/mL with guanidine buffer and the refolding was performed by a 5-fold drop-by-drop dilution in **renaturation buffer** (150 mM NaCl, 25 mM CaCl₂ and 100 mM TRIS pH 8) at 4°C under gentle stirring. The refolding product was dialyzed 3 times against **calcium buffer**, and the insoluble material was removed by a final ultracentrifugation step at 100000 g for 1 h at 4°C.

The wild type MGL-ECD proteins were purified with a two steps protocol⁸². First a 15 mL GalNAc-Agarose affinity column (Sigma) with an elution using **EDTA buffer** (150 mM NaCl, 25 mM TRIS pH 8, 10 mM EDTA). Second a 125 mL Toyopearl HW-50S gel filtration column (Tosoh Bioscience). Both columns are operating in tandem. On the other hand, purification of the D245H MGL-ECD mutant was carried out on an AktaXpress with a StrepTrap HP affinity column eluted with 2.5 mM D-desthiobiotin, followed by a 125 mL Toyopearl HW-50S gel filtration column (Tosoh Bioscience), both operating in tandem. Protein containing fractions were pooled together and concentrated by ultrafiltration using a vivaspin 20 PES, MWCO 10 KDa), flash frozen in liquid nitrogen and kept at -80°C.

- **NMR titration**

NMR titration experiments were performed on a 600 μM sample of the ^2H ^{15}N MGL ECD in calcium buffer with 1 and 2 molecular equivalents of both OS R1 and R3 ligands. ^1H - ^{15}N -BEST-TROSY correlation spectra were acquired at 308 K on a spectrometer operating at a magnetic field strength of 20 T corresponding to a proton Larmor frequency of 850 MHz. Acquisition times were typically 70 ms for direct dimension and 45 ms in the indirect dimension. A spectral width of 12 ppm and 36 ppm for the direct and indirect dimensions respectively, and a selective excited pulse centred at 8.5 ppm are applied. Spectra of 224 scans and a size of FID of 1452 and 280 for the direct and indirect dimensions, respectively, are recorded.

TopSpin 3.5 software was used for spectra processing. CcpNmr analysis 3.0 software was used for spectra analysis. CSPs in the ^1H - ^{15}N BEST-TROSY spectra upon OS titration were calculated $\text{CSP} = ((\Delta\delta^1\text{H})^2 + ([\Delta\delta^{15}\text{N}/6])^2)^{1/2}$. $\Delta\delta\text{H}$ represents chemical shift changes in amide proton, whereas $\Delta\delta\text{N}$ represents chemical shift changes in amide nitrogen, respectively. The threshold values are set to 2 times the standard deviation. Residues for which the shift change is 2 times greater than the standard deviation were considered as significantly perturbed.

- **LOS R1 micelles preparation**

E. coli R1 (F470) and *E. coli* R3 (F653) cells were grown in LB, and their respective LOSs were extracted following the phenol–chloroform–petroleum ether (PCP) method already described¹⁶⁴. DDM-LOS R1 and R3 micelles were prepared following the addition of 150 mM of DDM to purified LOS R1 (0.84 mM) and LOS R3 (0.6 mM) in 10 mM HEPES 150 mM NaCl 2 mM CaCl_2 pH 7.4. The mixture was kept under gentle rocking for 15 min at RT. Insoluble material was removed by ultracentrifugation at 100000 g for 30 min.

- **Surface plasmon resonance (SPR) analysis**

SPR direct interaction analysis was carried out using oriented surfaces of MGL ECD and DC-SIGN ECD that were beforehand specifically biotinylated on their N-terminus²¹⁵. Biotinylated proteins were immobilized onto the streptavidin functionalized surfaces allowing a uniform orientation of the lectins ECD. To do so, MGL Biot-ECD and DC-SIGN Biot-ECD were first diluted at 1 $\mu\text{g}/\text{mL}$ and 0.5 $\mu\text{g}/\text{mL}$, respectively, in running buffer (HBS-N, 2 mM CaCl_2 and 300 mM DDM buffer), then injected at 5 $\mu\text{L}/\text{min}$ till they reached a capture level of 1254 RU and 1264 RU, respectively. As for interaction measurements, LOS R1 and R3

micelles were injected at increasing concentrations ranging from 1.32 to 85 μM for LOS R1 micelles and 0.97 to 250 μM for LOS R3 micelles over lectin-oriented surfaces in running buffer at 20 $\mu\text{L}/\text{min}$. 50 mM EDTA, pH 8 was used for the regeneration of the surfaces. Streptavidin flow cell surface was used as reference for the binding response correction. Biacore T200 Evaluation Software 3.2.1 (GE Healthcare) was used for binding curves analysis. Lastly, data were fit using Steady State Affinity model.

- **MGL-ECD labelling with Alexa fluor 647nm**

MGL ECD at 2 mg/mL in PBS pH 7 was incubated in 100 mM sodium bicarbonate and 0.2 mg/mL Alexa Fluor 647 (AF647) NHS for 1 h at RT. A G25-PD10 desalting column (GE Healthcare) was used to remove excess of dye. Fractions enriched with MGL ECD were dialyzed against PBS pH 7 and concentrated.

- **FACS analysis of MGL-ECD binding onto R1, R3, O157:H7 and O113:H21 *E. coli* cells**

Cells were recovered by a centrifugation at 2500 g for 3 min and incubated with the same volume of 666 nM labelled MGL in PBS pH 7 2 mM Ca^{2+} and stirred for 15 min at RT. Two centrifugations at 2500 g for 3 min were performed and the pellets were finally resuspended in 150 μL PBS pH 7. The samples are then kept on ice and protected from light until analysis.

The measurements were carried out on a VYB device (Miltenyi biotech). The laser used is Y3 which emits at 661 nm and was used at 661 V. The analysis was done using the MACS Quant software. The fluorescence index was calculated by multiplying the percentage of labelled cells by the value of median fluorescence and normalized to 100% with R1.

- **Nanodiscs preparation from purified *E. coli* LOS/LPS**

^{13}C - ^{15}N *E. coli* F470 (R1) LOS and *E. coli* O157:H7 LPS were previously purified using the Phenol/Chloroform/light Petroleum (PCP) and phenol/hot water methods, respectively¹⁶⁴. *E. coli* R1 LOS and O157:H7 LPS at 1.8 mg/mL in **buffer C** (200 mM NaCl, 50 mM TRIS pH 7.4) were incubated under gentle stirring with 1% SMALP 200 (w/v) (purchased from Orbiscope) for 30 min at RT. Insoluble material was removed by ultracentrifugation at 100000 g for 1 h. Residual polymer was removed using a sucrose gradient (prepared in buffer C) prepared by layering sucrose solutions in the following order: 2 ml 45% (w/v), 3 ml 25%, 4 ml 5%, and 1 mL of SMA-LOS preparation layered on top. Sucrose gradients were centrifuged in a SW41Ti swinging-bucket rotor (Beckman) at 274000 g, overnight at 4 °C. 1 ml fractions

were collected, and the ones enriched in LPS-SMA nanodiscs are pooled, dialysed against buffer C, and concentrated using a Vivaspin Turbo 15, MWCO 100 KDa.

- **Outer membranes LPS Nanodiscs' preparation**

E. coli cells were grown in M9/H₂O minimal medium, enriched with ¹⁵N-NH₄Cl (1 g/L) U-[¹³C] -glucose (2 g/L) up to mid-exponential phase (1.1 OD_{260nm}). Cells were harvested by centrifugation at 5500 g at 4°C for 15 min. Outer membranes (OM) were prepared after high-pressure cell lysis and total membrane (TM) collection by ultracentrifugation at 100000 g for 1 h. The two membranes (IM and OM) were separated using an adjusted sucrose gradient to the studied strains: 20%/53%/73% (w/v), or 20%/45%/73% (w/v) (1/4/3 mL) on which 2mL of the resuspended TM is layered on top. Extracted OMs were solubilized using 1% SMALP 200 (w/v) for 2 h at RT. Insoluble material was removed by ultracentrifugation at 100000 g for 1 h followed by a second sucrose density gradient for residual polymer removal (45%/25%/5% (w/v)). The fractions enriched with nanodiscs were collected, dialysed 3 times against buffer C, and concentrated using a Vivaspin Turbo 15, MWCO 100 KDa.

- **ssNMR experiments**

For ssNMR analysis, the samples were diluted to 50% with buffer C, lyophilized and resuspended in 100% D₂O. Around 3 mg of the sample was sedimented into a 1.3 mm ssNMR rotor at 68000 g for 16 h. All hCH (CP- and INEPT-based correlations) and ¹³C-¹³C correlation spectra were collected on a 600 MHz spectrometer equipped with MAS 1.3 mm HCN probes. Pulse sequences from ssNMRLib were used for the experiments' setup²⁹⁰. An MAS frequency of 55 kHz and 15 kHz, for hCH and CC based experiments, respectively, were used and the sample temperature was set to 50°C.

For hCH-CP ¹H-¹³C correlation experiments of SMA-R1_{pur} nanodiscs, acquisition time were set to 14.9 ms for direct dimension and 8 ms in the indirect dimension. A spectral width of 30 ppm and 120 ppm for the direct and indirect dimensions, respectively, was used. A spectrum of 104 scans and a size of FID of 538 and 290 for the direct and indirect dimensions, respectively, is recorded.

For hCH-CP ¹H-¹³C correlation experiments of SMA-R3_{OM} nanodiscs, acquisition time of both direct and indirect dimensions is set to 14.9 ms and 6.8 ms, respectively. 30 ppm and 140 ppm for the direct and indirect dimensions, respectively, are used for the spectral width.

104 scans and a size of FID of 538 and 290 for the direct and indirect dimensions, respectively, are used.

hCH-CP ^1H - ^{13}C correlation experiments of SMA-O157_{pur} nanodiscs were recorded in duplicates then summed. Each experiment was recorded using the following parameters: acquisition time of 14.9 ms and 8 ms for direct and indirect dimensions, respectively. A spectral width of 30 ppm and 120 ppm for both the direct and indirect dimensions, respectively. Lastly, 208 scans and an FID size of 538 and 290 for both direct and indirect dimensions, respectively. On the other hand, for hCH-INEPT ^1H - ^{13}C correlation experiments, the acquisition time was set to 14.9 ms and 21 ms for direct and indirect dimensions, respectively. The same spectral width as in CP based experiment with 16 scans and a size of FID of 538 and 768 for the direct and indirect dimensions, respectively, are used. Lastly, ^{13}C - ^{13}C correlation experiment was recorded using an acquisition time of 14.9 ms and 8.5 ms for direct and indirect dimensions, respectively. A spectral width of 251 ppm and 140 ppm, and an FID size of 1132 and 360 for direct and indirect dimensions, respectively. The experiment was performed using 48 scans.

hCH-CP ^1H - ^{13}C correlation experiments of SMA-O157_{OM} nanodiscs were also recorded in duplicates and summed. Each experiment was acquired at an acquisition time of 14.9 ms and 6.8 ms for direct and indirect dimensions, respectively. A spectral width of 30 ppm and 140 ppm for the direct and indirect dimensions, respectively. Lastly, 144 scans and an FID size of 538 and 290 for the direct and indirect dimensions, respectively. hCH-INEPT ^1H - ^{13}C correlation experiment on the other hand was recorded using 32 scans and an acquisition time of 14.9 ms and 21 ms for direct and indirect dimensions, respectively. A size of FID of 538 and 888 for the direct and indirect dimensions, respectively.

TopSpin 3.5 and CcpNmr analysis 2.42 softwares were used for spectra processing and analysis, respectively.

- **Solution NMR experiments on LOS R3 DHPC micelles**

^{13}C , ^{15}N -labelled LOS R3 were extracted following the described phenol–chloroform–petroleum ether (PCP) method ¹⁶⁴. Labelled LOS R3 at 4.94 mM were solubilized following the addition of 374 mM DHPC. A 15 min sonication cycle was performed and was followed by an incubation at room temperature for 1 h. Insoluble material was removed by ultracentrifugation at 100000 g for 1 h. A final sample at 4.16 mM of LOS R3 and 58.8 mM DHPC is obtained. The sample was diluted to 50% with buffer C, lyophilized and resuspended in 100% D_2O .

The NMR experiments were collected at 50°C on a 700 MHz spectrometer. Pulse sequences from the NMRLib library were used²⁹¹. ¹H-¹³C HSQC correlation experiment was recorded using 104 scans, and an acquisition time of 121 ms and 7 ms for direct and indirect dimensions, respectively. A spectral width of 12 ppm and 120 ppm, and an FID size of 2048 and 300 for direct and indirect dimensions, respectively, were used. The 3D hCCH-TOCSY experiment was recorded using 4 scans and an acquisition time of 76 ms for the direct dimension F3 and 7.6 ms for the indirect dimensions F2 and F1. A spectral width of 13.9 ppm for F3 and 74 ppm for both F2 and F1. A size of FID of 1502 for F3 and 200 for F2 and F1.

TopSpin 3.5 software was used for spectra processing. CcpNmr analysis 2.42 software was used for spectra analysis.

- **Bio-Layer Interferometry experiments**

SMA R1_{pur} nanodiscs, in PBS buffer pH 7, were biotinylated on SMA polymer carboxyls using EDC and biotin hydrazide where 3% of carboxyl groups were targeted for modification. For a modification of 3% of the carboxyl groups in 1 g of polymer, 135 µmol of EDC and 324 µmole of biotin hydrazide were added. The mixture was incubated for 2 h at RT, and dialysed against the initial buffer.

The interferometry measurements were carried out on an octet Red96(Fortébio) using streptavidin coated biosensors (SA). The biotinylated nanodiscs were immobilized in HEPES buffered saline buffer, 0.02% Tween20 to reach 0.3 nm of response. The functionalized sensors were first immersed in well containing interaction buffer (50 mM Phosphate 150 mM NaCl pH 8) for equilibration, then immersed in the wells containing MGL-ECD protein at concentrations ranging from 10 to 75 nM. Each cycle of the kinetic have been programmed as follow: baseline (1000 s), association (600 s), and dissociation (300 s).

Nanodiscs interactions with Polymyxin B (PmB) antibiotic was also monitored. SA sensors were functionalized with nanodiscs up to 0.7 nm response. They were then immersed in wells containing PmB at concentrations ranging from 1.14 to 8 µM with agitation for 600 s (association), followed by an immersion in buffer for 400 s (dissociation).

Data of both interactions were analysed with the octet evaluation system after subtraction of the biosensor reference (SA) to remove the share of non-specific interactions.

- **AFM characterization**

AFM imaging of SMA R1_{pur}, SMA R1_{OM}, SMA O157:H7_{pur} and SMA O157:H7_{OM} was performed using multimode 8 atomic force microscope with a Nanoscope V controller (Bruker) operated in PeakForce Tapping mode with a ScanAsyst Air cantilever. ($k = 0.4$ N/m, $F_q = 67$ kHz). Images were recorded with a scan size of $2 \times 2 \mu\text{m}^2$, at a line rate of 1 Hz and a resolution of 512×512 px².

Nanodiscs samples were loaded on a freshly cleaved mica surface coated with Ni²⁺ ion as follows. First, 2 min incubation of 3 μl NiCl₂ at 2 mM followed by nitrogen flow drying. Afterwards, 3 μL of a 34 $\mu\text{g}/\text{mL}$ nanodiscs solutions were deposited and incubated for 2 min then dried under nitrogen flow. Regarding nanodiscs-MGL complex imaging, Ni²⁺-coated surfaces were prepared by deposition of 3 μL of a 50/50 (v/v) mixture of nanodiscs and MGL-ECD at 34 $\mu\text{g}/\text{mL}$ and 4.7 $\mu\text{g}/\text{mL}$, respectively, followed by a 10 min incubation and drying under nitrogen flow. Height and surface area were obtained from two different images.

- **QCM-D**

Quartz crystal microbalance-dissipation (QCM-D) crystals were first exposed to buffer (25 mM TRIS, 100 mM NaCl, pH 7.4). After establishing a baseline signal, SMA R1_{pur} at 0.5 mg/mL was injected at a flow rate of 50 $\mu\text{L}/\text{min}$ followed by a buffer rinse to remove unattached material. MGL-ECD at 10 $\mu\text{g}/\text{mL}$ or PmB at 100 $\mu\text{g}/\text{mL}$ were added, until the frequency stabilizes, followed by a final buffer rinse for removal of loosely bound molecules. SMA R1_{pur}-PmB bilayer was left overnight with no liquid flow then rinsed again with buffer. QCM-D measurement on SMA R1_{pur}-MGL ECD were carried out at 25°C, while SMA R1_{pur}-PmB experiments were carried out at 37°C.

- **Production of DepsiDBCO MGL ECD**

To produce DepsiDBCO MGL ECD, MGL ECD was first subjected to a Streptag cleavage using Factor Xa protease followed by a size exclusion chromatography on a Toyopearl HW-50S. Afterwards, sortase A ligation reaction using a mixture of cleaved MGL ECD (1 molar equivalent), DBCO-LPET(depsi)GG as a substrate (1.5 molar equivalent), and SrtA (0.2 molar equivalent) was performed. This reaction was carried out at 37°C for 6 h in 25 mM TRIS pH 8, 150 mM NaCl, 4 mM CaCl₂ buffer. Reaction product was purified on a GalNAc-agarose affinity chromatography followed with a SEC on a Toyopearl HW-50S as previously described in MGL ECD expression and purification.

References

1. Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2**, a000414 (2010).
2. Marquardt, D., Geier, B. & Pabst, G. Asymmetric Lipid Membranes: Towards More Realistic Model Systems. *Membranes (Basel)* **5**, 180 (2015).
3. Malanovic, N. & Lohner, K. Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. *Biochim Biophys Acta* **1858**, 936–946 (2016).
4. Neuhaus, F. C. & Baddiley, J. A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol Mol Biol Rev* **67**, 686–723 (2003).
5. Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol Rev* **40**, 133–159 (2016).
6. Garde, S., Chodisetti, P. K. & Reddy, M. Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus* **9**, (2021).
7. Vollmer, W., Blanot, D. & De Pedro, M. A. Peptidoglycan structure and architecture. *FEMS Microbiol Rev* **32**, 149–167 (2008).
8. Vollmer, W. Structural variation in the glycan strands of bacterial peptidoglycan. *FEMS Microbiol Rev* **32**, 287–306 (2008).
9. Rico-Pérez, G. *et al.* A novel peptidoglycan D,L-endopeptidase induced by Salmonella inside eukaryotic cells contributes to virulence. *Mol Microbiol* **99**, 546–556 (2016).
10. Humann, J. & Lenz, L. L. Bacterial Peptidoglycan-Degrading Enzymes and Their Impact on Host Muropeptide Detection. *J Innate Immun* **1**, 88 (2009).
11. Henderson, J. C. *et al.* The Power of Asymmetry: Architecture and Assembly of the Gram-Negative Outer Membrane Lipid Bilayer. *Annu Rev Microbiol* **70**, 255–278 (2016).
12. 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* **15**, 2561–2570 (1976).
13. Konovalova, A., Kahne, D. E. & Silhavy, T. J. Outer Membrane Biogenesis. *Annu Rev Microbiol* **71**, 539–556 (2017).
14. Konovalova, A. & Silhavy, T. J. Outer membrane lipoprotein biogenesis: Lol is not the end. *Philos Trans R Soc Lond B Biol Sci* **370**, (2015).
15. Wilson, M. M. & Bernstein, H. D. Surface-Exposed Lipoproteins: An Emerging Secretion Phenomenon in Gram-Negative Bacteria. *Trends Microbiol* **24**, 198–208 (2016).
16. Vogt, J. & Schulz, G. E. The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. *Structure* **7**, 1301–1309 (1999).
17. Dong, H. *et al.* Structural basis for outer membrane lipopolysaccharide insertion. *Nature* **511**, 52–56 (2014).

18. Ishida, H., Garcia-Herrero, A. & Vogel, H. J. The periplasmic domain of *Escherichia coli* outer membrane protein A can undergo a localized temperature dependent structural transition. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1838**, 3014–3024 (2014).
19. Konovalova, A., Perlman, D. H., Cowles, C. E. & Silhavy, T. J. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins. *Proc Natl Acad Sci U S A* **111**, E4350–E4358 (2014).
20. Yamashita, E., Zhalnina, M. V., Zakharov, S. D., Sharma, O. & Cramer, W. A. Crystal structures of the OmpF porin: function in a colicin translocon. *EMBO J* **27**, 2171–2180 (2008).
21. Di Lorenzo, F. *et al.* A Journey from Structure to Function of Bacterial Lipopolysaccharides. *Chem Rev* **122**, 15767–15821 (2021).
22. Paracini, N., Schneck, E., Imberty, A. & Micciulla, S. Lipopolysaccharides at Solid and Liquid Interfaces: Models for Biophysical Studies of the Gram-negative Bacterial Outer Membrane. *Adv Colloid Interface Sci* **301**, 102603 (2022).
23. Molinaro, A. *et al.* Chemistry of Lipid A: At the Heart of Innate Immunity. *Chemistry – A European Journal* **21**, 500–519 (2015).
24. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annual Review of Biochemistry* vol. 71 635–700 Preprint at <https://doi.org/10.1146/annurev.biochem.71.110601.135414> (2002).
25. Raetz, C. R. H., Reynolds, C. M., Trent, M. S. & Bishop, R. E. Lipid A modification systems in gram-negative bacteria. *Annu Rev Biochem* **76**, 295–329 (2007).
26. Amor, K. *et al.* Distribution of core oligosaccharide types in lipopolysaccharides from *Escherichia coli*. *Infect Immun* **68**, 1116–1124 (2000).
27. Preston, A., Mandrell, R. E., Gibson, B. W. & Apicella, M. A. The lipooligosaccharides of pathogenic gram-negative bacteria. *Crit Rev Microbiol* **22**, 139–180 (1996).
28. Knirel, Y. A. Structure of O-Antigens. *Bacterial Lipopolysaccharides* 41–115 (2011) doi:10.1007/978-3-7091-0733-1_3.
29. Whitfield, C. Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**, 39–68 (2006).
30. Kuhn, H.-M., Meier-Dieter, U. & Mayer, H. ECA, the enterobacterial common antigen. *FEMS Microbiol Rev* **4**, 195–222 (1988).
31. Stenutz, R., Weintraub, A. & Widmalm, G. The structures of *Escherichia coli* O-polysaccharide antigens. *FEMS Microbiol Rev* **30**, 382–403 (2006).
32. Murray, G. L., Attridge, S. R. & Morona, R. Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and complement. *J Bacteriol* **188**, 2735–2739 (2006).
33. Broeker, N. K. & Barbirz, S. Not a barrier but a key: How bacteriophages exploit host's O-antigen as an essential receptor to initiate infection. *Mol Microbiol* **105**, 353–357 (2017).
34. Bertani, B. & Ruiz, N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus* **8**, (2018).

35. Okuda, S., Sherman, D. J., Silhavy, T. J., Ruiz, N. & Kahne, D. Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model HHS Public Access. *Nat Rev Microbiol* **14**, 337–345 (2016).
36. Ho, H. *et al.* Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196–201 (2018).
37. Kalynych, S., Morona, R. & Cygler, M. Progress in understanding the assembly process of bacterial O-antigen. *FEMS Microbiol Rev* **38**, 1048–1065 (2014).
38. Tang, X. *et al.* Cryo-EM structures of lipopolysaccharide transporter LptB2FGC in lipopolysaccharide or AMP-PNP-bound states reveal its transport mechanism. *Nature Communications* **10**, 1–12 (2019).
39. Sperandio, P., Dehò, G. & Polissi, A. Lipopolysaccharide Export to the Outer Membrane. *Bacterial Lipopolysaccharides* 311–337 (2011) doi:10.1007/978-3-7091-0733-1_10.
40. Hicks, G. & Jia, Z. Structural Basis for the Lipopolysaccharide Export Activity of the Bacterial Lipopolysaccharide Transport System. *International Journal of Molecular Sciences* **19**, 2680 (2018).
41. Luo, Q. *et al.* Structural basis for lipopolysaccharide extraction by ABC transporter LptB2FG. *Nature Structural & Molecular Biology* **24**, 469–474 (2017).
42. Li, Y., Orlando, B. J. & Liao, M. Structural basis of lipopolysaccharide extraction by the LptB 2 FGC complex. *Nature* **567**, 486–490 (2019).
43. Bowyer, A., Baardsnes, J., Ajamian, E., Zhang, L. & Cygler, M. Characterization of interactions between LPS transport proteins of the Lpt system. *Biochem Biophys Res Commun* **404**, 1093–1098 (2011).
44. Sperandio, P. *et al.* New Insights into the Lpt Machinery for Lipopolysaccharide Transport to the Cell Surface: LptA-LptC Interaction and LptA Stability as Sensors of a Properly Assembled Transenvelope Complex. *J Bacteriol* **193**, 1042 (2011).
45. Suits, M. D. L., Sperandio, P., Dehò, G., Polissi, A. & Jia, Z. Novel Structure of the Conserved Gram-Negative Lipopolysaccharide Transport Protein A and Mutagenesis Analysis. *J Mol Biol* **380**, 476–488 (2008).
46. Botos, I. *et al.* Structural and Functional Characterization of the LPS Transporter LptDE from Gram-Negative Pathogens. *Structure* **24**, 965–976 (2016).
47. Botos, I., Noinaj, N. & Buchanan, S. K. Insertion of proteins and lipopolysaccharide into the bacterial outer membrane. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372**, (2017).
48. Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol Mol Biol Rev* **67**, 593–656 (2003).
49. H, N. & M, V. Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* **49**, 1–32 (1985).
50. Guo, L. *et al.* Lipid A Acylation and Bacterial Resistance against Vertebrate Antimicrobial Peptides. *Cell* **95**, 189–198 (1998).

51. Reynolds, C. M. *et al.* An Outer Membrane Enzyme Encoded by *Salmonella typhimurium* lpxR That Removes the 3'-Acylxyacyl Moiety of Lipid A. *J Biol Chem* **281**, 21974 (2006).
52. Trent, M. S., Pabich, W., Raetz, C. R. H. & Miller, S. I. A PhoP/PhoQ-induced Lipase (PagL) that catalyzes 3-O-deacylation of lipid A precursors in membranes of *Salmonella typhimurium*. *J Biol Chem* **276**, 9083–9092 (2001).
53. Gibbons, H. S., Lin, S., Cotter, R. J. & Raetz, C. R. H. Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in *Salmonella typhimurium* lipid A. Function of LpxO, A new Fe²⁺/α-ketoglutarate-dependent dioxygenase homologue. *J Biol Chem* **275**, 32940–32949 (2000).
54. Kawasaki, T. & Kawai, T. Toll-like receptor signaling pathways. *Front Immunol* **5**, (2014).
55. Rathinam, V. A. K., Zhao, Y. & Shao, F. Innate immunity to intracellular LPS. *Nature Immunology* 2019 20:5 **20**, 527–533 (2019).
56. Kagan, J. C. *et al.* TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-β. *Nat Immunol* **9**, (2008).
57. Botos, I., Segal, D. M. & Davies, D. R. The structural biology of Toll-like receptors. *Structure* **19**, 447–459 (2011).
58. 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* **46**, 38–50 (2017).
59. Needham, B. D. & Trent, M. S. Fortifying the barrier: The impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews Microbiology* vol. 11 467–481 Preprint at <https://doi.org/10.1038/nrmicro3047> (2013).
60. Ohto, U., Fukase, K., Miyake, K. & Shimizu, T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc Natl Acad Sci U S A* **109**, 7421–7426 (2012).
61. Seok Park, B. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature (London)* **458**, 1191–1195 (2009).
62. Rietschel, E. T. *et al.* Bacterial endotoxin: molecular relationships of structure to activity and function. *The FASEB Journal* **8**, 217–225 (1994).
63. Boonen, B., Alpizar, Y. A., Meseguer, V. M. & Talavera, K. TRP Channels as Sensors of Bacterial Endotoxins. *Toxins (Basel)* **10**, (2018).
64. Zhang, M. *et al.* TRP (transient receptor potential) ion channel family: structures, biological functions and therapeutic interventions for diseases. *Signal Transduction and Targeted Therapy* 2023 8:1 **8**, 1–38 (2023).
65. Mazgaeen, L. & Gurung, P. Recent Advances in Lipopolysaccharide Recognition Systems. *International Journal of Molecular Sciences* 2020, Vol. 21, Page 379 **21**, 379 (2020).
66. Dalsgaard, T., Sonkusare, S. K., Teuscher, C., Poynter, M. E. & Nelson, M. T. Pharmacological inhibitors of TRPV4 channels reduce cytokine production, restore endothelial function and increase survival in septic mice. *Sci Rep* **6**, (2016).

67. Vanaja, S. K. *et al.* Bacterial Outer Membrane Vesicles Mediate Cytosolic Localization of LPS and Caspase-11 Activation. *Cell* **165**, 1106–1119 (2016).
68. Deng, M. *et al.* The Endotoxin Delivery Protein HMGB1 Mediates Caspase-11-Dependent Lethality in Sepsis. *Immunity* **49**, 740–753.e7 (2018).
69. Yi, Y. S. Caspase-11 non-canonical inflammasome: a critical sensor of intracellular lipopolysaccharide in macrophage-mediated inflammatory responses. *Immunology* **152**, 207–217 (2017).
70. Erbs, G. & Newman, M. A. The role of lipopolysaccharide and peptidoglycan, two glycosylated bacterial microbe-associated molecular patterns (MAMPs), in plant innate immunity. *Mol Plant Pathol* **13**, 95–104 (2012).
71. Zipfel, C. Plant pattern-recognition receptors. *Trends Immunol* **35**, 345–351 (2014).
72. Ranf, S. *et al.* A lectin S-domain receptor kinase mediates lipopolysaccharide sensing in *Arabidopsis thaliana*. *Nat Immunol* **16**, 426–433 (2015).
73. Zeidler, D. *et al.* Innate immunity in *Arabidopsis thaliana*: lipopolysaccharides activate nitric oxide synthase (NOS) and induce defense genes. *Proc Natl Acad Sci U S A* **101**, 15811–15816 (2004).
74. Shang-Guan, K. *et al.* Lipopolysaccharides Trigger Two Successive Bursts of Reactive Oxygen Species at Distinct Cellular Locations. *Plant Physiol* **176**, 2543–2556 (2018).
75. Silipo, A. *et al.* The elicitation of plant innate immunity by lipooligosaccharide of *Xanthomonas campestris*. *J Biol Chem* **280**, 33660–33668 (2005).
76. Silipo, A. *et al.* The acylation and phosphorylation pattern of lipid A from *Xanthomonas campestris* strongly influence its ability to trigger the innate immune response in *Arabidopsis*. *Chembiochem* **9**, 896–904 (2008).
77. Newman, M. A., Sundelin, T., Nielsen, J. T. & Erbs, G. MAMP (microbe-associated molecular pattern) triggered immunity in plants. *Front Plant Sci* **4**, (2013).
78. Aslam, S. N. *et al.* Microbe-associated molecular pattern (MAMP) signatures, synergy, size and charge: influences on perception or mobility and host defence responses. *Mol Plant Pathol* **10**, 375 (2009).
79. Brown, G. D., Willment, J. A. & Whitehead, L. C-type lectins in immunity and homeostasis. *Nature Reviews Immunology* **18**, 374–389 (2018).
80. Hoving, J. C., Wilson, G. J. & Brown, G. D. Signalling C-type lectin receptors, microbial recognition and immunity. *Cell Microbiol* **16**, 185–194 (2014).
81. Mnich, M. E., van Dalen, R. & van Sorge, N. M. C-Type Lectin Receptors in Host Defense Against Bacterial Pathogens. *Front Cell Infect Microbiol* **10**, (2020).
82. Abbas, M. *et al.* The unique 3D arrangement of macrophage galactose lectin enables *Escherichia coli* lipopolysaccharide recognition through two distinct interfaces. *PNAS Nexus* **2**, (2023).
83. Cambi, A., Koopman, M. & Figdor, C. G. How C-type lectins detect pathogens. *Cell Microbiol* **7**, 481–488 (2005).
84. Drickamer, K. Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature* **360**, 183–186 (1992).

85. van Kooyk, Y. C-type lectins on dendritic cells: key modulators for the induction of immune responses. *Biochem Soc Trans* **36**, 1478–1481 (2008).
86. Zelensky, A. N. & Gready, J. E. The C-type lectin-like domain superfamily. *FEBS J* **272**, 6179–6217 (2005).
87. Dam, T. K. & Brewer, C. F. Effects of clustered epitopes in multivalent ligand-receptor interactions. *Biochemistry* **47**, 8470–8476 (2008).
88. 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*. *Journal of Biological Chemistry* **276**, 28939–28945 (2001).
89. Dennis, J. W. & Brewer, C. F. Density-dependent lectin-glycan interactions as a paradigm for conditional regulation by posttranslational modifications. *Mol Cell Proteomics* **12**, 913–920 (2013).
90. 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* **21**, 2999–3025 (2020).
91. Yan, H., Kamiya, T., Suabjakyong, P. & Tsuji, N. M. Targeting C-Type Lectin Receptors for Cancer Immunity. *Front Immunol* **6**, (2015).
92. Geijtenbeek, T. B. H. & Gringhuis, S. I. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* **9**, 465–479 (2009).
93. Gabba, A. *et al.* Crystal Structure of the Carbohydrate Recognition Domain of the Human Macrophage Galactose C-Type Lectin Bound to GalNAc and the Tumor-Associated Tn Antigen. *Biochemistry* **60**, 1327–1336 (2021).
94. van Vliet, S. J. *et al.* MGL-mediated internalization and antigen presentation by dendritic cells: a role for tyrosine-5. *Eur J Immunol* **37**, 2075–2081 (2007).
95. Engering, A. *et al.* The Dendritic Cell-Specific Adhesion Receptor DC-SIGN Internalizes Antigen for Presentation to T Cells. *The Journal of Immunology* **168**, 2118–2126 (2002).
96. Drickamer, K. & Taylor, M. E. Recent insights into structures and functions of C-type lectins in the immune system. *Curr Opin Struct Biol* **34**, 26–34 (2015).
97. Acosta, M. P. & Lepenies, B. Bacterial glycans and their interactions with lectins in the innate immune system. *Biochem Soc Trans* **47**, 1569–1579 (2019).
98. Nieto-Fabregat, F. *et al.* Molecular recognition of Escherichia coli R1-type core lipooligosaccharide by DC-SIGN. *iScience* **27**, (2024).
99. Imai, T. *et al.* Lipoteichoic acid anchor triggers Mincle to drive protective immunity against invasive group A Streptococcus infection. *Proc Natl Acad Sci U S A* **115**, E10662–E10671 (2018).
100. Van Vliet, S. J. *et al.* Variation of Neisseria gonorrhoeae Lipooligosaccharide Directs Dendritic Cell-Induced T Helper Responses. *PLoS Pathog* **5**, e1000625 (2009).
101. Maalej, M. M. *et al.* The human macrophage galactose-type lectin, MGL, recognizes the outer core of E. coli lipooligosaccharide. *ChemBioChem* **20**, cbic.201900087 (2019).

102. Simeone, R. *et al.* Cytosolic Access of Mycobacterium tuberculosis: Critical Impact of Phagosomal Acidification Control and Demonstration of Occurrence In Vivo. *PLoS Pathog* **11**, e1004650 (2015).
103. Van Kooyk, Y. & Geijtenbeek, T. B. H. DC-SIGN: ESCAPE MECHANISM FOR PATHOGENS. *Nat Rev Immunol* **3**, (2003).
104. Sahly, H., Keisari, Y. & Ofek, I. Manno(rhamno)biose-containing capsular polysaccharides of Klebsiella pneumoniae enhance opsono-stimulation of human polymorphonuclear leukocytes. *J Innate Immun* **1**, 136–144 (2009).
105. Devi, S., Rajakumara, E. & Ahmed, N. Induction of Mincle by Helicobacter pylori and consequent anti-inflammatory signaling denote a bacterial survival strategy. *Scientific Reports 2015 5:1* **5**, 1–13 (2015).
106. 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* **23**, 853–864 (2013).
107. 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. *Journal of Biological Chemistry* **294**, 1300–1311 (2019).
108. 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. *Journal of Biological Chemistry* **277**, 20686–20693 (2002).
109. Szczykutowicz, J. Ligand Recognition by the Macrophage Galactose-Type C-Type Lectin: Self or Non-Self?—A Way to Trick the Host's Immune System. *International Journal of Molecular Sciences 2023, Vol. 24, Page 17078* **24**, 17078 (2023).
110. Bulteau, F. *et al.* Targeting Tn-Antigen-Positive Human Tumors with a Recombinant Human Macrophage Galactose C-Type Lectin. *Mol Pharm* **19**, 235–245 (2022).
111. 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* **17**, 661–669 (2005).
112. 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* **11**, 1768–1781 (2009).
113. Di Lorenzo, F., De Castro, C., Silipo, A. & Molinaro, A. Lipopolysaccharide structures of Gram-negative populations in the gut microbiota and effects on host interactions. *FEMS Microbiol Rev* **43**, 257–272 (2019).
114. Di Lorenzo, F. *et al.* Pairing Bacteroides vulgatus LPS Structure with Its Immunomodulatory Effects on Human Cellular Models. *ACS Cent Sci* **6**, 1602–1616 (2020).
115. Zhu, Q. *et al.* Chemical synthesis of glycans up to a 128-mer relevant to the O-antigen of Bacteroides vulgatus. *Nature Communications 2020 11:1* **11**, 1–7 (2020).
116. Control of Infectious Diseases, 1900-1999. *JAMA* **282**, 1029–1032 (1999).
117. Waksman, S. A., Schatz, A. & Reynolds, D. M. Production of antibiotic substances by actinomycetes. *Ann N Y Acad Sci* **1213**, 112–124 (2010).

118. Wright, P. M., Seiple, I. B. & Myers, A. G. The evolving role of chemical synthesis in antibacterial drug discovery. *Angew Chem Int Ed Engl* **53**, 8840–8869 (2014).
119. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: from targets to networks. *Nature Reviews Microbiology* **8**, 423–435 (2010).
120. Kapoor, G., Saigal, S. & Elongavan, A. Action and resistance mechanisms of antibiotics: A guide for clinicians. *J Anaesthesiol Clin Pharmacol* **33**, 300 (2017).
121. Tomasz, A. THE MECHANISM OF THE IRREVERSIBLE ANTIMICROBIAL EFFECTS OF PENICILLINS: How the Beta-Lactam Antibiotics Kill and Lyse Bacteria. *Ann. Rev. Microbiol* **33**, 113–150 (1979).
122. De Rosa, M., Verdino, A., Soriente, A. & Marabotti, A. The Odd Couple(s): An Overview of Beta-Lactam Antibiotics Bearing More Than One Pharmacophoric Group. *Int J Mol Sci* **22**, 1–21 (2021).
123. Zeng, X. & Lin, J. Beta-lactamase induction and cell wall metabolism in Gram-negative bacteria. *Front Microbiol* **4**, (2013).
124. Tipper, D. J. & Strominger, J. L. Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. *Proc Natl Acad Sci U S A* **54**, 1133–1141 (1965).
125. Zeng, D. *et al.* Approved Glycopeptide Antibacterial Drugs: Mechanism of Action and Resistance. *Cold Spring Harb Perspect Med* **6**, (2016).
126. Hutchings, M., Truman, A. & Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **51**, 72–80 (2019).
127. Schar, D., Klein, E. Y., Laxminarayan, R., Gilbert, M. & Van Boeckel, T. P. Global trends in antimicrobial use in aquaculture. *Scientific Reports 2020 10:1* **10**, 1–9 (2020).
128. Centers for Disease Control and Prevention (CDC), U. D. of H. and H. S. P. H. S. C. for D. C. and P. (CDC). ANTIBIOTIC RESISTANCE THREATS in the United States. (2013).
129. Fleming, A. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae. *Br J Exp Pathol* **10**, 226 (1929).
130. Levy, S. B. & Bonnie, M. Antibacterial resistance worldwide: Causes, challenges and responses. *Nat Med* **10**, S122–S129 (2004).
131. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat Rev Microbiol* **21**, 280–295 (2023).
132. Wilson, D. N., Hauryliuk, V., Atkinson, G. C. & O'Neill, A. J. Target protection as a key antibiotic resistance mechanism. *Nat Rev Microbiol* **18**, 637–648 (2020).
133. Cong, Y., Yang, S. & Rao, X. Vancomycin resistant Staphylococcus aureus infections: A review of case updating and clinical features. *J Adv Res* **21**, 169–176 (2020).
134. Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Resist Updat* **13**, 151–171 (2010).
135. Tooke, C. L. *et al.* β -Lactamases and β -Lactamase Inhibitors in the 21st Century. *J Mol Biol* **431**, 3472–3500 (2019).

136. Du, D. *et al.* Multidrug efflux pumps: structure, function and regulation. *Nat Rev Microbiol* **16**, 523–539 (2018).
137. Zhang, F. & Cheng, W. The Mechanism of Bacterial Resistance and Potential Bacteriostatic Strategies. *Antibiotics* **11**, (2022).
138. Wise, M. G., Horvath, E., Young, K., Sahm, D. F. & Kazmierczak, K. M. Global survey of *Klebsiella pneumoniae* major porins from ertapenem non-susceptible isolates lacking carbapenemases. *J Med Microbiol* **67**, 289–295 (2018).
139. Mishra, N. N. *et al.* Daptomycin resistance in enterococci is associated with distinct alterations of cell membrane phospholipid content. *PLoS One* **7**, (2012).
140. Draper, P. The outer parts of the mycobacterial envelope as permeability barriers. *Front Biosci* **3**, (1998).
141. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim Biophys Acta Proteins Proteom* **1794**, 808–816 (2009).
142. Romano, K. P. & Hung, D. T. Targeting LPS biosynthesis and transport in gram-negative bacteria in the era of multi-drug resistance. *Biochim Biophys Acta Mol Cell Res* **1870**, 119407 (2023).
143. Onishi, H. R. *et al.* Antibacterial agents that inhibit lipid A biosynthesis. *Science* **274**, 980–982 (1996).
144. Sabnis, A. & Edwards, A. M. Lipopolysaccharide as an antibiotic target. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1870**, 119507 (2023).
145. Williams, A. H., Immormino, R. M., Gewirth, D. T. & Raetz, C. R. H. Structure of UDP-N-acetylglucosamine acyltransferase with a bound antibacterial pentadecapeptide. *Proceedings of the National Academy of Sciences* **103**, 10877–10882 (2006).
146. Jenkins, R. J. & Dotson, G. D. Dual targeting antibacterial peptide inhibitor of early lipid A biosynthesis. *ACS Chem Biol* **7**, 1170–1177 (2012).
147. Nayar, A. S. *et al.* Novel antibacterial targets and compounds revealed by a high-throughput cell wall reporter assay. *J Bacteriol* **197**, 1726–1734 (2015).
148. Ho, H. *et al.* Structural basis for dual-mode inhibition of the ABC transporter MsbA. *Nature* **557**, 196–201 (2018).
149. Zhang, G. *et al.* Cell-based screen for discovering lipopolysaccharide biogenesis inhibitors. *Proc Natl Acad Sci U S A* **115**, 6834–6839 (2018).
150. Gronenberg, L. S. & Kahne, D. Development of an activity assay for discovery of inhibitors of lipopolysaccharide transport. *J Am Chem Soc* **132**, 2518–2519 (2010).
151. Zhang, X. *et al.* Identification of an anti-Gram-negative bacteria agent disrupting the interaction between lipopolysaccharide transporters LptA and LptC. *Int J Antimicrob Agents* **53**, 442–448 (2019).
152. Martin-Loeches, I., Dale, G. E. & Torres, A. Murepavadin: a new antibiotic class in the pipeline. *Expert Rev Anti Infect Ther* **16**, 259–268 (2018).
153. Mohapatra, S. S., Dwibedy, S. K. & Padhy, I. Polymyxins, the last-resort antibiotics: Mode of action, resistance emergence, and potential solutions. *J Biosci* **46**, (2021).

154. Poirel, L., Jayol, A. & Nordmanna, P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. *Clin Microbiol Rev* **30**, 557–596 (2017).
155. Sabnis, A. *et al.* Colistin kills bacteria by targeting lipopolysaccharide in the cytoplasmic membrane. *Elife* **10**, (2021).
156. Janssen, A. B. & van Schaik, W. Harder, better, faster, stronger: Colistin resistance mechanisms in *Escherichia coli*. *PLoS Genet* **17**, (2021).
157. Vaara, M. *et al.* Characterization of the lipopolysaccharide from the polymyxin-resistant *pmrA* mutants of *Salmonella typhimurium*. *FEBS Lett* **129**, 145–149 (1981).
158. Roberts, K. D. *et al.* A synthetic lipopeptide targeting top-priority multidrug-resistant Gram-negative pathogens. *Nat Commun* **13**, (2022).
159. Bruss, J. *et al.* Single- and Multiple-Ascending-Dose Study of the Safety, Tolerability, and Pharmacokinetics of the Polymyxin Derivative SPR206. *Antimicrob Agents Chemother* **65**, (2021).
160. Luther, A. *et al.* Chimeric peptidomimetic antibiotics against Gram-negative bacteria. *Nature* **576**, 452–458 (2019).
161. Wang, X., Zhang, C., Shi, F. & Hu, X. Purification and characterization of lipopolysaccharides. *Subcell Biochem* **53**, 27–51 (2010).
162. Westphal, O. ; J. K. Bacterial lipopolysaccharides. Extraction with phenol-water and further applications of the procedure. *Method Carbohyd Chem* **5**, 83–91 (1965).
163. Galanos, C., Lüderitz, O. & Westphal, O. A new method for the extraction of R lipopolysaccharides. *Eur J Biochem* **9**, 245–249 (1969).
164. 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* vol. 480 89–115 (Academic Press Inc., 2010).
165. Marolda, C. L., Lahiry, P., Vinés, E., Saldías, S. & Valvano, M. A. Micromethods for the characterization of lipid A-core and O-antigen lipopolysaccharide. *Methods Mol Biol* **347**, 237–252 (2006).
166. Hardy, E., Pupo, E., Castellanos-Serra, L., Reyes, J. & Fernández-Patrón, C. Sensitive reverse staining of bacterial lipopolysaccharides on polyacrylamide gels by using zinc and imidazole salts. *Anal Biochem* **244**, 28–32 (1997).
167. Hardy, E., Pupo, E., Santana, H., Guerra, M. & Castellanos-Serra, L. R. Elution of lipopolysaccharides from polyacrylamide gels. *Anal Biochem* **259**, 162–165 (1998).
168. Steinberg, T. H. *et al.* Rapid and simple single nanogram detection of glycoproteins in polyacrylamide gels and on electroblots. *Proteomics* **1**, 841–55 (2001).
169. Chafchaoui-Moussaoui, I. *et al.* A new rapid and micro-scale hydrolysis, using triethylamine citrate, for lipopolysaccharide characterization by mass spectrometry. *Rapid Commun Mass Spectrom* **25**, 2043–2048 (2011).
170. BLIGH, E. G. & DYER, W. J. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* **37**, 911–917 (1959).

171. Holst, O. Deacylation of lipopolysaccharides and isolation of oligosaccharide phosphates. *Methods Mol Biol* **145**, 345–353 (2000).
172. Wollenweber, H. W. & Rietschel, E. T. Analysis of lipopolysaccharide (lipid A) fatty acids. *J Microbiol Methods* **11**, 195–211 (1990).
173. Sawabdekeb, J. S., Sloneker, J. H. & Jeanes, A. Quantitative Determination of Monosaccharides as Their Alditol Acetates by Gas Liquid Chromatography. *Anal Chem* **37**, 1602–1604 (1965).
174. Ciucanu, I. & Kerek, F. A simple and rapid method for the permethylation of carbohydrates. *Carbohydr Res* **131**, 209–217 (1984).
175. 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* **62**, 359–362 (1978).
176. Kilar, A., Dörnyei, Á. & Kocsis, B. Structural characterization of bacterial lipopolysaccharides with mass spectrometry and on- and off-line separation techniques. *Mass Spectrom Rev* **32**, 90–117 (2013).
177. Kaltashov, I. A., Doroshenko, V., Cotter, R. J., Takayama, K. & Qureshi, N. Confirmation of the Structure of Lipid A Derived from the Lipopolysaccharide of *Rhodobacter sphaeroides* by a Combination of MALDI, LSIMS, and Tandem Mass Spectrometry. *Anal Chem* **69**, 2317–2322 (1997).
178. Aussel, L., Brisson, J. R., Perry, M. B. & Caroff, M. Structure of the lipid A of *Bordetella hinzii* ATCC 51730. *Rapid Commun Mass Spectrom* **14**, 595–599 (2000).
179. Sturiale, L. *et al.* New conditions for matrix-assisted laser desorption/ionization mass spectrometry of native bacterial R-type lipopolysaccharides. *Rapid Commun Mass Spectrom* **19**, 1829–1834 (2005).
180. 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* **266**, 460–466 (1999).
181. Banoub, J. H., El Aneed, A., Cohen, A. M. & Joly, N. Structural investigation of bacterial lipopolysaccharides by mass spectrometry and tandem mass spectrometry. *Mass Spectrom Rev* **29**, 606–650 (2010).
182. Duus, J., Gotfredsen, C. H. & Bock, K. Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev* **100**, 4589–4614 (2000).
183. Laguri, C. *et al.* Solid State NMR Studies of Intact Lipopolysaccharide Endotoxin. *ACS Chem Biol* **13**, (2018).
184. Polenova, T., Gupta, R. & Goldbourt, A. Magic angle spinning NMR spectroscopy: A versatile technique for structural and dynamic analysis of solid-phase systems. *Anal Chem* **87**, 5458–5469 (2015).
185. Mnich, M. E., van Dalen, R. & van Sorge, N. M. C-Type Lectin Receptors in Host Defense Against Bacterial Pathogens. *Front Cell Infect Microbiol* **10**, 1–15 (2020).
186. Geissner, A. *et al.* Microbe-focused glycan array screening platform. *Proc Natl Acad Sci U S A* **116**, 1958–1967 (2019).

187. Stowell, S. R. *et al.* Microbial Glycan Microarrays Define Key Features of Host-Microbial Interactions. *Nat Chem Biol* **10**, 470–476 (2014).
188. Hanske, J. *et al.* Bacterial polysaccharide specificity of the pattern recognition receptor langerin is highly species-dependent. *Journal of Biological Chemistry* **292**, 862–871 (2017).
189. Zheng, R. B. *et al.* Insights into Interactions of Mycobacteria with the Host Innate Immune System from a Novel Array of Synthetic Mycobacterial Glycans. *ACS Chem Biol* **12**, 2990–3002 (2017).
190. Jégouzo, S. A. F. *et al.* Mammalian lectin arrays for screening host-microbe interactions. *Journal of Biological Chemistry* **295**, 4541–4555 (2020).
191. Touhami, A., Hoffmann, B., Vasella, A., Denis, F. A. & Dufrêne, Y. F. Probing specific lectin-carbohydrate interactions using atomic force microscopy imaging and force measurements. *Langmuir* **19**, 1745–1751 (2003).
192. Lakshminarayanan, A., Richard, M. & Davis, B. G. Studying glycobiology at the single-molecule level. *Nature Reviews Chemistry 2018 2:8* **2**, 148–159 (2018).
193. Gour, N. & Verma, S. Synthesis and AFM studies of lectin-carbohydrate self-assemblies. *Tetrahedron* **64**, 7331–7337 (2008).
194. Touhami, A., Hoffmann, B., Vasella, A., Denis, F. A. & Dufrêne, Y. F. Aggregation of yeast cells: Direct measurement of discrete lectin-carbohydrate interactions. *Microbiology (N Y)* **149**, 2873–2878 (2003).
195. Chen, J. L., Lin, S. & Lin, L. P. Rhizobial surface biopolymers and their interaction with lectin measured by atomic force microscopy. *World J Microbiol Biotechnol* **22**, 565–570 (2006).
196. McKinnon, K. M. Flow Cytometry: An Overview. *Curr Protoc Immunol* **120**, 5.1.1 (2018).
197. 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* **201**, 313–324 (2019).
198. Vishweswaraiyah, R. H. *et al.* Monitoring of Microbial Safety of Foods Using Lectins: A Review. *Frontiers in Food Science and Technology* **2**, 842063 (2022).
199. McPherson, R. L. *et al.* Lectin-Seq: A method to profile lectin-microbe interactions in native communities. *Sci Adv* **9**, (2023).
200. Man-Kupisinska, A. *et al.* Interaction of mannose-binding lectin with lipopolysaccharide outer core region and its biological consequences. *Front Immunol* **9**, 385385 (2018).
201. Maity, S., Gundampati, R. K. & Kumar, T. K. S. NMR methods to characterize protein-ligand interactions. *Nat Prod Commun* **14**, (2019).
202. Cala, O., Guillière, F. & Krimm, I. NMR-based analysis of protein-ligand interactions. *Anal Bioanal Chem* **406**, 943–956 (2014).
203. Haselhorst, T., Lamerz, A. C. & Itzstein, M. Von. Saturation transfer difference NMR spectroscopy as a technique to investigate protein-carbohydrate interactions in solution. *Methods in Molecular Biology* **534**, 375–386 (2009).

204. Furukawa, A., Konuma, T., Yanaka, S. & Sugase, K. Quantitative analysis of protein-ligand interactions by NMR. *Prog Nucl Magn Reson Spectrosc* **96**, 47–57 (2016).
205. 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* **123**, 6108–6117 (2001).
206. Angulo, J. & Nieto, P. M. STD-NMR: Application to transient interactions between biomolecules—a quantitative approach. *European Biophysics Journal* **40**, 1357–1369 (2011).
207. Mayer, M. & Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew Chem Int Ed Engl* **38**, 1784–1788 (1999).
208. Quintana, J. I., Atxabal, U., Unione, L., Ardá, A. & Jiménez-Barbero, J. Exploring multivalent carbohydrate–protein interactions by NMR. *Chem Soc Rev* **52**, 1591–1613 (2023).
209. Miller, M. C., Nesmelova, I. V., Platt, D., Klyosov, A. & Mayo, K. H. The carbohydrate-binding domain on galectin-1 is more extensive for a complex glycan than for simple saccharides: implications for galectin-glycan interactions at the cell surface. *Biochem J* **421**, 211–221 (2009).
210. Laigre, E., Goyard, D., Tiertant, C., Dejeu, J. & Renaudet, O. The study of multivalent carbohydrate–protein interactions by bio-layer interferometry. *Org Biomol Chem* **16**, 8899–8903 (2018).
211. Fee, C. J. Label-free, real-time interaction and adsorption analysis 1: Surface plasmon resonance. *Methods in Molecular Biology* **996**, 287–312 (2013).
212. Kabiri, M. & Unsworth, L. D. Application of isothermal titration calorimetry for characterizing thermodynamic parameters of biomolecular interactions: Peptide self-assembly and protein adsorption case studies. *Biomacromolecules* **15**, 3463–3473 (2014).
213. Dam, T. K. & Brewer, C. F. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem Rev* **102**, 387–429 (2002).
214. Mandal, D. K., Kishore, N. & Brewer, C. F. Thermodynamics of Lectin-Carbohydrate Interactions. Titration Microcalorimetry Measurements of the Binding of N-Linked Carbohydrates and Ovalbumin to Concanavalin A. *Biochemistry* **33**, 1149–1156 (1994).
215. 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* **18**, (2020).
216. 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* **28**, 163–171 (2000).
217. Dörr, J. M. *et al.* The styrene–maleic acid copolymer: a versatile tool in membrane research. *European Biophysics Journal* vol. 45 3–21 Preprint at <https://doi.org/10.1007/s00249-015-1093-y> (2016).
218. Dörr, J. M. *et al.* The styrene–maleic acid copolymer: a versatile tool in membrane research. *European Biophysics Journal* **45**, 3–21 (2016).

219. Jamshad, M. *et al.* Structural analysis of a nanoparticle containing a lipid bilayer used for detergent-free extraction of membrane proteins. *Nano Res* **8**, 774–789 (2015).
220. Bjørnestad, V. A., Orwick-Rydmark, M. & Lund, R. Understanding the Structural Pathways for Lipid Nanodisc Formation: How Styrene Maleic Acid Copolymers Induce Membrane Fracture and Disc Formation. *Langmuir* **37**, 6178–6188 (2021).
221. Scheidelaar, S. *et al.* Molecular Model for the Solubilization of Membranes into Nanodisks by Styrene Maleic Acid Copolymers. *Biophys J* **108**, 279–290 (2015).
222. Xue, M., Cheng, L., Faustino, I., Guo, W. & Marrink, S. J. Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic Acid Copolymers. *Biophys J* **115**, 494–502 (2018).
223. Overduin, M. & Esmaili, M. Memtein: The fundamental unit of membrane-protein structure and function. *Chem Phys Lipids* **218**, 73–84 (2019).
224. Angelisová, P. *et al.* The use of styrene-maleic acid copolymer (SMA) for studies on T cell membrane rafts. *Biochim Biophys Acta Biomembr* **1861**, 130–141 (2019).
225. Teo, A. C. K. *et al.* Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein. *Sci Rep* **9**, (2019).
226. Sinha, S. *et al.* Immunochemical characterisation of styrene maleic acid lipid particles prepared from Mycobacterium tuberculosis plasma membrane. *PLoS One* **18**, e0280074 (2023).
227. Scheidelaar, S. *et al.* Effect of Polymer Composition and pH on Membrane Solubilization by Styrene-Maleic Acid Copolymers. *Biophys J* **111**, 1974–1986 (2016).
228. Vliet, S. J. van *et al.* Variation of Neisseria gonorrhoeae Lipooligosaccharide Directs Dendritic Cell-Induced T Helper Responses. *PLoS Pathog* **5**, e1000625 (2009).
229. 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* **11**, 1768–1781 (2009).
230. Achilli, S. *et al.* TETRALEC, Artificial Tetrameric Lectins: A Tool to Screen Ligand and Pathogen Interactions. *Int J Mol Sci* **21**, 1–20 (2020).
231. Porkolab, V. *et al.* Powerful Avidity with a Limited Valency for Virus-Attachment Blockers on DC-SIGN: Combining Chelation and Statistical Rebinding with Structural Plasticity of the Receptor. *ACS Cent Sci* **9**, 709–718 (2023).
232. Sijmons, D., Guy, A. J., Walduck, A. K. & Ramsland, P. A. Helicobacter pylori and the Role of Lipopolysaccharide Variation in Innate Immune Evasion. *Front Immunol* **13**, (2022).
233. Lerouge, I. & Vanderleyden, J. O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol Rev* **26**, 17–47 (2002).
234. Duerr, C. U. *et al.* O-Antigen Delays Lipopolysaccharide Recognition and Impairs Antibacterial Host Defense in Murine Intestinal Epithelial Cells. *PLoS Pathog* **5**, e1000567 (2009).

235. Chabrol, E. *et al.* Glycosaminoglycans are interactants of Langerin: comparison with gp120 highlights an unexpected calcium-independent binding mode. *PLoS One* **7**, (2012).
236. Furukawa, A. *et al.* Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL. *Proc Natl Acad Sci U S A* **110**, 17438–17443 (2013).
237. Wawrzinek, R. *et al.* A Remote Secondary Binding Pocket Promotes Heteromultivalent Targeting of DC-SIGN. *J Am Chem Soc* **143**, 18977–18988 (2021).
238. Hanske, J. *et al.* Calcium-Independent Activation of an Allosteric Network in Langerin by Heparin Oligosaccharides. *ChemBioChem* **18**, 1183–1187 (2017).
239. Gabba, A., Murphy, P. V., Kiessling, L. L. & Birrane, G. Beyond the Crystal Structure of Human Macrophage C-Type Lectin. *Biochemistry* **16**, 50 (2024).
240. Laguri, C. *et al.* Solid State NMR Studies of Intact Lipopolysaccharide Endotoxin. *ACS Chem Biol* **13**, 2106–2113 (2018).
241. Garcia-Vello, P. *et al.* Structure of the O-Antigen and the Lipid A from the Lipopolysaccharide of *Fusobacterium nucleatum* ATCC 51191. *ChemBioChem* **22**, 1252–1260 (2021).
242. Silipo, A. *et al.* Full Structural Characterisation of the Lipooligosaccharide of a *Burkholderia pyrocinia* Clinical Isolate. (2006) doi:10.1002/ejoc.200600520.
243. Kaufmann, S. *et al.* Supported Lipopolysaccharide Bilayers. (2012) doi:10.1021/la3020223.
244. Bahl, N. *et al.* Delineation of lipopolysaccharide (LPS)-binding sites on hemoglobin: From in silico predictions to biophysical characterization. *Journal of Biological Chemistry* **286**, 37793–37803 (2011).
245. Paracini, N., Clifton, L. A., Skoda, M. W. A. & Lakey, J. H. Liquid crystalline bacterial outer membranes are critical for antibiotic susceptibility. *Proc Natl Acad Sci U S A* **115**, E7587–E7594 (2018).
246. Günsel, U. & Hagn, F. Lipid Nanodiscs for High-Resolution NMR Studies of Membrane Proteins. *Chemical Reviews* vol. 122 9395–9421 Preprint at <https://doi.org/10.1021/acs.chemrev.1c00702> (2022).
247. Noh, I. *et al.* Cellular Nanodiscs Made from Bacterial Outer Membrane as a Platform for Antibacterial Vaccination. *ACS Nano* **17**, 1120–1127 (2023).
248. Klöpfer, K. & Hagn, F. Beyond detergent micelles: The advantages and applications of non-micellar and lipid-based membrane mimetics for solution-state NMR. *Prog Nucl Magn Reson Spectrosc* **114–115**, 271–283 (2019).
249. Cian, M. B., Giordano, N. P., Mettlach, J. A., Minor, K. E. & Dalebroux, Z. D. Separation of the cell envelope for gram-negative bacteria into inner and outer membrane fractions with technical adjustments for *acinetobacter baumannii*. *Journal of Visualized Experiments* **2020**, 1–8 (2020).
250. Dörr, J. M. *et al.* Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K⁺ channel: The power of native nanodiscs. *Proc Natl Acad Sci U S A* **111**, 18607–18612 (2014).

251. López, C. A., Zgurskaya, H. & Gnanakaran, S. Molecular characterization of the outer membrane of *Pseudomonas aeruginosa*. *Biochim Biophys Acta Biomembr* **1862**, (2020).
252. Zhang, R. *et al.* Characterizing the structure of lipodisq nanoparticles for membrane protein spectroscopic studies. *Biochim Biophys Acta Biomembr* **1848**, 329–333 (2015).
253. D’errico, G. *et al.* Mesoscopic and microstructural characterization of liposomes formed by the lipooligosaccharide from *Salmonella minnesota* strain 595 (Re mutant). (2009) doi:10.1039/b816248d.
254. D’errico, G. *et al.* Characterization of liposomes formed by lipopolysaccharides from *Burkholderia cenocepacia*, *Burkholderia multivorans* and *Agrobacterium tumefaciens*: from the molecular structure to the aggregate architecture. *Phys. Chem. Chem. Phys* **12**, 13574–13585 (2010).
255. Schanda, P. & Ernst, M. Studying dynamics by magic-angle spinning solid-state NMR spectroscopy: Principles and applications to biomolecules. *Prog Nucl Magn Reson Spectrosc* **96**, 1–46 (2016).
256. Manioglu, S. *et al.* Antibiotic polymyxin arranges lipopolysaccharide into crystalline structures to solidify the bacterial membrane. *Nat Commun* **13**, (2022).
257. Oh, Y. J., Plochberger, B., Rechberger, M. & Hinterdorfer, P. Characterizing the effect of polymyxin B antibiotics to lipopolysaccharide on *Escherichia coli* surface using atomic force microscopy. *Journal of Molecular Recognition* **30**, (2017).
258. Schmidt, V. & Sturgis, J. N. Modifying styrene-maleic acid co-polymer for studying lipid nanodiscs. *Biochim Biophys Acta Biomembr* **1860**, 777–783 (2018).
259. Thomas, C. J. & Surolia, A. Kinetics of the interaction of endotoxin with polymyxin B and its analogs: a surface plasmon resonance analysis. *FEBS Lett* **445**, 420–424 (1999).
260. Gales, A. C., Jones, R. N. & Sader, H. S. Global assessment of the antimicrobial activity of polymyxin B against 54 731 clinical isolates of Gram-negative bacilli: Report from the SENTRY antimicrobial surveillance programme (2001-2004). *Clinical Microbiology and Infection* **12**, 315–321 (2006).
261. Wang, K. F., Nagarajan, R., Mello, C. M. & Camesano, T. A. Characterization of Supported Lipid Bilayer Disruption By Chrysothysin-3 Using QCM-D. *J. Phys. Chem. B* **115**, 15228–15235 (2011).
262. Ravula, T., Hardin, N. Z. & Ramamoorthy, A. Polymer Nanodiscs: Advantages and Limitations. *Chem Phys Lipids* **219**, 45 (2019).
263. Ravula, T., Ramadugu, S. K., Di Mauro, G. & Ramamoorthy, A. Bioinspired, Size-Tunable Self-Assembly of Polymer-Lipid Bilayer Nanodiscs. *Angew Chem Int Ed Engl* **56**, 11466–11470 (2017).
264. Ravula, T., Hardin, N. Z., Ramadugu, S. K., Cox, S. J. & Ramamoorthy, A. Formation of pH-Resistant Monodispersed Polymer-Lipid Nanodiscs. *Angew Chem Int Ed Engl* **57**, 1342–1345 (2018).
265. Marconnet, A. *et al.* Solubilization and Stabilization of Membrane Proteins by Cycloalkane-Modified Amphiphilic Polymers. *Biomacromolecules* **21**, 3459–3467 (2020).

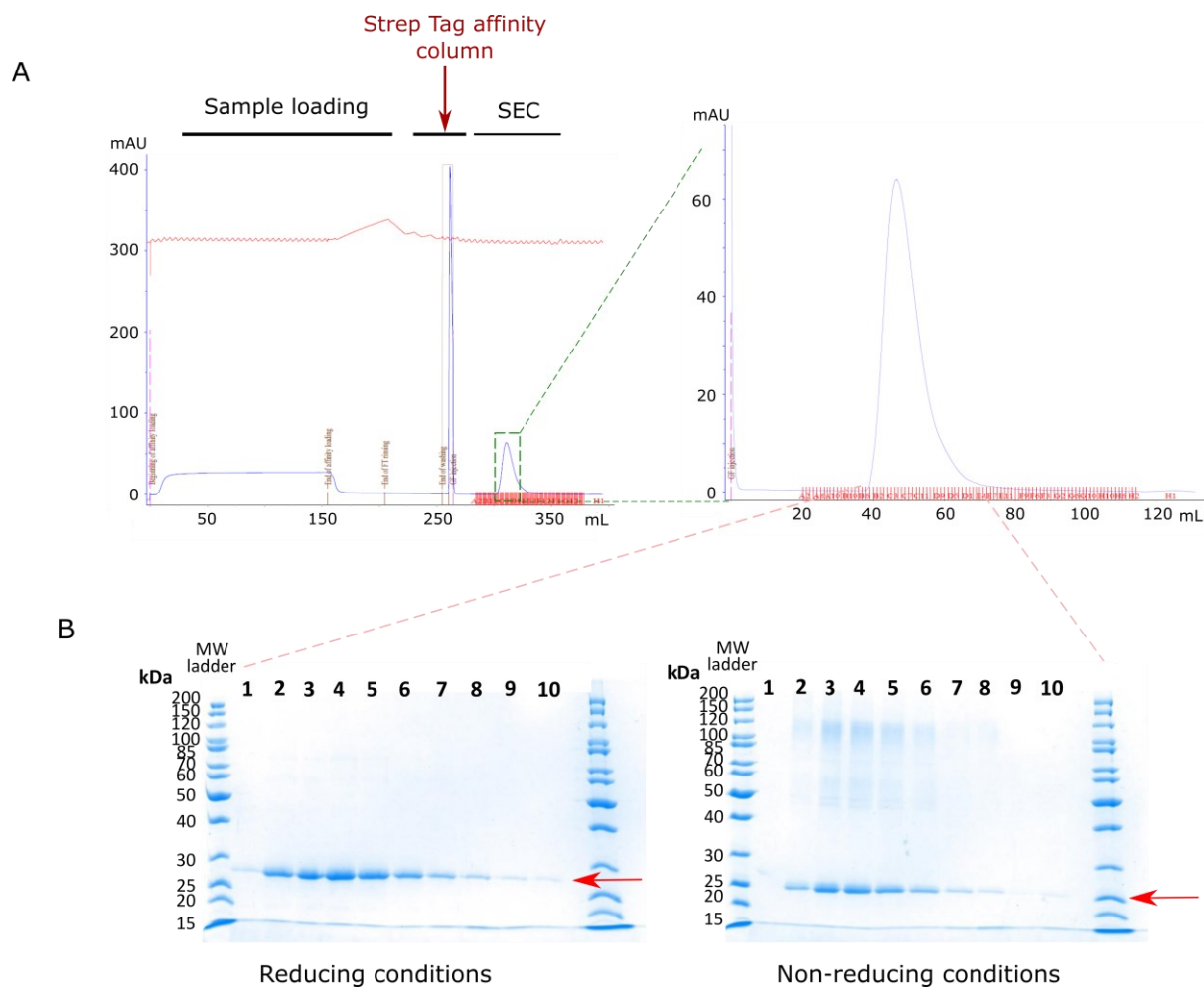
266. Marconnet, A. *et al.* Influence of Hydrophobic Groups Attached to Amphipathic Polymers on the Solubilization of Membrane Proteins along with Their Lipids. *Anal Chem* **94**, 14151–14158 (2022).
267. Thépaut, M. *et al.* DC/L-SIGN recognition of spike glycoprotein promotes SARS-CoV-2 trans-infection and can be inhibited by a glycomimetic antagonist. *PLoS Pathog* **17**, (2021).
268. Williamson, D. J., Fascione, M. A., Webb, M. E. & Turnbull, W. B. Efficient N-terminal labeling of proteins by use of sortase. *Angewandte Chemie - International Edition* **51**, 9377–9380 (2012).
269. Williamson, D. J., Webb, M. E. & Turnbull, W. B. Depsipeptide substrates for sortase-mediated N-terminal protein ligation. *Nature Protocols* **2014 9:2** **9**, 253–262 (2014).
270. Tabarani, G. *et al.* DC-SIGN neck domain is a pH-sensor controlling oligomerization. Saxs and hydrodynamic studies of extracellular domain. *Journal of Biological Chemistry* **284**, 21229–21240 (2009).
271. Feinberg, H., Powlesland, A. S., Taylor, M. E. & Weis, W. I. Trimeric Structure of Langerin. *Journal of Biological Chemistry* **285**, 13285–13293 (2010).
272. Carugo, O. & Djinović-Carugo, K. Structural biology: A golden era. *PLoS Biol* **21**, e3002187 (2023).
273. Jeffries, C. M. *et al.* Preparing monodisperse macromolecular samples for successful biological small-angle X-ray and neutron-scattering experiments. *Nature Protocols* **2016 11:11** **11**, 2122–2153 (2016).
274. Lakey, J. H. Recent advances in neutron reflectivity studies of biological membranes. *Curr Opin Colloid Interface Sci* **42**, 33–40 (2019).
275. Le Brun, A. P., Clifton, L. A., Holt, S. A., Holden, P. J. & Lakey, J. H. Deuterium Labeling Strategies for Creating Contrast in Structure–Function Studies of Model Bacterial Outer Membranes Using Neutron Reflectometry. *Methods Enzymol* **566**, 231–252 (2016).
276. Elbein, A. D. & Heath, E. C. The Biosynthesis of Cell Wall Lipopolysaccharide in *Escherichia coli*: I. THE BIOCHEMICAL PROPERTIES OF A URIDINE DIPHOSPHATE GALACTOSE 4-EPIMERASELESS MUTANT. *Journal of Biological Chemistry* **240**, 1919–1925 (1965).
277. Madl, T., Gabel, F. & Sattler, M. NMR and small-angle scattering-based structural analysis of protein complexes in solution. *J Struct Biol* **173**, 472–482 (2011).
278. Dias Mirandela, G. *et al.* Merging In-Solution X-ray and Neutron Scattering Data Allows Fine Structural Analysis of Membrane-Protein Detergent Complexes. *Journal of Physical Chemistry Letters* **9**, 3910–3914 (2018).
279. Schwieters, C. D. *et al.* Solution Structure of the 128 kDa Enzyme I Dimer from *Escherichia coli* and its 146 kDa Complex With HPr Using Residual Dipolar Couplings and Small and Wide Angle X-Ray Scattering. *J Am Chem Soc* **132**, 13026 (2010).
280. Caporaletti, F. *et al.* Small-angle x-ray and neutron scattering of MexR and its complex with DNA supports a conformational selection binding model. *Biophys J* **122**, 408–418 (2023).

281. World Health Organization. Global Antimicrobial Resistance and Use Surveillance System (GLASS) Report 2022. *World Health Organization 2003–2005* (2022).
282. Paracini, N., Clifton, L. A. & Lakey, J. H. Studying the surfaces of bacteria using neutron scattering: finding new openings for antibiotics. *Biochem Soc Trans* **48**, 2139–2149 (2020).
283. Foglia, F., Lawrence, M. J. & Barlow, D. J. Studies of model biological and bio-mimetic membrane structure: Reflectivity vs diffraction, a critical comparison. *Curr Opin Colloid Interface Sci* **20**, 235–243 (2015).
284. Nilsson, I. *et al.* Molecular characterization and verification of azido-3,8-dideoxy-D-manno-oct-2-ulosonic acid incorporation into bacterial lipopolysaccharide. *Journal of Biological Chemistry* **292**, 19840–19848 (2017).
285. Saïdi, F. *et al.* Evaluation of Azido 3-Deoxy- d - Manno-oct-2-ulosonic Acid (Kdo) Analogues for Click Chemistry-Mediated Metabolic Labeling of *Myxococcus xanthus* DZ2 Lipopolysaccharide. *ACS Omega* **7**, 34997–35013 (2022).
286. Wang, Y. J. *et al.* Imaging of *Escherichia coli* K5 and glycosaminoglycan precursors via targeted metabolic labeling of capsular polysaccharides in bacteria. *Sci Adv* **9**, (2023).
287. Lostao, A., Lim, K. S., Pallarés, M. C., Ptak, A. & Marcuello, C. Recent advances in sensing the inter-biomolecular interactions at the nanoscale - A comprehensive review of AFM-based force spectroscopy. *Int J Biol Macromol* **238**, (2023).
288. Li, M., Dang, D., Liu, L., Xi, N. & Wang, Y. Imaging and Force Recognition of Single Molecular Behaviors Using Atomic Force Microscopy. *Sensors 2017, Vol. 17, Page 200* **17**, 200 (2017).
289. Naqvi, K. F. *et al.* Novel Role for Macrophage Galactose-Type Lectin-1 to Regulate Innate Immunity against *Mycobacterium tuberculosis*. *J Immunol* **207**, 221–233 (2021).
290. Vallet, A., Favier, A., Brutscher, B. & Schanda, P. ssNMRLib: a comprehensive library and tool box for acquisition of solid-state nuclear magnetic resonance experiments on Bruker spectrometers. *Magnetic resonance* **1**, 331 (2020).
291. Favier, A. & Brutscher, B. NMRLib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J Biomol NMR* **73**, 199–211 (2019).
292. Brown, L., Wolf, J. M., Prados-Rosales, R. & Casadevall, A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nature Reviews Microbiology 2015 13:10* **13**, 620–630 (2015).
293. Viegas, A., Ao Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A Simple and Fast Method for Ligand Screening and Characterization of Protein Binding. *J. Chem. Educ* **88**, 990–994 (2011).

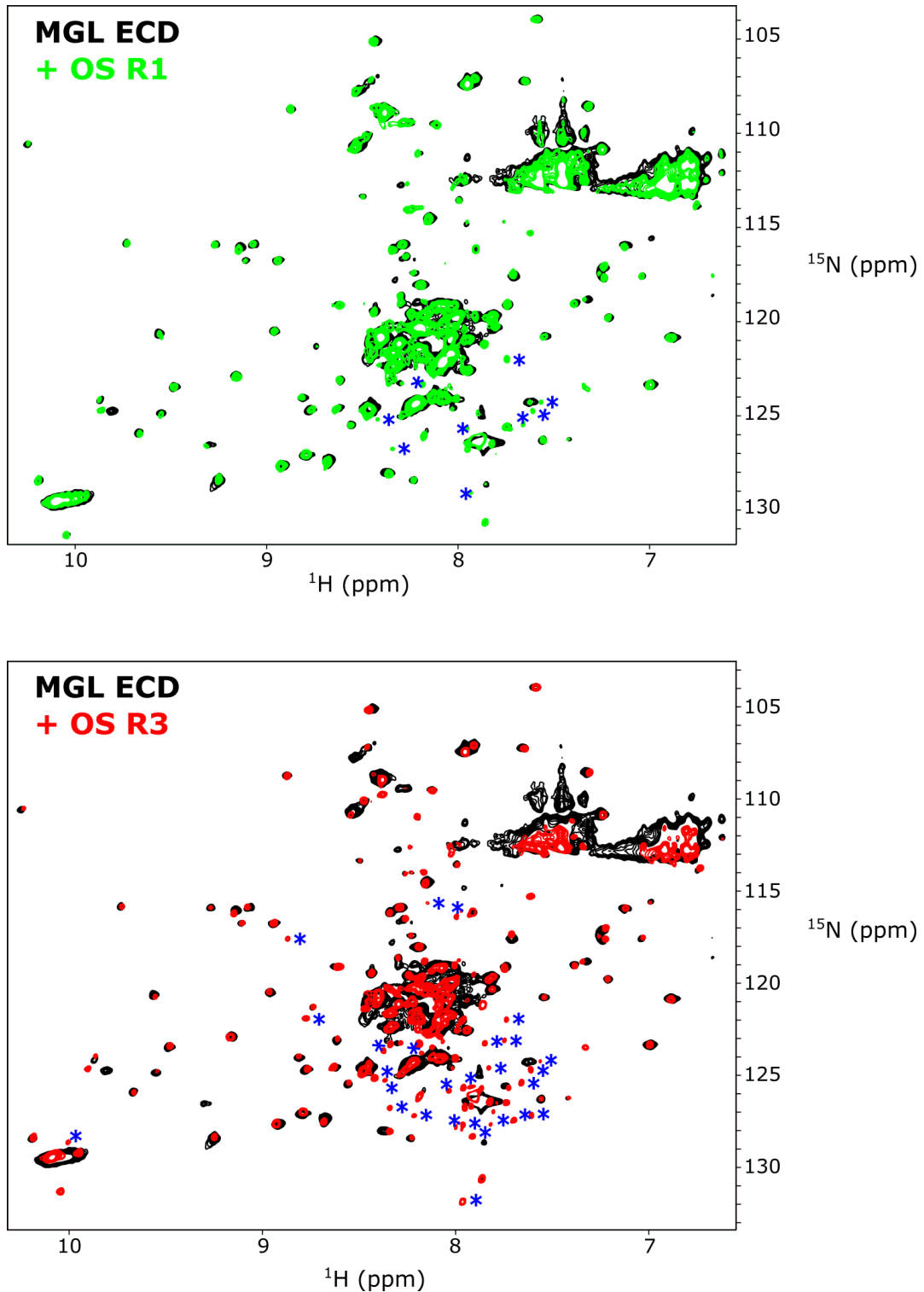
Annexes

Table of annexes

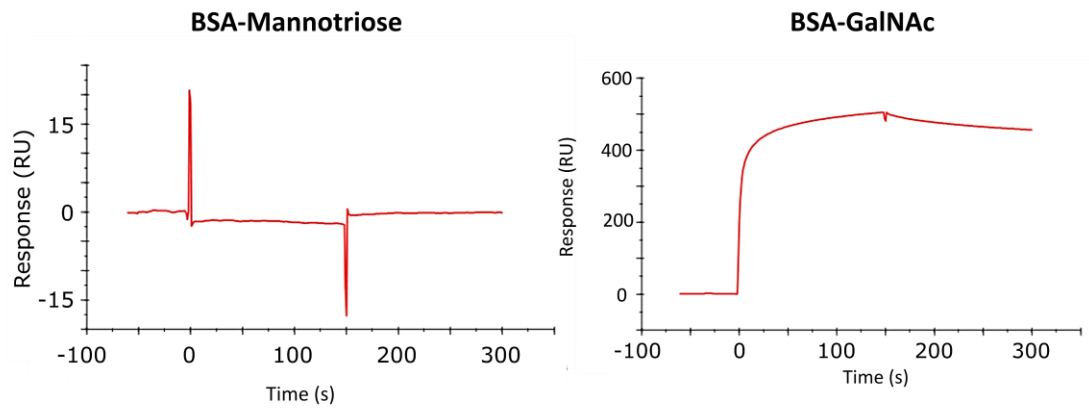
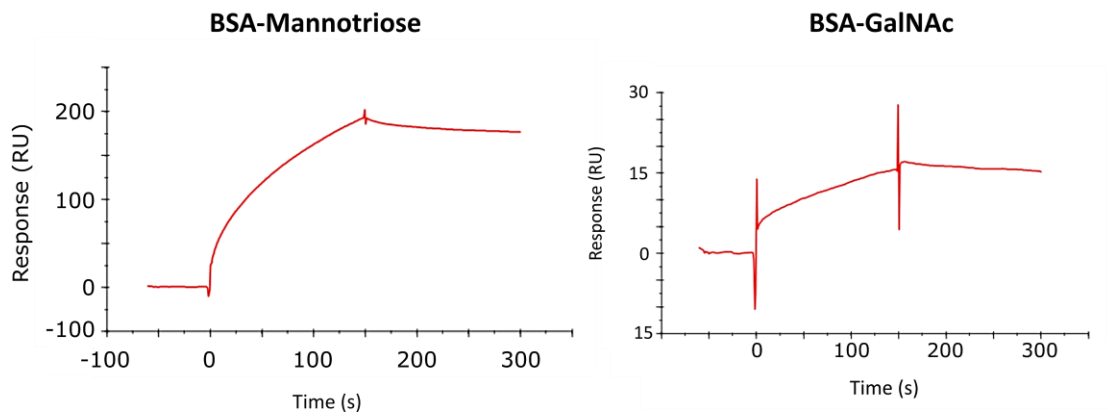
Annex 1: Purification of MGL ECD ^{D245H}	206
Annex 2: Full ¹⁵ N- ¹ H correlation spectra of MGL-ECD before and after addition of 2:1 Glycan-OS R1/R3 ratio.....	207
Annex 3: SPR-oriented surfaces functionality tests.....	208
Annex 4: Flow Cytometry of <i>E. coli</i> R1 and R3 cells without labelled MGL ECD with Alexafluor 647	209
Annex 5: Flow cytometry quantification of MGL ECD labelled with AF647 bound to R1 and R3 cells.....	210
Annex 6: Flow Cytometry of <i>E. coli</i> O113:21 and O157:H7 cells without labelled MGL ECD with Alexafluor 647	211
Annex 7: AFM Surface measurements of the different nanodiscs alone and with interaction with MGL ECD and their corresponding statistical data	212
Annex 8: Overlaid spectra of SMA R1 _{pur} with the purified delipidated R1 oligosaccharide OS	213
Annex 9: ¹ H- ¹³ C INEPT based correlation spectra of SMA O157:H7 _{pur} nanodiscs.....	214
Annex 10: ¹³ C- ¹³ C Single Quantum J-correlation experiment of SMA O157:H7 _{pur} nanodiscs	215
Annex 11: ¹ H- ¹³ C CP based correlation spectrum of SMA O157:H7 _{pur} nanodiscs	216
Annex 12: Strips of an HCC experiment of the assigned core OS residues on the H1 region recorded on LOS R3 reconstituted in DHPC micelles.....	217
Annex 13: Interactions studies of immobilized SMA R1 _{pur} with PmB by BLI.....	218
Annex 14: Interaction of immobilized SMA-R1 _{pur} nanodiscs with MGL ECD by BLI	219
Annex 15: Interaction of immobilized MGL ECD with purified R1 LOS liposomes by BLI	220



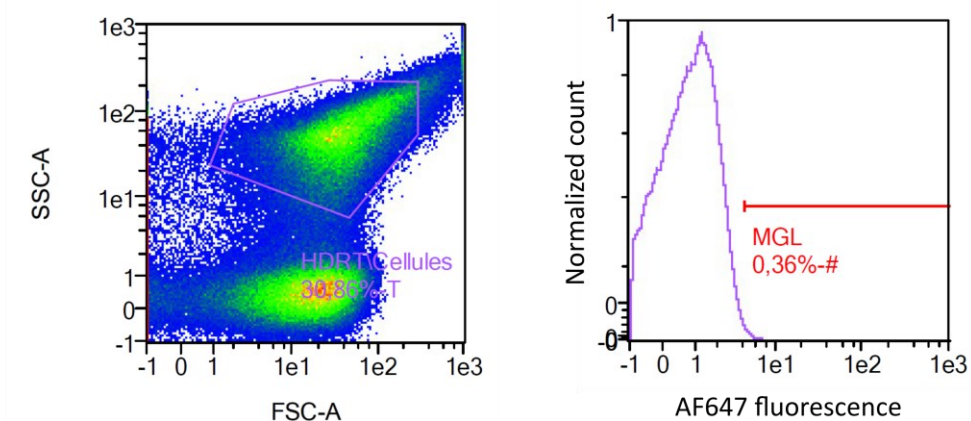
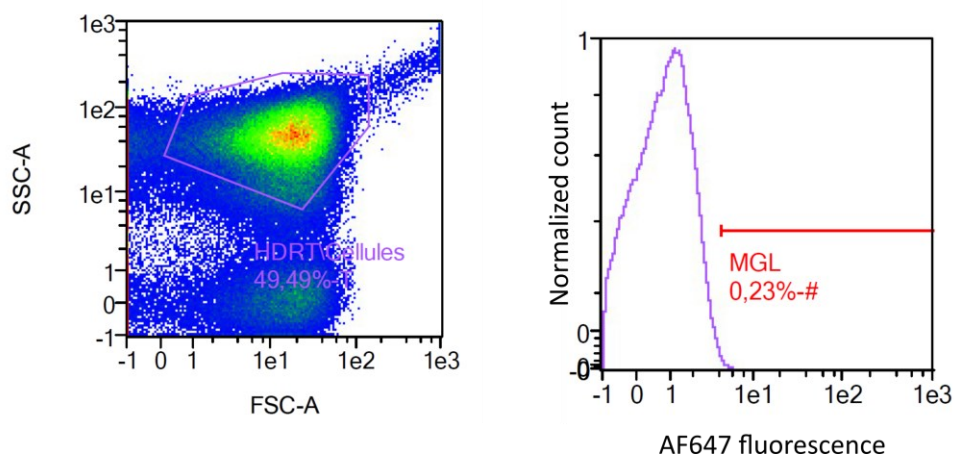
Annex 1: Purification of MGL ECD^{D245H}. (A) Full chromatogram of MGL ECD two-steps purification (left) consisting of a first StrepTag affinity chromatography followed with a size-exclusion chromatography on a Toyopearl HW-50S with a zoom on a SEC Chromatogram profile (right). (B) Representative 12% SDS-PAGE of fractions obtained following the SEC chromatography in both reduced and non-reduced conditions.



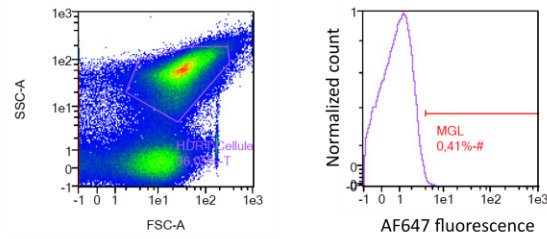
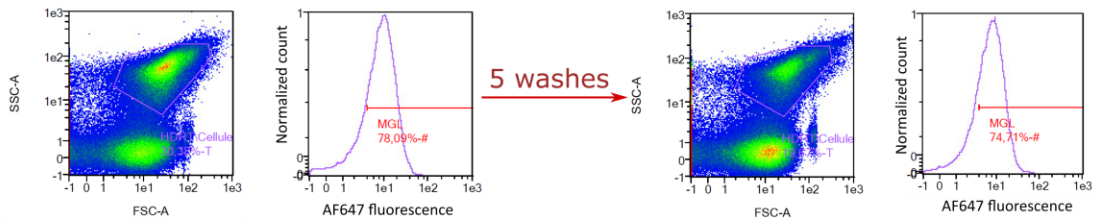
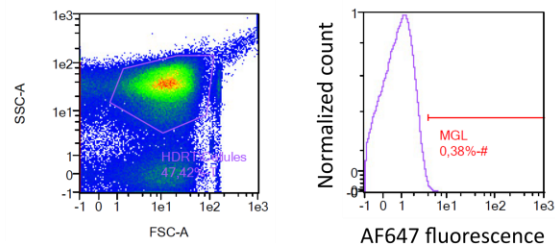
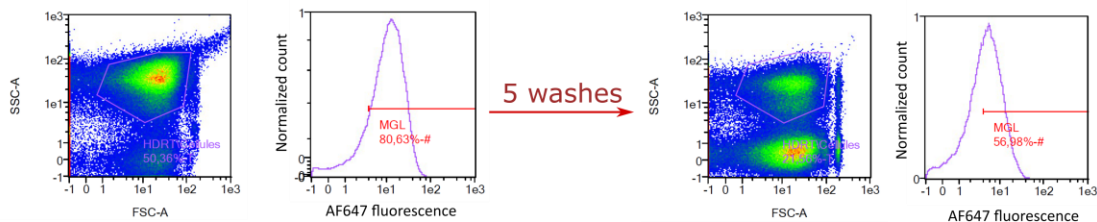
Annex 2: Full ^{15}N - ^1H correlation spectra of MGL-ECD before (black) and after addition of 2:1 Glycan-OS R1/R3 ratio. New peaks appearing from minor protein degradation over time are indicated with blue asterisks.

A MGL ECD Surface functionality test**B** DC-SIGN ECD Surface functionality test

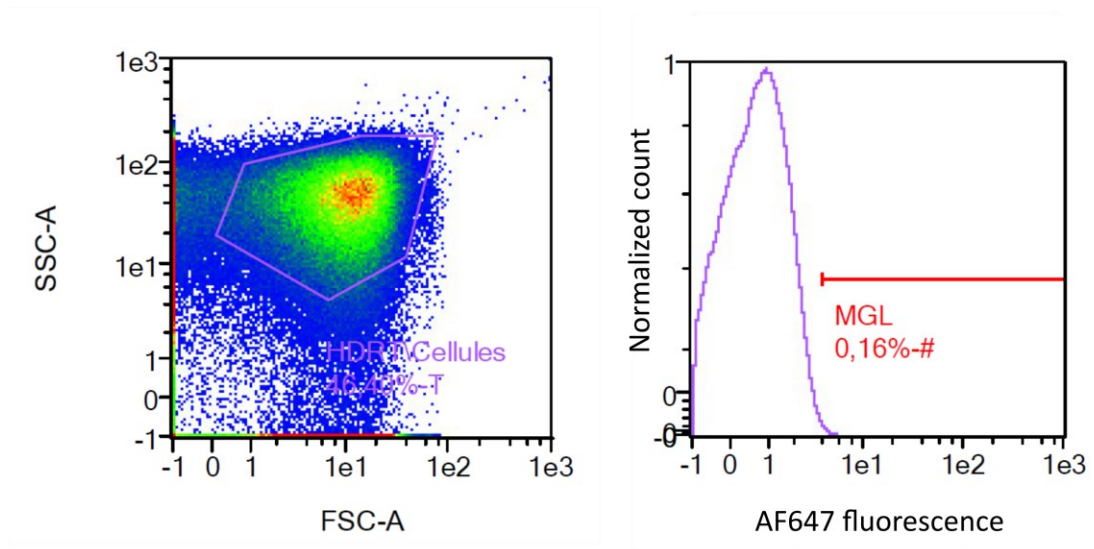
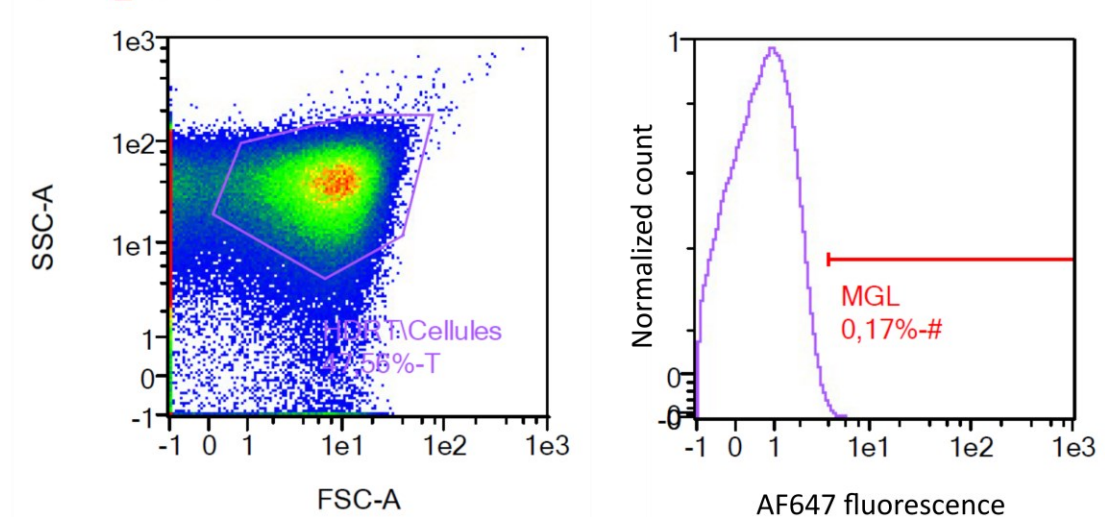
Annex 3: SPR-oriented surfaces functionality tests. (A) MGL ECD surface functionality tests with BSA-Mannotriose and BSA-GalNAc as negative and positive controls, respectively. (B) DC-SIGN ECD surface functionality tests with BSA-Mannotriose and BSA-GalNAc controls.

E. coli R1_no MGL*E. coli R3_no MGL*

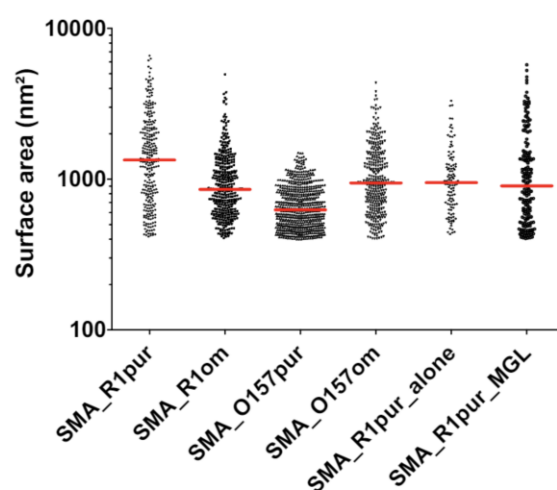
Annex 4: Flow Cytometry of *E. coli R1* and *R3* cells without labelled MGL ECD with Alexafluor 647. The corresponding flow cytometry density plots are presented together with the count vs fluorescence plot of the selected bacterial populations. Associated % of fluorescent population is indicated. FSC: Forward scattering; SSC: Side scattering; AF647: Alexafluor647.

*E. coli R1**E. coli R1_no MGL**E. coli R1_MGL**E. coli R3**E. coli R3_no MGL**E. coli R3_MGL*

Annex 5: Flow cytometry quantification of MGL ECD labelled with AF647 bound to R1 and R3 cells. Quantification of MGL binding to cells with and without washing steps, top and bottom panels, respectively. The corresponding 2D representation of Forward (FSC) and side (SSC) scattering together with the count vs fluorescence plot of the selected bacterial population are shown. Experiments were done in duplicates and only data of one of the replicates is shown.

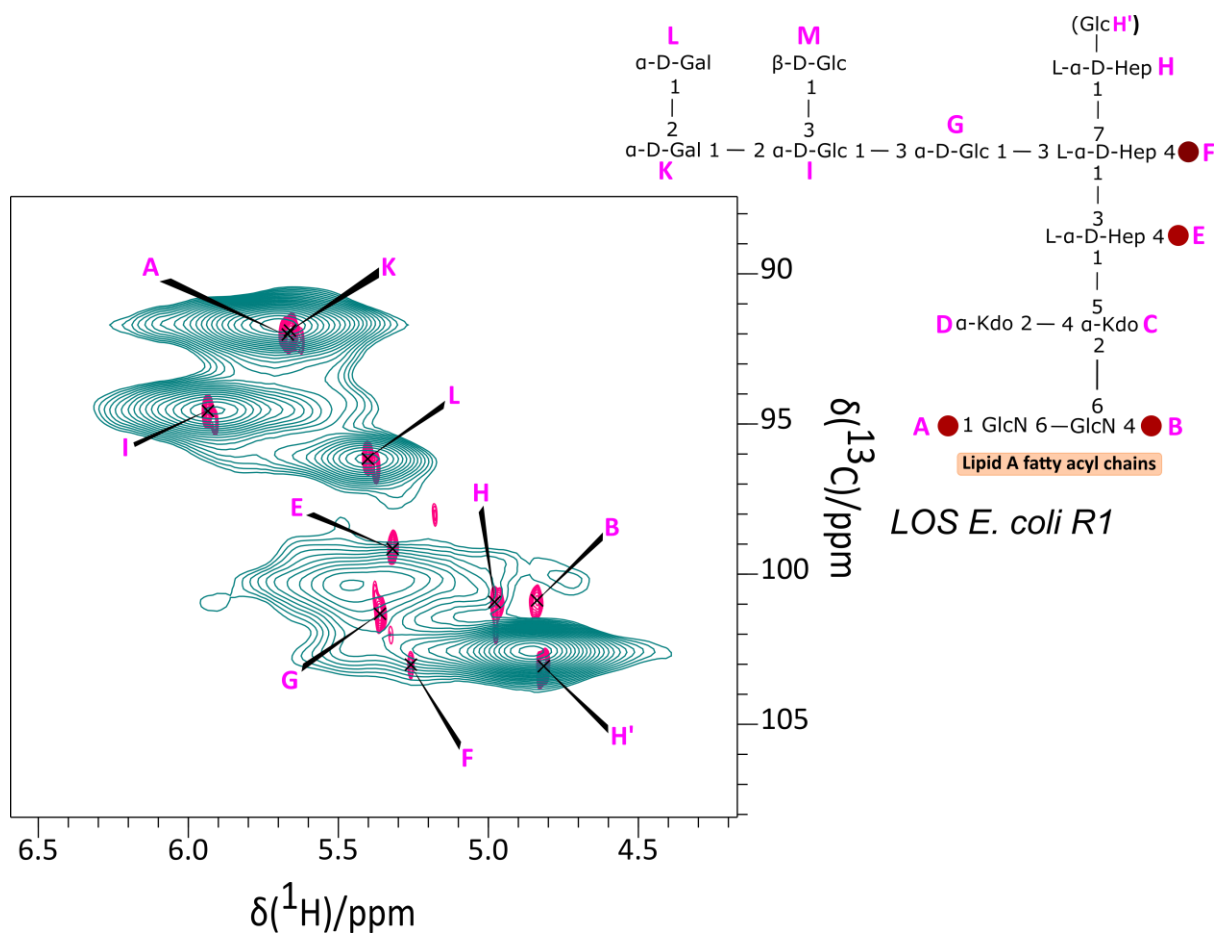
E. coli O113:H21_no MGL*E. coli* O157:H7_no MGL

Annex 6: Flow Cytometry of *E. coli* O113:H21 and O157:H7 cells without labelled MGL ECD with Alexafluor 647. The corresponding flow cytometry density plots are presented together with the count vs fluorescence plot of the selected bacterial populations. Associated % of fluorescent population is indicated. FSC: Forward scattering; SSC: Side scattering; AF647: Alexafluor647.

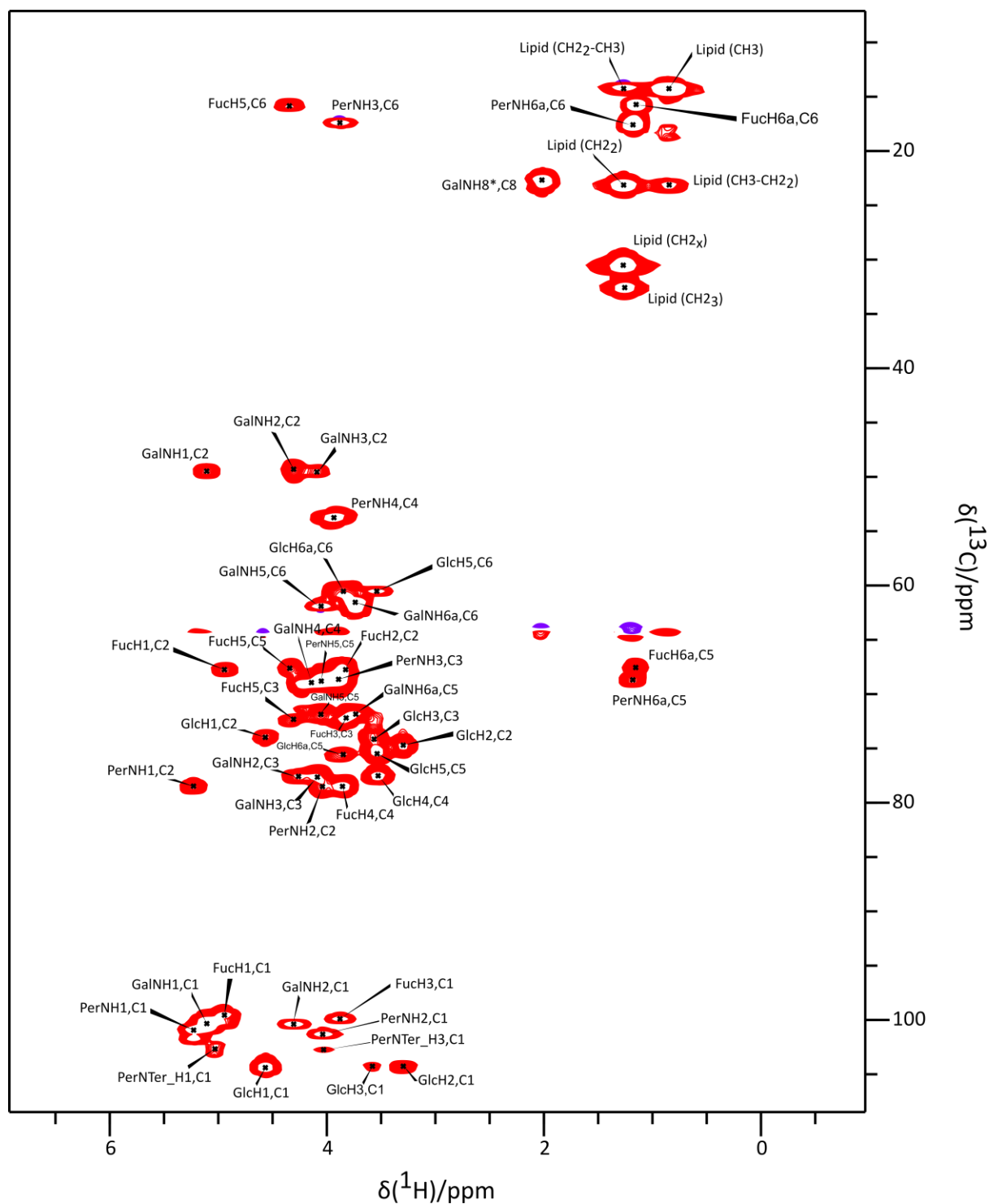


	SMA-R1 _{pur}	SMA-R1 _{OM}	SMA-O157 _{pur}	SMA-O157 _{OM}	SMA-R1 _{pur_alone}	SMA-R1 _{pur_MGL}
Number of values	278	360	656	345	108	163
Median (m²)	1.34E-15	8.55E-16	6.265E-16	9.43E-16	9.48E-16	9.02E-16
Mean (m²)	1.712E-15	1.028E-15	6.856E-16	1.124E-15	1.092E-15	1.293E-15
Std. Deviation (m²)	1.396E-15	6.045E-16	2.343E-16	6.418E-16	5.939E-16	1.048E-15
Std. Error	8.375E-17	3.186E-17	9.147E-18	3.455E-17	5.715E-17	8.207E-17

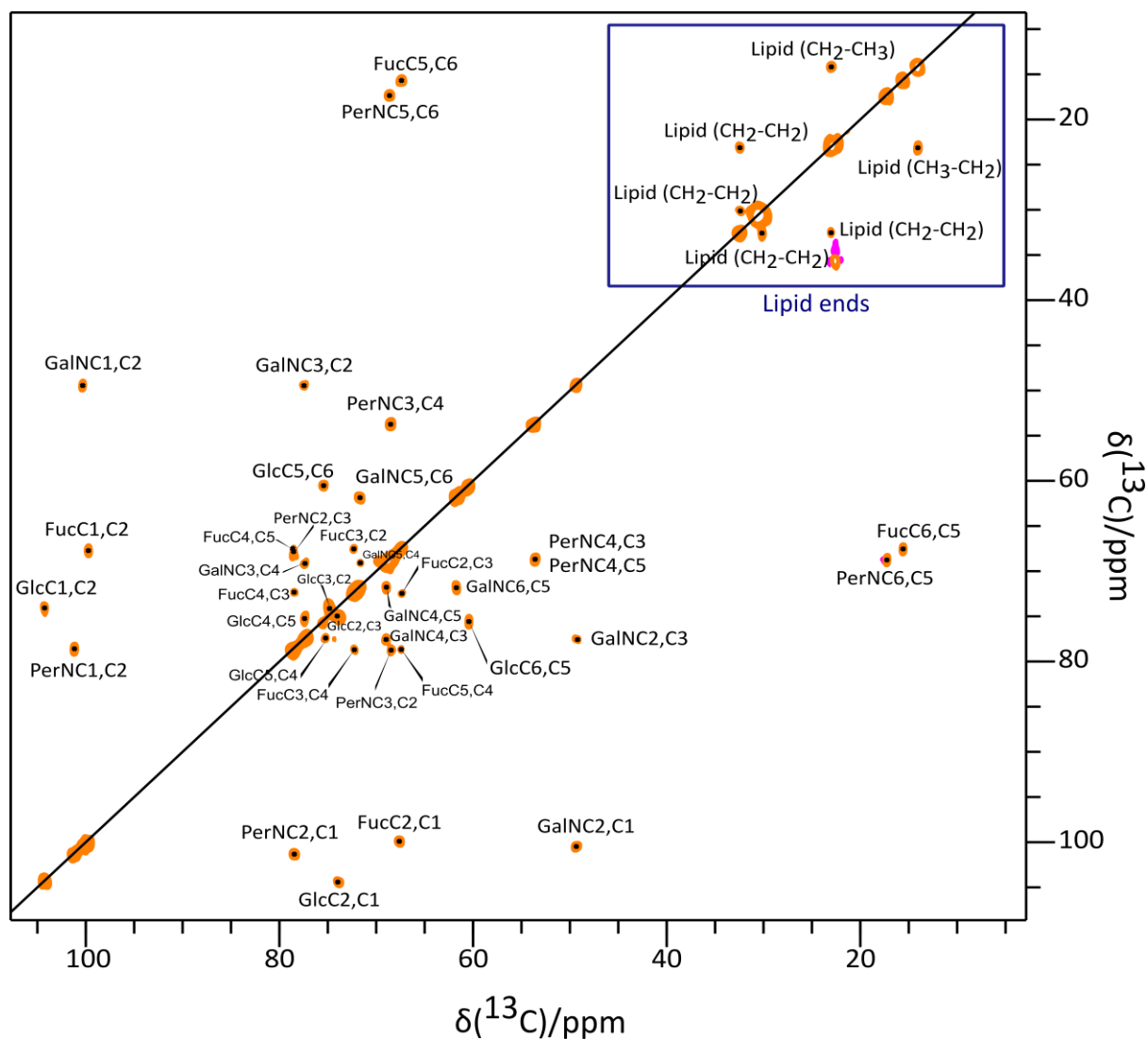
Annex 7: AFM Surface measurements of the different nanodiscs alone and with interaction with MGL ECD and their corresponding statistical data. Samples were imaged on a mica surface coated with Ni²⁺ ion. Imaging was performed in air using the PeakForce Tapping mode.



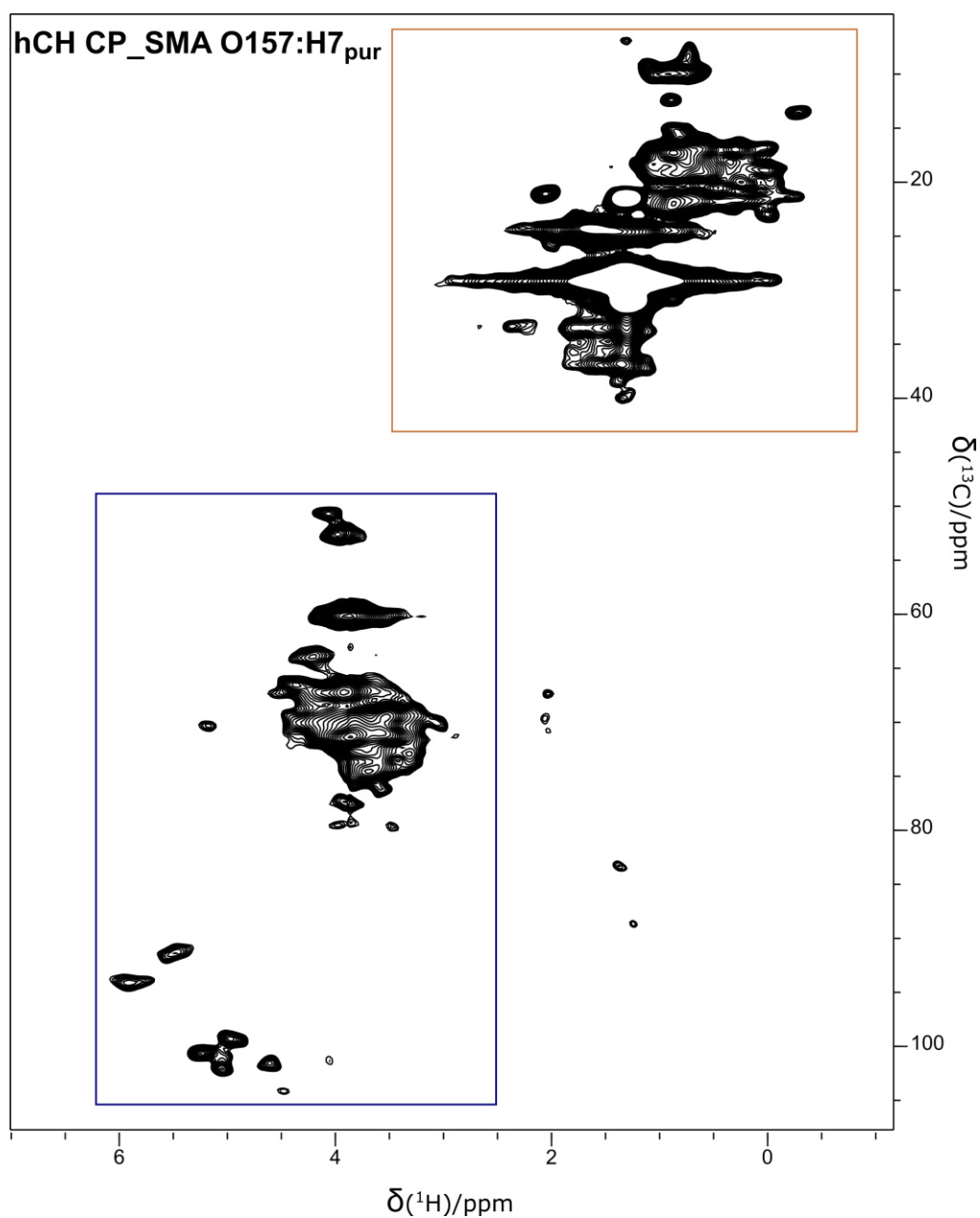
Annex 8: Overlaid spectra of SMA R1_{pur} with the purified delipidated R1 oligosaccharide OS. Zoom on the anomeric region of overlaid spectra of: SMA R1_{pur} (teal) recorded using ssNMR with the purified delipidated R1 oligosaccharide OS recorded using solution NMR (pink). The chemical structure of LOS R1 is indicated with the annotations used for the assigned residues.



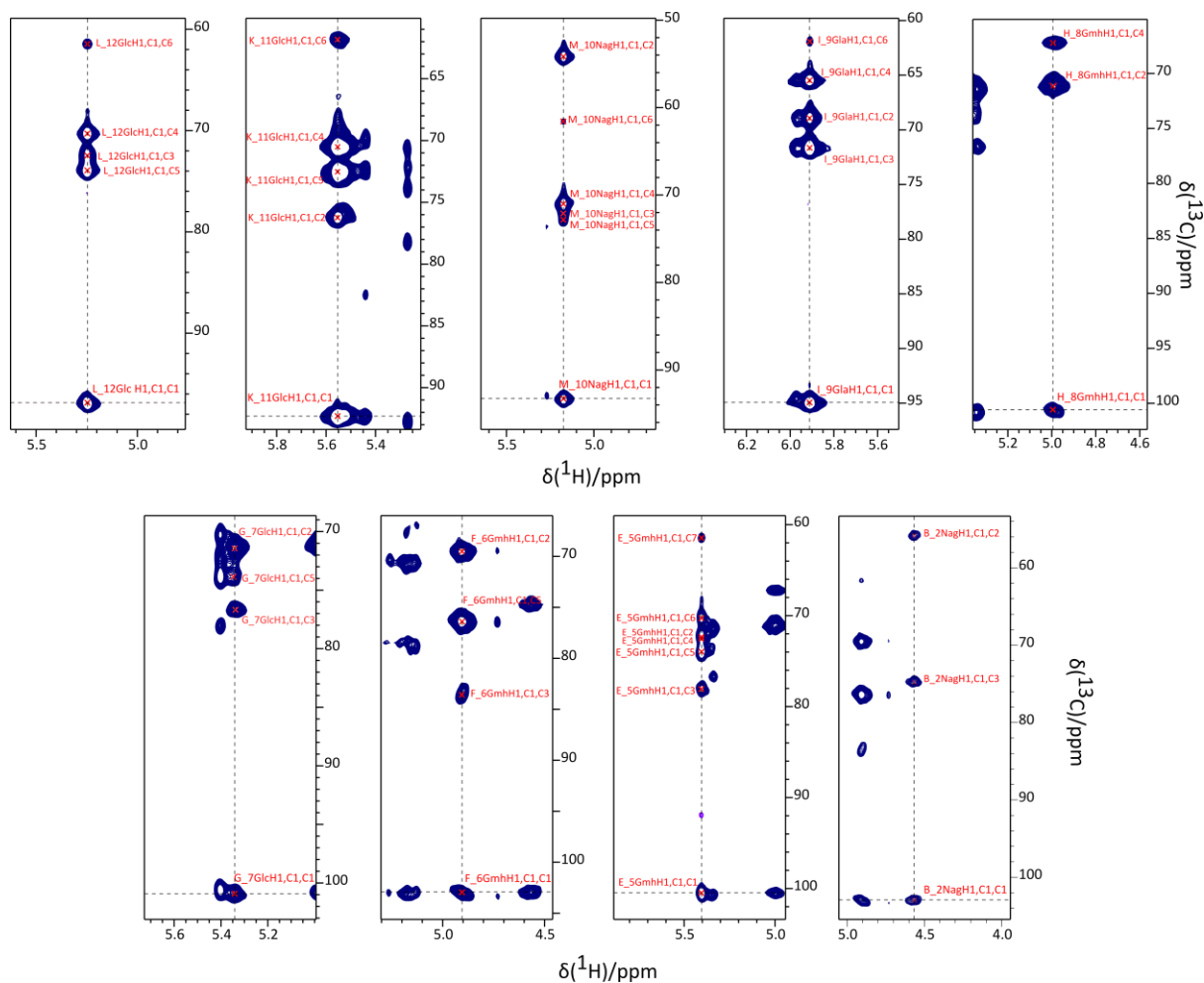
Annex 9: ^1H - ^{13}C INEPT based correlation spectra of SMA O157:H7_{pur} nanodiscs. The sample was span at 55 kHz, at a sample temperature of 50°C on a 600 MHz spectrometer equipped with MAS 1.3 mm HCN probes. The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O.



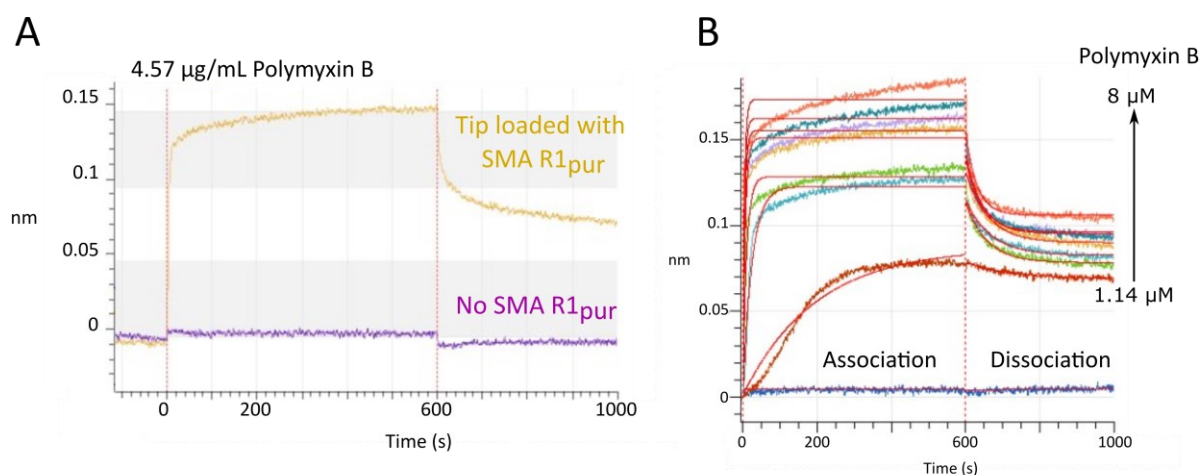
Annex 10: ^{13}C - ^{13}C Single Quantum J -correlation experiment of SMA O157:H7_{pur} nanodiscs. The sample was spun at 15 kHz, at a sample temperature of 50°C on a 600 MHz spectrometer equipped with MAS 1.3 mm HCN probe. The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D₂O.



Annex 11: ^1H - ^{13}C CP based correlation spectrum of SMA O157:H7_{pur} nanodiscs. Fatty acid signals are highlighted in orange, while sugar signals (ring and anomeric) are highlighted in blue. The sample was prepared in 25 mM Tris pH 7.4 100 mM NaCl in 100% D_2O . The NMR experiment was recorded on a 600 MHz spectrometer equipped with MAS 1.3 mm HCN probes at 55 kHz, and a sample temperature of 50°C.



Annex 12: Strips of an HCC experiment of the assigned core OS residues on the H1 region recorded on LOS R3 reconstituted in DHPC micelles. The experiment was performed on a 600MHz spectrometer equipped with a cryoprobe.



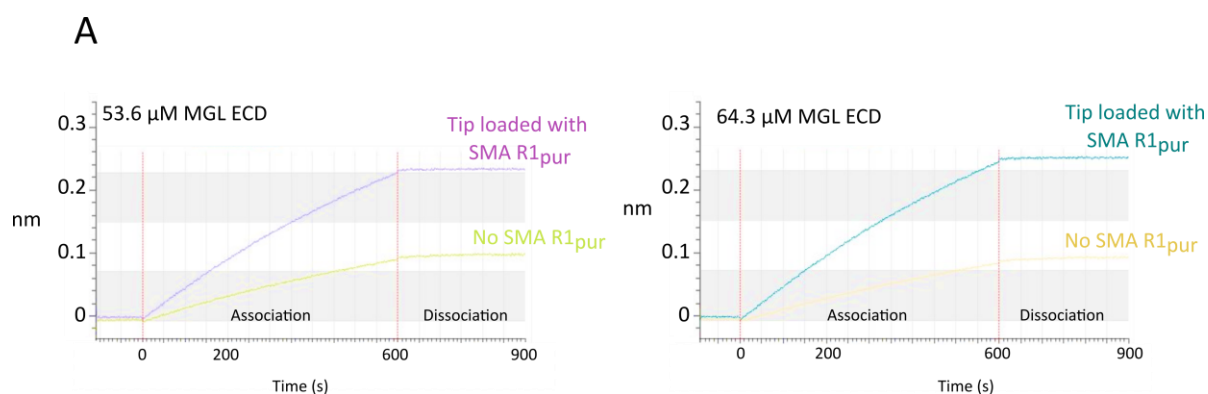
B

Polymyxin B (μM)	K_D (M)	k_a (1/Ms)	k_a Error	k_d (1/s)	K_d Error
1.14				8.22E-03	2.21E-04
2.28	1.28E-06	1.14E+04	1.97E+02	1.46E-02	1.57E-04
3.42	7.43E-07	2.30E+04	3.76E+02	1.71E-02	1.93E-04
4.57	5.83E-07	2.82E+04	4.86E+02	1.64E-02	1.97E-04
5.71	7.83E-07	2.91E+04	6.67E+02	2.28E-02	2.72E-04
6.86	7.99E-07	2.78E+04	7.03E+02	2.23E-02	2.62E-04
8	1.44E-06	2.26E+04	7.23E+02	3.27E-02	3.84E-04

C

Steady State Results	
χ^2	0.0002
R^2	0.9767
Rmax	0.2255
Rmax Error	± 0.009
K_D (M)	2.10E-06
K_D Error	$\pm 2.5E-07$

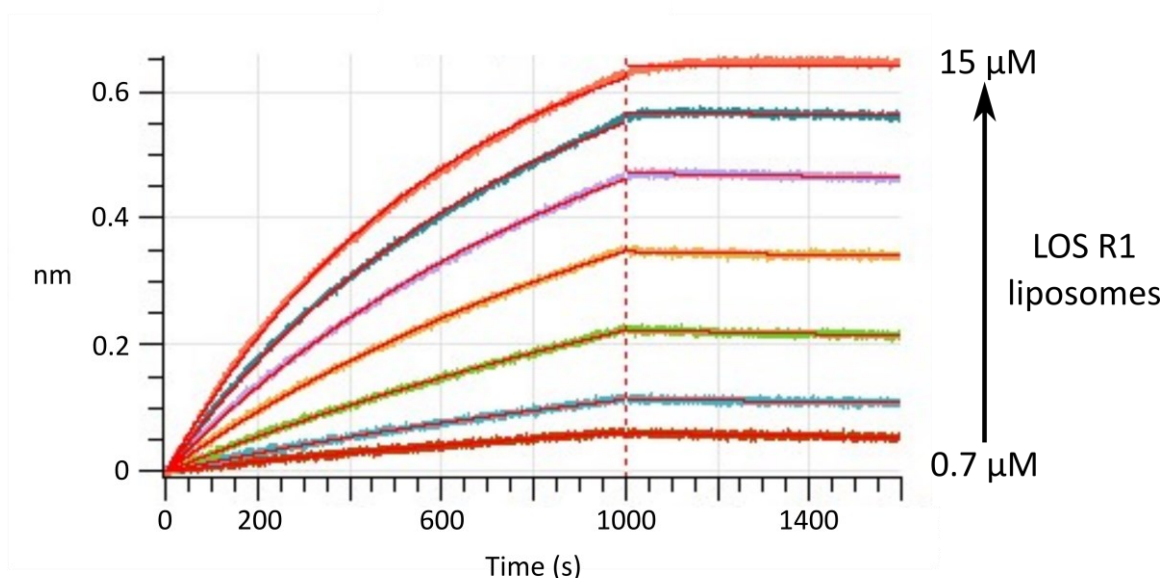
Annex 13: Interactions studies of immobilized SMA R1_{pur} with PmB by BLI. (A) Sensorgram showing the interaction response of Polymyxin B with tips loaded or not with SMA-R1_{pur} nanodiscs at a PmB concentration of 4.57 $\mu\text{g/mL}$. (B) Local fit of the association and dissociation steps following SMA R1_{pur} interaction with PmB. A table summarizing the kinetic parameters determined for some data points following this fit is presented. This includes association (k_a) and dissociation (k_d) rates with the dissociation constant (K_D). (C) Table summarizing the results of the steady state analysis. This includes the chi-squared (χ^2) value of the fitted model, the coefficient of determination (R^2), maximum binding response (Rmax) and dissociation constant (K_D).

**B**

MGL ECD [nM]	K_D (M)	k_a (1/Ms)	k_a Error	k_d (1/s)	k_d Error
10.7				5.35E-02	6.32E-03
21.4	7.02E-12	7.04E+04		4.94E-07	
32.1				2.59E-02	7.65E-03
42.9	3.18E-08	1.34E+04	6.65E+04	4.27E-04	2.85E-03
53.6	6.51E-09	2.35E+04	1.42E+04	1.53E-04	7.56E-04
64.3	2.02E-08	1.58E+04	1.31E+04	3.18E-04	8.39E-04
75	7.35E-08	6.26E+03	1.62E+04	4.60E-04	1.21E-03

K_D average	K_D standard deviation
2.64088E-08	2.90794E-08

Annex 14: Interaction of immobilized SMA-R1_{pur} nanodiscs with MGL ECD by BLI. (A) Sensorgram showing the interaction of MGL ECD with tips loaded or not with SMA-R1_{pur} nanodiscs at two MGL concentrations 53.6 μ M and 64.3. (B) Following MGL ECD binding to SMA R1_{pur} tips at different concentrations at 25°C, a local fit of the association and dissociation steps was performed, and their respective association (k_a) and dissociation (k_d) rates were estimated from the fit. An apparent dissociation constant (K_{Dapp}) is calculated by averaging the obtained dissociation constants (K_D) from the fit and their standard deviation (bottom).



$$K_{Dapp} = 180 \text{ nM } \pm 60$$

LOSR1 (μM)	K_D (M)	k_a (1/Ms)	k_a Error	k_d (1/s)	k_d Error
1,17	1,87E-07	4,88E+02	8,19E+00	9,11E-05	2,68E-06
1,94	2,78E-07	2,52E+02	2,51E+00	6,99E-05	1,39E-06
3,2	1,40E-07	2,51E+02	1,04E+00	3,52E-05	8,97E-07
5,4	1,47E-07	1,86E+02	5,09E-01	2,72E-05	6,67E-07

Annex 15: Interaction of immobilized MGL ECD with purified R1 LOS liposomes by BLO. MGL was immobilized on streptavidin sensors then immersed in wells containing R1 LOS liposomes at increasing concentrations ranging from 0.7 to 15 μM for 1000 s (association). The tips were next immersed in buffer for 600 s (dissociation). A local fit of the association and dissociation steps was performed, and their respective association (k_a) and dissociation (k_d) rates were estimated from the fit. An apparent dissociation constant (K_{Dapp}) is calculated by averaging the obtained dissociation constants (K_D) from the local fit and their standard deviation.