UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Química Orgánica y Farmacéutica

TESIS DOCTORAL

TLR4 modulation: molecular recognition studies and drug design by molecular modelling

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Jean-Marc Billod

Directora

Sonsoles Martín-Santamaría

Madrid, 2019

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA ORGÁNICA Y FARMACÉUTICA

TESIS DOCTORAL

TLR4 MODULATION: MOLECULAR RECOGNITION STUDIES AND DRUG DESIGN BY MOLECULAR MODELLING

MODULACIÓN DEL TLR4: ESTUDIOS DE RECONOCIMIENTO MOLECULAR Y DISEÑO DE FÁRMACOS POR MODELADO MOLECULAR

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

Jean-Marc Billod

DIRECTORA
Sonsoles Martín-Santamaría

Madrid, 2018

Acknowledgements

To me, realizing this thesis marks the completion of a journey that began many years ago. I am so grateful for all the people that have helped me, sometimes without knowing it, to fulfill my goal. I would like to acknowledge all of those people here.

Five years ago, Bruno Cardey helped me perform my first-year master's research project abroad. I traveled to Trømso, a city in northern Norway, where I worked with Prof. Clemens Woywod. Because of Bruno and Clemens, I learned a great deal during my time there.

A year later, motivated by my previous experience in Scandinavia I moved to Gothenburg in Sweden for my master's thesis. There, I was welcomed by Prof. Leif Eriksson who has been a great mentor. I would like to thank him for the tremendous amount of support he has given me. His attitude toward science inspired me then, and it still inspires me now.

Another year later, I moved to Madrid, Spain, where I joined the lab of Prof. Sonsoles Martín-Santamaría. I spent the next three years studying and doing research, but I also traveled extensively, discovered many new things, and met countless new incredible people. I am immensely grateful for the entire experience. I would like to thank Sonsoles for allowing me to work in her lab with a Marie Curie Early Stage Researcher fellowship, and for her excellent scientific and personal guidance throughout my Ph.D. thesis. I also would like to thank her for the numerous opportunities she gave me to meet and interact with fellow researchers, to learn new things every day, and to travel and work at different laboratories across Europe.

Regarding these work travels, I would like to thank all of those people who warmly welcomed me into their labs: Prof. Jesus Jimenez-Barbero and Helena Coelho in Derio, Spain. Dr. Alba Silipo and Mateusz Pallach in Naples, Italy. Prof. S.J. Marrink and Dr. Paulo Telles de Souza in Groningen, Netherlands. Dr. Manuel Fresno, Dr. Miguel Angel Llamas Matías and Sara Isabel Vaz Francisco in Madrid, Spain.

I would like to thank my colleagues for their constant presence and support during these three years: Alessandra, Laura, Lucía, Javi, Joan, and all of the students who came to the lab.

Next, I would like to acknowledge the European Union and the European community for the Marie Sklodowska-Curie grant, and for continuously pushing forward the existence of a society in which a part of the population can dedicate their time to doing fundamental research. Although the benefits are not always immediate, society will be rewarded in the future.

I would like to thank my friends in France. They were always within reach despite the physical distance.

Additionally, I would like to thank all of the people I have met in various countries through the past five years that have impacted my life in a positive way.

Finally, and most importantly, I would like to thank my parents for always supporting my scientific aspirations. From helping me build a lab in the attic when I was a child to supporting my decisions to work abroad, first in Norway, then in Sweden, and again in Spain. They have always done everything they can for me, as well as my sister and my five brothers, whom I also want to acknowledge here.

Without all these people I would not have accomplished any of these things or been able to write this thesis.

Table of contents

Table of abbreviations

The following abbreviations are used in this thesis.

AGPs	Aminoalkyl Glucosaminide 4-Phosphates
AMPs	Antimicrobial Peptides
APBS	Adaptive Poisson-Boltzmann Solver
Ara4N	4-amino-4-deoxy-arabinose
CD14	Cluster of Differentiation 14
CPK	Corey-Pauling-Koltun
DAMPs	Damage-Associated Molecular Patterns
DC	Dendritic Cell
ED	Ectodomain
EtN	2-aminoethanol group
FA	Fatty Acid
GLA	Gifu Lipid As
GlcN	Glucosamine
HEK293	Human embryonic kidney cells 293
HM	Homology Modeling
HMGB1	High Mobility Group Box 1
h/m/eqMD-2	human /mouse/equine MD-2
h/m/eqTLR4	human /mouse/equine TLR4
hPBMCs	human Peripheral Blood Mononuclear Cells
ID	Intracellular Domain
IL	Interleukin
INF	Interferon
IRF-3	Interferon Regulatory Factor 3
KML-C	Korean Mistletoe Lectin
LBP	Lipopolysaccharide-Binding Protein
Ld	Liquid-ordered
Lo	Liquid-disordered
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
Mal	MyD88-adapter-like
mCD14	mouse CD14
mWBCs	murine White Blood Cells
MD simulations	Molecular Dynamics simulations
MD-2	Myeloid Differentiation factor 2
MD-2*	partner MD-2
MyD88	Myeloid Differentiation factor 88
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
OM	Outer Membrane
PAMPs	Pathogen-Associated Molecular Patterns
PRR	Pattern Recognition Receptor
PTX	Paclitaxel
PDB	Protein Data Bank
R-LPS	Rough-type LPS

RMSD	Root-Mean-Square Deviation
RMSF	Root-Mean-Square Fluctuation
S-LPS	Smooth-type LPS
SAR	Structure-activity relationship
SASA	Solvent-Accessible Surface Area
TD	Transmembrane Domain
TICAM-1	TIR-domain-Containing Adapter Molecule-1
TICAM-2	TIR-domain-Containing Adapter Molecule-2
TIR	Toll/Interleukin-1 Receptor
TIRAP	TIR-domain-containing Adapter Protein
TIRP	TIR-containing Protein
TLR	Toll-Like-Receptor
TLR4*	partner TLR4
TNF-α	Tumor Necrosis Factor alpha
TRAM	TRIF-Related Adapter Molecule
TRIF	TIR-domain-containing adapter-inducing Interferon-β
VS	Virtual Screening
WT	Wild Type

Abstract

Introduction

The heterodimeric complex, formed by Toll-Like Receptor 4 (TLR4) and its accessory protein Myeloid Differentiation factor 2 (MD-2) is responsible of activating the innate immune system when sensing the presence of particular pathogen-associated molecular patterns (PAMPs) from bacteria. The outer membrane of Gram-negative bacteria is primarily populated by lipopolysaccharides (LPS) which are essential for their growth and survival. These LPSs are specifically recognized by the TLR4/MD-2 complex as follows: an LPS binds to MD-2 inside a deep molecular hydrophobic pocket causing molecular rearrangements of the receptorial complex resulting in the dimerization of another TLR4/MD-2 unit. TLR4 ectodomains dimerization event brings together the TLR4 intercellular domains initiating the activation of innate immune system signaling pathways. Interestingly, this activation is not only modulated by naturally occurring LPSs from many different Gram-negative bacteria but also by non-naturally occurring glycolipids and other non-LPS like molecules.

Objectives

TLR4 attracted lots of attention for the finding of new modulators with important applications in biomedicine. Several new compounds modulating TLR4 are undergoing preclinical and clinical evaluation, for the treatment of sepsis, inflammatory diseases, rheumatoid arthritis, and as vaccines and cancer immunotherapeutic agents. However, a TLR4 modulator to effectively treat septic shock is yet to be discovered and approved for commercialization. Also, TLR4 agonists are sought to develop co-adjuvants for antitumoral treatments. Elucidating the molecular determinants that make a given molecule to be an agonist or an antagonist of TLR4, and understanding the mechanism of the TLR4/MD-2 system, would greatly help the design of new TLR4 modulators.

The relatively recent elucidation of the X-ray crystallographic structure of the extracellular domain of TLR4 in complex with MD-2 has opened new perspectives for the research around this challenging receptor. Towards that end, this thesis can be divided into three major parts.

The first one is to assist, by computational techniques, the design of synthetic LPS-like and non LPS-like TLR4 modulators by fine-tuning their relative agonist or antagonist

potency through subtle molecular changes. An important effort is made to predict their effect on TLR4 and to assess their mode of action.

A second part is to computationally explain the effect of natural compounds (LPSs) by understanding how they interact selectively with some component of the TLR4 activation pathways, most relevantly MD-2 in complex with TLR4, MD-2 alone, and CD14. To unravel atomic details about the molecular recognition mechanism of the receptor and about the ligand-receptor interactions of these natural modulators by applying molecular modeling and computational chemistry techniques.

A third part is dedicated to gaining a deeper understanding on the molecular aspects of TLR4 activation and signaling by computational approaches. To clarify how minute molecular rearrangements on the ectodomain of TLR4 in complex with MD-2 is translated into intracellular signaling. This part includes the transmembrane domain of TLR4 and its intracellular domain as they both play an important role in the signal transmission. Apart from the fundamental knowledge they provide, these findings can guide the future development of novel agonists and antagonists of the TLR4/MD-2 system with promising biomedical applications in sepsis, inflammation, vaccines and cancer immunotherapy, among others.

Results and conclusions

Regarding naturally occurring modulators, we studied the LPS from *Bradyrhizobium* species. Rhizobia are Gram-negative bacteria able to establish symbiotic relationship with legumes and to reduce atmospheric nitrogen into ammonium, thus providing nitrogen nutrition for the host plants. Bacteria belonging to the *Bradyrhizobium* genus promote nitrogen-fixing nodules development on roots and stems of both wild-growing and cultivated *Aeschynomene* legumes. It was previously demonstrated that the lipopolysaccharide (LPS) macromolecule in Rhizobia plays a key role throughout the symbiotic process and that its structural features are altered in response to plant signals. Different lipid A structures from *Bradyrhizobium* were recently elucidated. They are highly heterogeneous regarding the number, length and nature of their acyl chains. Some contained very long-chains fatty acids and, more surprisingly, a covalently linked hopanoid molecule. That novelty prompted us to evaluate the activity these *Bradyrhizobium* lipid As may have on the innate immune system. Experimental studies, including cell assays on both murine and human bone marrow-derived macrophages and

HEK 293-TLR4/MD-2/CD14 cells, revealed an extremely low capability to elicit an immune response. More intriguingly, a potent antagonistic activity towards the toxic *E. coli* LPS was observed. Our computational studies allowed the proposal of plausible binding modes of two of these *Bradyrhizobium* lipid As to the TLR4/MD-2 system. These binding modes account for the potent activity antagonizing the binding of *E. coli* LPS to the MD-2/TLR4 complex thus inhibiting its toxic effects. It is likely that the TLR4 signaling modulation occurs by direct interaction with the TLR4/MD-2 complex, both in its hopanoid-containing and hopanoid-free forms. Our studies do not point toward a primary role of the hopanoid moiety in the biological activity regarding TLR4 signaling.

With respect to LPS-like synthesis modulators, we studied a group of glucosamine derivative. FP7, a glucosamine derivative with two phosphate groups and two myristic (C_{14}) FA chains, is active in inhibiting in a dose-dependent way human and murine TLR4 activation by LPS. NMR experiments suggest that FP7 interact with MD-2, probably inserting its FA chains into hydrophobic binding cavity. We designed new TLR4 modulators, based on FP7, and performed structure-activity relationship (SAR) studies to understand how their FA chains length determine their potency as TLR4 modulators. These FP7 variants differ only in FA chains lengths (10, 12, 14 and 16 carbon atoms). In this study we took into account both the interaction with MD-2 and the aggregation properties of the molecules. We reported structural and functional biological data demonstrating the ability of novel FP variants to negatively regulate TLR4 signaling in different cell model systems. Our computational studies were relevant in the context of the SAR study and to propose the rationale for the mechanism of binding. Our models suggest that there is an optimum length for the FA chains for an appropriate TLR4 antagonist activity related to the binding mode and to the physical-chemical properties of the FP variants.

On the subject of non LPS-like modulators we studied amphiphilic guanidinocalixarenes. To block abnormal TLR4 signaling in bacterial sepsis, two different strategies have been developed. The first one is based on LPS neutralization by the formation of noncovalent adducts with cationic compounds thus preventing LPS from interacting with the receptors. The second strategy is based on the use of molecules that compete with endotoxic LPS in binding to the same site on CD14 and

MD-2, thereby inhibiting the induction of signal transduction by impairing LPS-initiated receptor dimerization. Among the amphiphilic guanidinocalixarenes studied, we included one whose activity in this biological context had previously been reported as reference compound. Its biological activity was associated with its capacity to bind and neutralize LPS as topomimetic of LPS-binding peptides. Our computational studies challenged this view. We hypothesized that calixarene-based facial amphiphiles could also be suitable as scaffolds to obtain TLR4 ligands with antagonist activity. In a biological context, amphiphilic calixarenes showed remarkable properties significantly related to their amphiphilicity. Since we hypothesized that calixarene derivatives could directly bind to human and murine MD-2 and CD14 in a similar fashion than LPS, we preliminarily performed docking calculations to support this mode of interaction. In addition, we studied whereas the TLR4 antagonist activity is a rather general property of positively charged amphiphilic calixarenes and if this antagonist effect also derives from the direct interaction of calixarenes with the receptors and not exclusively from LPS neutralizing action, as it was suggested. Experimental evidences showed that some of these calixarenes were active in inhibiting, in a dose-dependent way, the LPS-stimulated TLR4 activation and TLR4-dependent cytokine production in human and mouse cells. Moreover, guanidinocalixarenes also inhibited TLR4 signaling when TLR4 was activated by a non-LPS stimulus, the plant lectin PHA. These results point at the calixarene moiety as a potential scaffold for the development of new TLR4-directed therapeutics.

As for the activation of TLR4, computational studies of the different domains composing the TLR4 were undertaken aiming at uncovering details of the mechanism of activation of the receptor. Understanding, at the atomic scale, the dimerization of both the transmembrane domain and the intracellular domain of TLR4 permitted to favor certain binding modes and specific secondary structures, increasing the knowledge available regarding the activation.

Resumen

Introducción

El complejo heterodimérico, formado por el receptor *Toll-like* 4 (*Toll-like receptor 4*, TLR4) y su proteína accesoria, el *Myeloid Differentiation factor 2* (MD-2), es responsable de activar la respuesta del sistema inmune innato cuando detecta la presencia de patrones moleculares asociados a patógenos (*pathogen associated molecular patterns*, PAMPs), que provienen de bacterias y virus. En concreto, la membrana externa de bacterias Gram-negativas está poblada principalmente por lipopolisacáridos (*lipopolisaccharides*, LPS), compuestos que son esenciales para su crecimiento y supervivencia. Estos LPS son reconocidos de forma específica por el complejo TLR4/MD-2 de la siguiente manera: una molécula de LPS se une a la proteína MD-2 dentro de un profundo bolsillo hidrofóbico dando lugar al reordenamiento molecular del complejo resultando en la dimerización de otra unidad de TLR4/MD-2. El evento de dimerización de los ectodominios del TLR4 hace que se acercan los dominios intracelulares que inician la activación de las vías de señalización del sistema inmune innato. Curiosamente, esta activación no sólo está modulada por LPS naturales de muchas bacterias Gram-negativas distintas, sino también por glicolípidos no naturales y otras moléculas de estructura química diferente a los LPS.

Objetivos

El TLR4 es una diana terapéutica de gran interés para la búsqueda de nuevos moduladores con aplicaciones importantes en biomedicina. Varios compuestos moduladores del TLR4 se encuentran en proceso de evaluación preclínica y clínica, para el tratamiento de sepsis, enfermedades inflamatorias, artritis reumatoide, y también como vacunas y agentes inmunoterapéuticos contra el cáncer. Sin embargo, hasta el momento no hay ningún modulador del TLR4 aprobado para tratar eficazmente el *shock* séptico. Además, se buscan agonistas del TLR4 para desarrollar coadyuvantes de tratamientos antitumorales. Elucidar los factores moleculares que hacen que una molécula determinada sea un agonista o un antagonista del TLR4, y comprender el mecanismo del sistema TLR4/MD-2, sería de gran ayuda para el diseño de nuevos moduladores del TLR4.

La elucidación, relativamente reciente, de la estructura tridimensional del dominio extracelular del TLR4 en complejo con MD-2 por cristalografía de rayos X ha abierto nuevas perspectivas para la investigación en torno a este receptor. Con ese fin, esta tesis se puede dividir en tres partes principales.

La primera es llevar a cabo, mediante técnicas computacionales, el diseño de moduladores sintéticos del TLR4 con estructura, bien similar a los LPSs por un lado, y también con otro tipo de estructuras químicas, ajustando sus propiedades como agonista o antagonista por medio de la modificación molecular. Se proporcionan los estudios computacionales realizados para predecir su efecto sobre el TLR4 y proponer su mecanismo de acción.

En una segunda parte, se trata de explicar la actividad de compuestos naturales (LPS) por medio de estudios del modo de unión a algunas de las proteínas implicadas en las vías de activación del TLR4: MD-2 en complejo con TLR4, la proteína MD-2 sola, y la proteína CD14. El objetivo es desentrañar los detalles atómicos del mecanismo de reconocimiento molecular de estos receptores y de las interacciones ligando-receptor de estos moduladores naturales mediante la aplicación de técnicas de modelado molecular y química computacional.

La tercera parte está dedicada a obtener una comprensión más profunda de los aspectos moleculares de la activación y señalización del TLR4 mediante enfoques computacionales. Hemos analizado cómo pequeños reordenamientos moleculares en el ectodominio del TLR4 en complejo con MD-2 se traducen en señalización intracelular. Esta parte incluye el estudio del dominio de transmembrana del TLR4 y su dominio intracelular, ya que ambos son importantes en la transmisión de la señal. Además del conocimiento fundamental que proporcionan, estos descubrimientos pueden guiar el desarrollo futuro de nuevos agonistas y antagonistas del sistema TLR4/MD-2 con aplicaciones biomédicas prometedoras en sepsis, inflamación, vacunas e inmunoterapia contra el cáncer, entre otras.

Resultados y conclusiones

Respecto a los moduladores naturales, hemos estudiado los LPS de las especies de *Bradyrhizobium*. Los rizobios son bacterias Gram-negativas capaces de establecer una relación simbiótica con las leguminosas y reducir el nitrógeno atmosférico generando

amonio, proporcionando así nutrición nitrogenada a la planta huésped. Las bacterias pertenecientes al género *Bradyrhizobium* promueven el desarrollo de nódulos fijadores de nitrógeno en las raíces y tallos de las leguminosas de *Aeschynomene*, tanto de crecimiento silvestre como cultivado. Se ha demostadro que el LPS de *Rhizobia* juega un papel clave en todo el proceso simbiótico y que sus características estructurales se alteran en respuesta a las señales de la planta. Recientemente, se ha elucidado la estructura de diferentes lípidos A de *Bradyrhizobium*, mostrando ser muy heterogéneos con respecto al número, la longitud y la naturaleza de sus cadenas de acilo. Algunos contenían ácidos grasos de cadena muy larga y, más sorprendentemente, una molécula de hopanoide unida covalentemente. Esa novedad nos llevó a evaluar la actividad que estos lípidos de *Bradyrhizobium* pueden tener en el sistema inmune innato. Los estudios experimentales, que incluyen ensayos celulares en macrófagos derivados de médula ósea de ratón y humanos, y células HEK 293-TLR4/MD-2/CD14, revelaron una capacidad extremadamente baja para provocar una respuesta inmune. Más curiosamente, se observó una actividad potente como antagonista del LPS tóxico de *E. coli*. Nuestros estudios computacionales permitieron la propuesta de modos de unión plausibles de dos de estos lípidos A de *Bradyrhizobium* hacia el sistema TLR4/MD-2. Estos modos de unión pueden explicar la potente actividad como antagonistas de la unión del LPS de *E. coli* al complejo MD-2/TLR4, inhibiendo así sus efectos tóxicos. Es probable que la modulación de la señalización del TLR4 se produzca por interacción directa de los lípidos A de *Bradyrhizobium* con el complejo TLR4/MD-2, tanto en su forma hopanoidea como sin hopanoide. Además, nuestros estudios indican que el anillo de hopanoide no juega un papel primordial en la actividad biológica con respecto a la señalización del TLR4.

Con respecto a los moduladores sintéticos de tipo LPS, estudiamos un grupo de derivados de glucosamina. El FP7, un derivado de glucosamina con dos grupos fosfatos y dos cadenas de ácido mirístico (C_{14}), es activo en la inhibición de manera dosis-dependiente de la activación por LPS del TLR4 humano y de ratón. Los experimentos de RMN sugieren que el FP7 interacciona con MD-2, probablemente insertando sus cadenas de ácido graso en la cavidad de unión hidrofóbica. Hemos diseñado nuevos moduladores del TLR4, basados en la estructura química del FP7, y hemos realizado estudios de relación estructura-actividad (*structure-activity relationship*, SAR) para comprender cómo la longitud de las cadenas de los ácidos grasos determina su potencia

como moduladores del TLR4. Estos derivados del FP7 difieren solo en las longitudes de las cadenas del ácido graso (10, 12, 14 y 16 átomos de carbono). En este estudio, hemos tenido en cuenta, tanto la interacción con MD-2, como las propiedades de agregación de las moléculas. Los datos biológicos estructurales y funcionales demuestran la capacidad de estos análogos nuevos de FP7 para regular negativamente la señalización de TLR4 en diferentes sistemas de modelos celulares. Nuestros estudios computacionales han sido de gran relevancia en el contexto del estudio SAR y para proponer la justificación del mecanismo de unión. Nuestros modelos sugieren que existe una longitud óptima para las cadenas de ácido graso para una actividad antagonista de TLR4 apropiada, relacionada con el modo de unión y con las propiedades físico-químicas de los derivados de FP7.

En cuanto a los moduladores con estructuras de tipo no lipopolisacarídica, hemos estudiado guanidino-calixarenos anfifílicos. Para bloquear la señalización anormal del TLR4 en la sepsis bacteriana, se han desarrollado dos estrategias distintas. La primera se basa en la neutralización del LPS mediante la formación de aductos no covalentes con compuestos catiónicos, lo que impide que los LPS interaccionen con los receptores. La segunda estrategia se basa en el uso de moléculas que compiten con el LPS endotóxico en la unión al mismo sitio en CD14 y MD-2, inhibiendo así la inducción de la transducción de señal al alterar la dimerización del receptor iniciado por LPS. Entre los guanidino-calixarenos anfifílicos estudiados, como compuesto de referencia, se incluyó uno cuya actividad en este contexto biológico había sido publicada anteriormente. Su actividad biológica se asoció con su capacidad de unirse y neutralizar LPS como topomimético de péptidos de unión a LPS. Nuestros estudios computacionales nos han permitido proponer una alternativa a esta propuesta de mecanismo. Nuestra hipótesis se basa en que los anfífilos faciales basados en calixarenos también pueden ser andamios apropiados para obtener ligandos de TLR4 con actividad antagonista. En un contexto biológico, los calixarenos anfifílicos mostraron propiedades notables relacionadas con su anfifilicidad. Dado que la hipótesis de que los derivados de calixareno podrían unirse directamente a los receptores MD-2 y CD14 de una manera similar a LPS, hemos realizado cálculos de *docking* para apoyar este modo de interacción. Además, hemos estudiado si la actividad antagonista de TLR4 es una propiedad general de los calixarenos anfifílicos cargados positivamente y si este efecto antagonista también se podría derivar de la interacción directa de calixarenos con

los receptores y no exclusivamente de la acción neutralizante de LPS, como se sugirió. Las evidencias experimentales mostraron que algunos de estos calixarenos eran activos en la inhibición, de una manera dependiente de la dosis, de la activación del TLR4 estimulada por LPS y de la producción de citoquina dependiente de TLR4 en células humanas y de ratón. Además, estos guanidino-calixarenos también inhibieron la señalización del TLR4 cuando el TLR4 se activó mediante un estímulo que no era de LPS, sino la lectina de planta PHA. Estos resultados apuntan a que el esqueleto de calixareno puede ser un andamio estructural potencial para el desarrollo de nuevos moduladores dirigidos al TLR4.

En cuanto a la activación del TLR4, se han realizado estudios computacionales de los diferentes dominios que componen el TLR4 con el objetivo de descubrir detalles del mecanismo de activación del receptor insertado en la membrana celular. La comprensión a escala atómica del proceso de dimerización, tanto del dominio de transmembrana como del dominio intracelular del TLR4, nos puede permitir el diseño de nuevos moduladores que favorezcan ciertos modos de unión que modulen su activación.

CHAPTER 1

Introduction

1.1 Scientific background about TLR4

1.1.1 Historical background

The lipopolysaccharide-binding protein (LBP) was the first important component of the LPS-sensing bimolecular machinery to be discovered; it was reported as early as 1986, isolated from rabbit serum[1]. Three years later, in 1989, the same research team performed a number of binding assays using LPS from various origins providing the first insights into the propensity of lipid A-containing molecules to bind to LBP.[2] A year later, Schumann *et al.* reported the structure and function of LBP.[3] Recently, in 2013, the crystal structure of mouse LPB was reported and deposited in the Protein Data Bank (PDB) under the accession code 4M4D.[4]

The cluster of differentiation 14 (CD14) was the second component of the LPS receptor complex to be found and characterized. The mouse and human CD14 primary structure were reported in a 1989 paper, revealing its leucine-rich nature.[5] The mouse crystal structure of CD14 was reported in 2005 and deposited in the PDB under the accession code 1WWL and the human crystal structure in 2012 (accession code 4GLP).[6-7]

From a mechanistic point of view, the missing piece of the puzzle, and arguably the most important one, was discovered only at the end of the last century. In a 1996 study, Lemaitre *et al.* observed that particular Drosophila mutants reacted differently to fungal infection and traced this difference down to a single gene, leading to the understanding that the Toll protein of Drosophila is indispensable to the activation of its immune system.[8] Two years later, Poltorak *et al.* studied several sub-strains of mice and noted that the ones presenting mutation in a gene resembling the Toll gene of *Drosophila*, had lost their LPS responsiveness. Due to its high resemblance and providing that TLR1, 2 and 3 had already been reported, the gene and the protein it codes were dubbed Toll-like-receptor 4 (TLR4).[9] Together these discoveries revealed that mammals and fruit flies are equipped with similar tools to detect the presence of pathogens. The evasive protein that activates the innate immunity was finally discovered. Bruce A. Beutler and Jules A. Hoffmann were awarded the Nobel Prize in Physiology or Medicine of 2011 for their findings.[10] In 2009, a crystal structure of the TLR4/MD-2/LPS dimer complex was reported, known as the active, or agonist, conformation of TLR4. This achievement constituted the structural basis for TLR4 activation.[11] A more in-depth review of the history of the TLRs and their associated adaptors can be found here.[12]

1.1.2 Biological role

Toll-like receptors (TLRs) are classified as pattern recognition receptors (PRRs) and have a primordial role in the activation of the innate immunity. TLRs are trans-membrane proteins located extra and intracellularly and are specialized in the recognition of pathogen-associated molecular patterns (PAMPs).[13] TLR4 is the mammalian endotoxin sensor.[14]

TLR4 represents an interesting case study for several reasons: (i) it is the only TLR that requires the presence of an accessory protein (MD-2) to function; (ii) it can activate the immune response through two different signaling pathways (cf. *1.1.3 TLR4 activation pathways*); and (iii) it reacts differently to specific PAMP lipopolysaccharides (LPS), a component of the outer membrane of Gram-negative bacteria, that are either agonist or antagonist of this receptor depending on minute variation in their structures.[15]

Early in the LPS recognition, lipopolysaccharide binding protein (LPB) binds to LPS and brings it to cell surface PRR CD14, which delivers it to MD-2 inducing the formation of the TLR4/MD-2/TLR4*/MD-2* heterodimer. Both LPB and CD14 are essential for detecting small amount of circulating LPS,[16] however, greater amount of LPS can activate TLR4 in the absence of both LPB and CD14.[3, 17] The dimerization of two TLR4 ectodomains induces the dimerization of their associated intracellular domains, which leads to the recruitment of downstream adaptors and to the activation of the intracellular signaling events triggering the immune response (Figure 1.1). The binding of an antagonist ligand to the extracellular domain prevents the formation of the dimer, and consequently, the intracellular signaling events to occur.

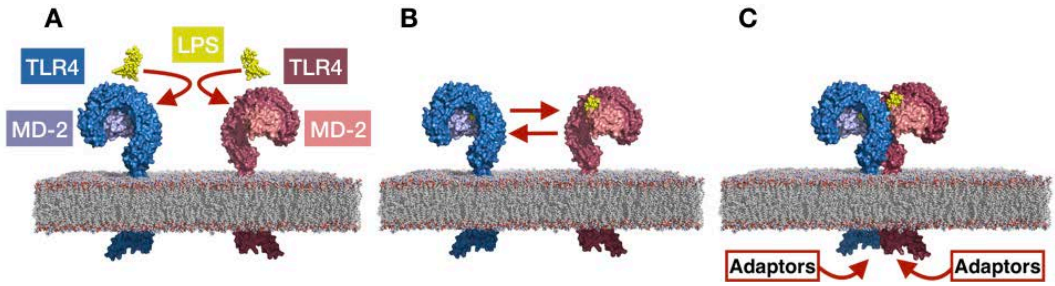


Figure 1.1. Schematic representation of the LPS-induced dimerization of the TLR4/MD-2 complex leading to immune system activation. Red arrows indicate motion and mutual recognition. (A) Two LPSs are engaged by two distinct TLR4/MD-2 systems; (B) two TLR4/MD-2/LPS complexes dimerize by protein-protein interactions; (C) dimerization brings together the two intracellular TIR-containing domains providing a suitable molecular surface for recruiting downstream adaptors.

1.1.3 TLR4 activation pathways

Activation of the Toll-like receptor 4 by lipopolysaccharide or other TLR4 ligands, triggers the release of pro-inflammatory cytokines that provoke strong immune responses through induction of dendritic cell (DC) maturation and expression of type-1 interferon genes and of IFN-regulated genes.[18] At least two known distinct intracellular signaling pathways lead to this immune response, namely the MyD88-dependent pathway and the TRIF-dependent pathway. Intracellular TLR4 signaling is mainly governed by the Toll/interleukin-1 receptor (TIR) homology domain that is conserved in the four major adaptors involved in the signaling (further mentioned by their abbreviations in bold):

- Myeloid differentiation factor 88 (**MyD88**)
- MyD88-adapter-like (**Mal**) protein, also known as TIR-domain-containing adapter protein (TIRAP)
- TIR-domain-containing adapter-inducing interferon-β (**TRIF**) also called TIR-domain-containing adapter molecule-1 (TICAM-1)
- TRIF-related adapter molecule (**TRAM**), also called TIR-containing protein (TIRP), or TIR-containing adapter molecule-2 (TICAM-2)

The MyD88-dependent pathway is known to regulate early NF-κB activation and related inflammatory cytokine production.[19] In this pathway, MAL and MyD88 are immediately recruited on the molecular surface newly created from the dimerization of two TLR4 intracellular domains.

The TRIF-dependent pathway triggers the secretion of TNF-α which then binds to their receptor activating NF-κB, known as late phase NF-κB activation. This second activation pathway requires endocytosis of the activated dimer of the TLR4/MD-2/ligand complex. TRAM and TRIF are recruited by the intracellular domain of TLR4.

A lot more adaptors are involved in these two pathways; their roles have been extensively reviewed.[20-21]

1.1.4 Natural LPSs

Lipopolysaccharides (LPSs) naturally occur as a primary constituent of the outer membrane (OM) of Gram-negative bacteria playing an important role toward its physicochemical properties. These properties depend on the self-aggregating behavior

of LPSs, which is in turn finely tuned by the LPSs molecular structures. The occurrence of negatively charged phosphate groups, decorating most LPS, contributes to the structure and the low fluidity of the LPS monolayer, because they engage in electrostatic interactions with divalent cations such as calcium (Ca^{2+}) and magnesium (Mg^{2+}) normally present on the surface of the OM.[22-23] This is reflected in bacterial resistance to harsh conditions both in the case of extremophiles and in the case of bacteria exposed to dangerous external compounds such as antibiotics.[24] A plethora of other features characterizing the LPS structure have been demonstrated to be involved in the capability of the bacteria to resist stress factors.[22-23] Given the wide structural heterogeneity of LPS structures, it is reasonable to assume that the external environment is one of the key drivers in promoting such high structural diversity. This was further corroborated by the observation of bacteria belonging to the same species producing diverse LPS compounds under different growth conditions.[23]

Despite the structural heterogeneity, in general, three different structural domains can be identified in most LPSs: a lipid moiety, known as lipid A, and a polysaccharide composed of a core oligosaccharide (outer and inner part) and an *O*-specific polysaccharide (Figure 1.2).

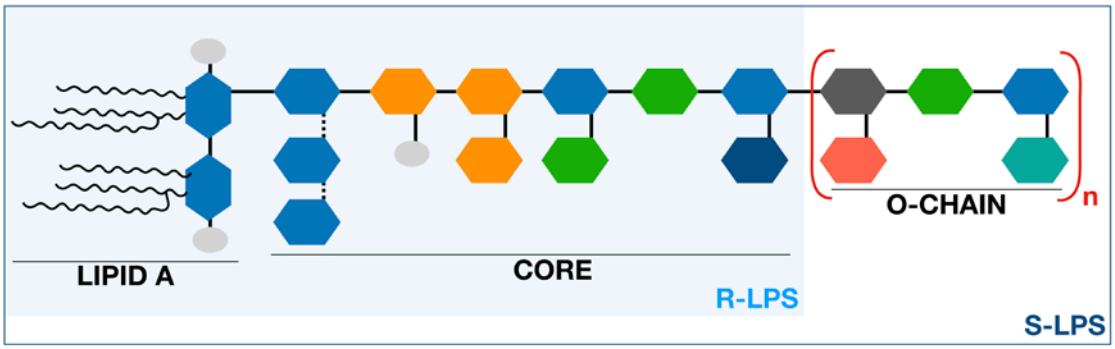


Figure 1.2. Schematic representation of the general chemical structure of LPSs possessing phosphorylated core regions. S-type LPSs are built up of three distinct moieties, termed lipid A, core OS, and the O-chain. Lipid A is embedded in the bacterial OM and is the most conservative part of the LPS, whereas the core OS and O-chain regions are more exposed to the environment and structurally variable. In cases of absent or truncated O-chains, the terminology employed is R-type LPS or LOS.

In the case of absence of the polysaccharide portion, the terminology currently used to designate the LPS is lipooligosaccharide (LOS) or rough-type LPS (R-LPS), whereas the complete form with all three domains is termed smooth-type LPS (S-LPS).[22-23, 25] The lipid A is the most conserved part of the LPS.[26] Despite the general conservative

structure, microheterogeneity has also been observed both in the acylation and phosphorylation pattern of the lipid A moiety. Briefly, it is worth underlining that fatty acids (FAs) can be attached to the glucosamine disaccharide backbone either in an "asymmetric" (4+2 e.g. as in *E. coli* lipid A, Figure 1.4 and Figure 1.7) or a "symmetric" (3+3 e.g. as in *Neisseria meninigitidis* lipid A) fashion. With regard to the phosphorylation pattern a plethora of different substituents directly linked to the phosphate groups have been detected including a further phosphate unit or additional sugar residues, such as aminoarabinose or uronic acid units, as well as a 2-aminoethanol group (EtN).[26]

1.1.5 TLR4 activation by LPS/lipid A and structural insights

The quest to find novel TLR4 modulators for clinical applications, especially modulators with known LPS-like structure, has been going on for a while and is still ongoing. The techniques computational approaches used toward that aim include virtual screening (VS) techniques following drug design methodologies.[27-28] There are several X-ray crystallographic structures of TLR4 from various species and in particular for human, several structures are reported in complex with different ligands, details are given in Table 1.1. These structures provide many insights into the structural variation between the agonist and antagonist conformations, helping us to understand the changes taking places at the atomic scale.[15]

In the X-ray crystallographic structure of TLR4/MD-2 in complex with the most potent agonist known to date, *E. coli* LPS (PDB ID 3FXI), MD-2 large hydrophobic cavity accommodates five fatty acid (FA) chains (Figure 1.3 and Figure 1.4). The sixth FA chain, the one in magenta in Figure 1.4b, protrudes from the MD-2 hydrophobic pocket toward the partner TLR4 (TLR4*) completing the dimerization interface. The phosphate groups of the LPS are anchored to Arg90, Lys91, Ser118 and Lys122 (Figure 1.5) from the polar rim of MD-2, and the polysaccharide moiety establishes a network of polar interactions with TLR4. Two distinct protein-protein interaction regions have been defined between TLR4 and MD-2, termed patches A and B (Figure 1.3), which respectively contain residues close to the N-terminal, and to the central domain of TLR4. Therefore, all the structural components of the LPS molecule are important for binding to, and recognition by, the TLR4/MD-2 complex.[29]

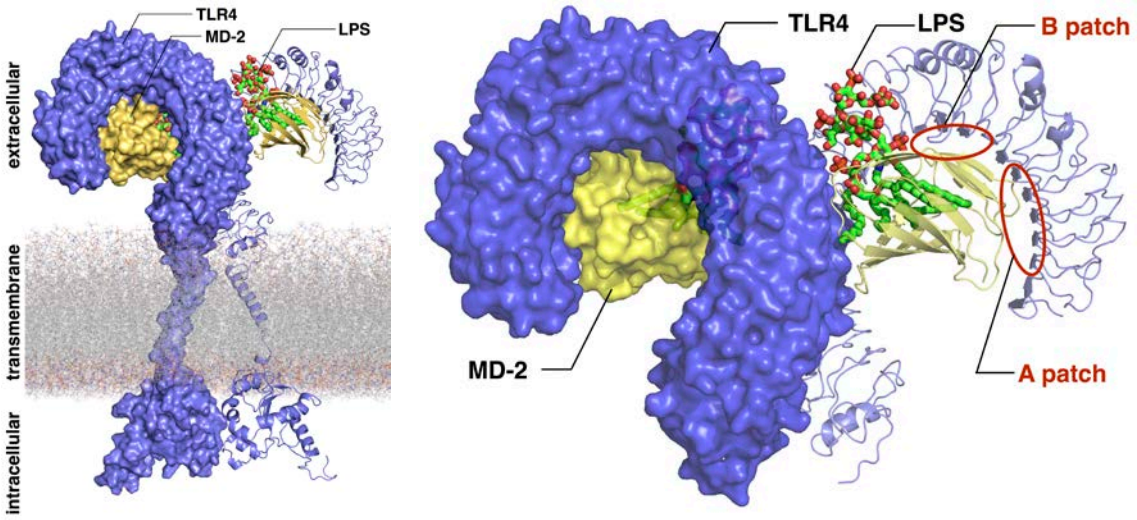


Figure 1.3. Representation of the 3D structure of TLR4/MD-2/LPS. On the left: large-scale representation showing the intracellular, transmembrane and extracellular domains of TLR4/MD-2 in complex with *E. coli* LPS. 3D Structures correspond to the X-ray crystallographic structure for the extracellular domain (PDB ID 3FXI) and homology modeling for the transmembrane and intracellular domains. On the right: a close-up look at the TLR4 extracellular domain (purple) along with MD-2 (yellow) and LPS (CPK colors with C atoms in green) from PDB ID 3FXI.

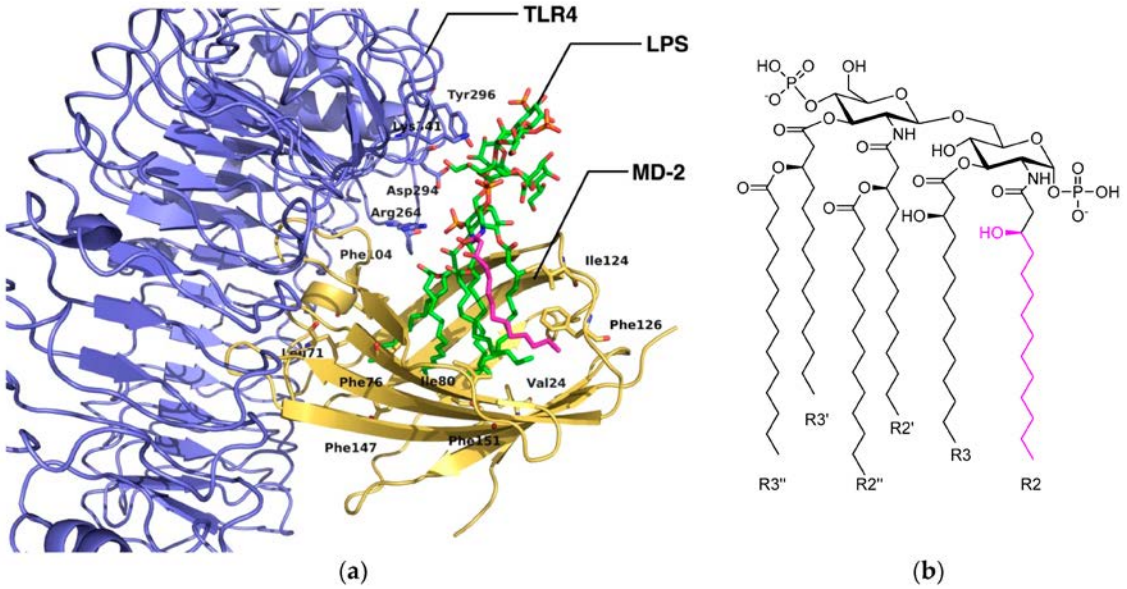


Figure 1.4. Representation of the LPS in complex with TLR4/MD-2. (**a**) Detail of the 3D structure of the complex between TLR4/MD-2 and *E. coli* LPS (CPK colors with C atoms in green and R2 C atoms in magenta) from the X-ray crystallographic structure (PDB ID 3FXI); (**b**) chemical structure of *E. coli* lipid A. The R2 FA chain (magenta) placed at the channel of MD-2 completes the dimerization interface.

Lipid A is composed of FA chains of different lengths attached to a 1,4-β-diphosphorylated diglucosamine backbone.[26] The agonistic activity of lipid A has been mainly attributed to the number, length and chemical structure of its FA chains, as well as to its phosphorylation degree and the number and types of substituted groups

attached to the phosphates.[26] Recent findings have questioned this paradigm as an occurrence of immunostimulatory LPSs bearing penta-acylated lipid As and positively-charged residues decorating their lipid As have been reported.[30-31] These data suggest that subtle changes in lipid A structure may profoundly impact the innate immune response.

Regarding LPS recognition by MD-2, the placement of the sixth FA chain into a specific hydrophobic channel of MD-2 assists the binding with TLR4* (Figure 1.4).[32-33] The general assumption is that Phe126 is the "molecular switch" in endotoxic signaling. Upon lipid A/LPS binding, the flexible MD-2 protein experiences a local conformational change involving the side chain of Phe126 and the surrounding residues: the loop formed by residues 123-129 (Figure 1.5).[34]

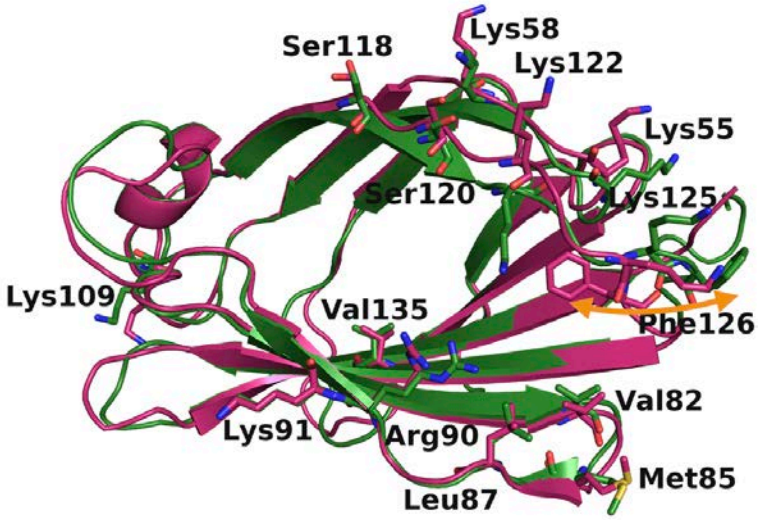


Figure 1.5. Superimposition of the X-ray crystallographic structures of the agonist (magenta) and the antagonist (green) conformations of MD-2 from PDB ID 3FXI and 2E56, respectively. Bound ligands have been hidden for clarity (*E. coli* LPS in 3FXI; three myristic acids in 2E56). Conformational change of the molecular switch Phe126 is marked by an orange double arrow.

Interestingly, tetraacylated lipid IVa, a lipid A precursor, binds in an antagonistic manner to human MD-2, whereas it binds in an agonist manner to mouse MD-2.[35] This causes a proinflammatory effect in mouse cells (agonism), but no effect in human. This behavior was also studied in other mammalian species. For instance, lipid IVa is a weak agonist of equine cells, while it antagonizes lipid A-induced activation in dogs. X-ray crystallographic structures of hMD-2/lipid IVa (PDB ID 2E59) and mMD-2/lipid IVa (PDB ID 3VQ1) complexes, reveal that four FA chains of lipid IVa are inserted into the

MD-2 pocket, occupying a similar volume in both human and mouse TLR4/MD-2, although with different consequences: exerting an agonist activity in human and an antagonist activity in mouse. From the crystal structures, one can see that the orientation of the lipid IVa is rotated by 180° in the di-saccharide plan thus lipid IVa presents two different molecular patterns of interaction for human and mouse (Figure 1.6).[34, 36-37] These different binding modes of lipid IVa, which determine how the phosphate groups interact with the TLR4/MD-2 complex, may be crucial to explaining its distinct behavior in different species. This illustrates the importance of scrutinizing the key ligand/receptor interaction to rationalize the mechanism for TLR4 modulation.

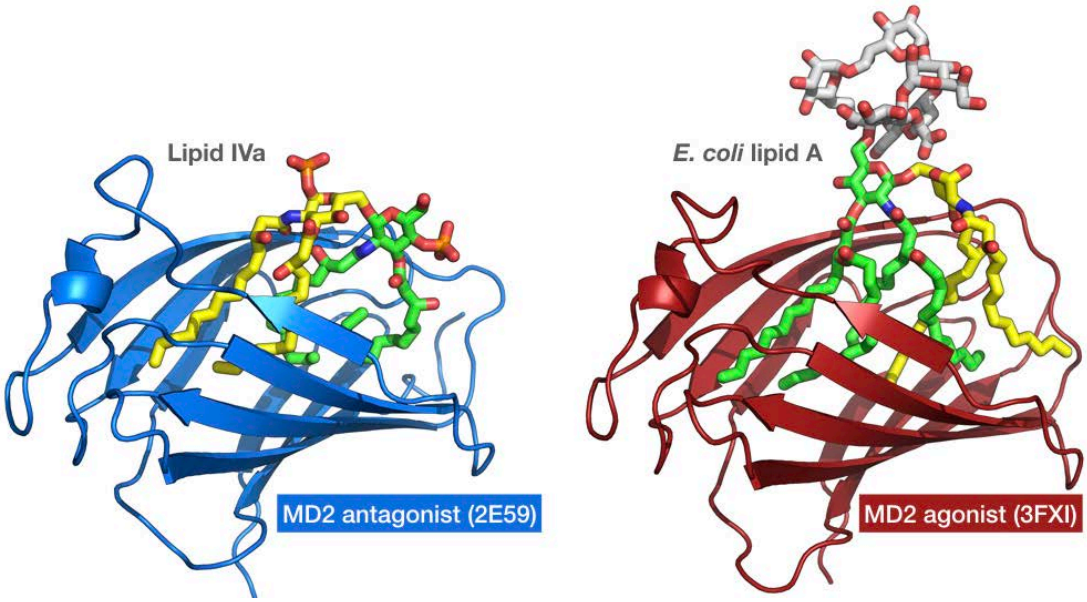


Figure 1.6. Representation of type A (antagonist-like) binding mode as known from lipid IVa in PDB ID 2E59 (on the left) and type B (agonist-like) binding mode as for *E. coli* lipid A in PDB ID 3FXI (on the right).

Table 1.1. X-Ray crystallographic structures of TLR4 deposited in the Protein Data Bank. ([a]) The structures are noted as multimer when two TLR4/MD-2 heterodimers are found in the structure. ([b]) Non-natural complexes. ([c]) *Escherichia coli* LPS. ([d]) Re-chemotype of *Escherichia coli* LPS. ([e]) Human TLR4 polymorphism D299G and T399I. ([f]) Human TLR4 decoy receptors constructed by combining LRR modules from TLR4 ectodomain and variable lymphocyte receptors (VLR) from jawless fish. ([g]) Three units of myristic acid.

PDB ID	Organism	Proteins	Ligands	Structure[a]	MD-2 conformation	Resolution (Å)
3FXI[11]	Human	TLR4/MD-2	LPS[c]	Multimer	agonist	3.10
4G8A[38]	Human	TLR4/MD-2[e]	Re-LPS[d]	Multimer	agonist	2.40
3ULA[39]	Human-inshore hagfish hybrid[f]	TLR4 fragment/ MD-2	E55	Heterodimer[b]	antagonist	3.60
3UL7, 3UL8, 3UL9[39]	Human-inshore hagfish hybrid[f]	TLR4 fragment	None	monomer	-	2.37, 2.50, 2.45
2Z65[36]	Human-inshore hagfish hybrid	TLR4 fragment/ MD-2	E55	heterodimer[b]	antagonist	2.70
2Z63[36]	Human-inshore hagfish hybrid	TLR4	-	monomer	-	2.00
2Z62[36]	Human-inshore hagfish hybrid	TLR4 fragment	-	monomer	-	1.70
2Z66[36]	Human-inshore hagfish hybrid	TLR4 fragment	-	tetramer[b]	-	1.90
2E56[34]	Human	MD-2	Myristic acid[g]	monomer	antagonist	2.00
2E59[34]	Human	MD-2	Lipid IVa	monomer	antagonist	2.21

1.2 Computational studies about TLR4: ligand recognition

1.2.1 Natural LPSs

Rhodobacter sphaeroides lipid A (*Rs*LA, Figure 1.7)[40] has five acyl chains, with one unsaturated and two shorter chains than *E. coli* lipid A. The penta-acyl chain-containing *Rs*LA is midway between the agonist (six FA chains) and the antagonist (four FA chains) structures. It activates the TLR4 pathway in horses and hamsters while inhibiting it in humans and mice, raising questions about the molecular recognition process as the horse TLR4/MD-2 sequence is more closely related to the human sequence than to the mouse one.

To clarify the species-specific response, a computational-aided study of the 3D structures from the three species was undertaken. Homology models were built for horse and hamster TLR4/MD-2 with MODELLER[41] using human and murine X-ray crystallographic structures as templates (PDB ID 3FXI and 2Z64). This study showed that Arg385 plays an important role in horse TLR4 complex activation by lipid IVa through polar interactions between the guanidinium moiety and the phosphate group of lipid IVa.[42] In human and hamster TLR4 this residue is substituted by a glycine and by an alanine in murine. Autodock Vina generated a docked structure of the horse TLR4/MD-2/*Rs*La complex that closely resembles the pose of lipid IVa in the murine crystal structure of TLR4/MD-2. On the contrary, the binding pose calculated for hamster MD-2 was similar to the one of lipid IVa in the crystal structure of chicken (PDB ID 3MU3) and human MD-2. The difference noticed over the species was mainly attributed to the compositional variation of their corresponding proteins. Autodock docking studies on hMD-2 revealed that the longest chain of *Rs*LPS could be accommodated in MD-2 by undergoing a fold like in the case of Eritoran. The diglucosamine polar head is always exposed to the solvent.

Through molecular modeling, Irvine *et al.* showed that the different human/horse TLR4 responses towards *Rs*LA can be attributed to two different amino acids, Gly384 and Ser441 in human TLR4 that are Arg385 and Pro442 in horse.[43] TLR4 Arg385 in horse, although located at around 9 Å from the docked *Rs*LA, is shown to establish a critical long-range electrostatic interaction with a phosphate group of *Rs*LA, while Pro442, situated near the TLR4* dimerization interface, interacts with a FA chain of *Rs*LA through van der Waals interactions. This hypothesis was confirmed by experimental

assays using HEK293 cells transfected with G384R/S441P hTLR4 and eqMD-2, and R385G/P442S eqTLR4 and hMD-2. It was observed that the R385G/P442S mutations in horse caused a complete loss of activity, and that in human the double mutant G384R/P441S TLR4 was unable to activate the signaling event. Since the double mutation did not restore the activity, other residues must be required. The docking of *Rs*LA in human TLR4/MD-2 shares some similarity with the Eritoran crystal structure, such as the folding of the longest acyl chain and the polar interaction with charged residues of MD-2. *Rs*LPS can adopt two orientations depending on the position of 1-PO4 (primarily oriented towards TLR4 in horse and towards TLR4* in human). Superimposition of docked *Rs*LA with X-ray crystallography structures of lipid A and lipid IVa showed that *Rs*LA and lipid A acyl chains occupy more volume than lipid IVa. More importantly, the R2 chain of *Rs*LA and lipid A protrudes from MD-2 establishing interactions with TLR4* unlike the R2 chain of lipid IVa that is folded into the MD-2 pocket (Figure 1.4).

The severe pathogen *B. cenocepacia* LPS (Figure 1.7) was reported by Di Lorenzo *et al.* to strongly activate human TLR4/MD-2 despite that its lipid A has only five acyl chains.[30] LPS-induced endotoxic shock experiments in mouse confirmed the proinflammatory *B. cenocepacia* lipid A activity. A combination of docking calculations and MD simulations, backed-up by evidence from an experimental study involving mutations in the TLR4/MD-2 protein-protein interface, suggested that the longer acyl chains allow reaching deeper regions inside the MD-2 pocket, thus compensating for the absence of a sixth FA chain and permitting the fifth FA chain to be exposed on the MD-2 surface where it interacts with TLR4* to promote dimerization. The replacement of Val82 by a phenylalanine enhanced the inflammatory response. This was explained by the conversion of Van der Waals interactions into stronger CH-π interactions with *B. cenocepacia* lipid A FA chain, longer than its *E. coli* lipid A counterpart. The molecular model also showed that Ara4N residues provide additional polar interactions affecting *B. cenocepacia* LPS binding to TLR4/MD-2, and contribute to the anchoring of the lipid A into the receptor complex through interactions with both TLR4 and TLR4*. Interestingly, the presence of the positively-charged ammonium groups in the Ara4N seems to favor the electrostatic interactions and, consequently, the binding, whereas uncharged amino acids are critical for responses to *B. pertussis* lipid A.[31] This TLR4/MD-2/LPSBC model was used to generate a

computational mutant, D294A, R322A, S415A* and S416A*, which was submitted to MD simulations and energy analysis for quantification of the per-residue contributions to the final binding energy using a method called MM-ISMSA.[44] This study permitted the identification of the mutated residues as major contributors to the total binding energy of *B. cenocepacia* LPS and suggested that the ammonium groups of Ara4N stabilize the complex by providing additional anchorage interactions. Altogether, these results provide a molecular explanation for the activation of the human TLR4/MD-2 complex by a penta-acylated lipid A.

1.2.2 Synthetic LPS mimetics

TLR4 ligands, inspired by the LPS structure, have been designed and synthesized. Eritoran was one of the first LPS-derivative to enter clinical trials and to reach phase III as an antisepsis agent (TLR4 antagonist), but failed to reach approval as the eritoran-treated group did not outperform the group that received a placebo.[45] Eritoran is a tetraacylated lipid A derived from the structure of *R. sphaeroides* Lipid A, although antagonist of human TLR4, it is an agonist of mouse and horse TLR4. The species-dependent activity of Eritoran was investigated in a series of docking calculations, with Autodock, in human, mouse and horse TLR4. Homology models based on crystal structures were built with the SCWRL4 program[46] for the TLR4 of species lacking experimentally resolved structure.[37] This study revealed non-conserved amino acids that are important for the binding of Eritoran: Lys58 in human, which corresponds to an asparagine in mouse and a glutamic acid in horse, Lys388, a serine in mouse and a lysine in horse, and Gln436, an arginine in mouse and a glutamine in horse. These residues are part of non-conserved interaction patterns that are primordial for the ligand TLR4-TLR4* bridging role, permitting the dimerization thus determinant in the agonist activity of the ligand.

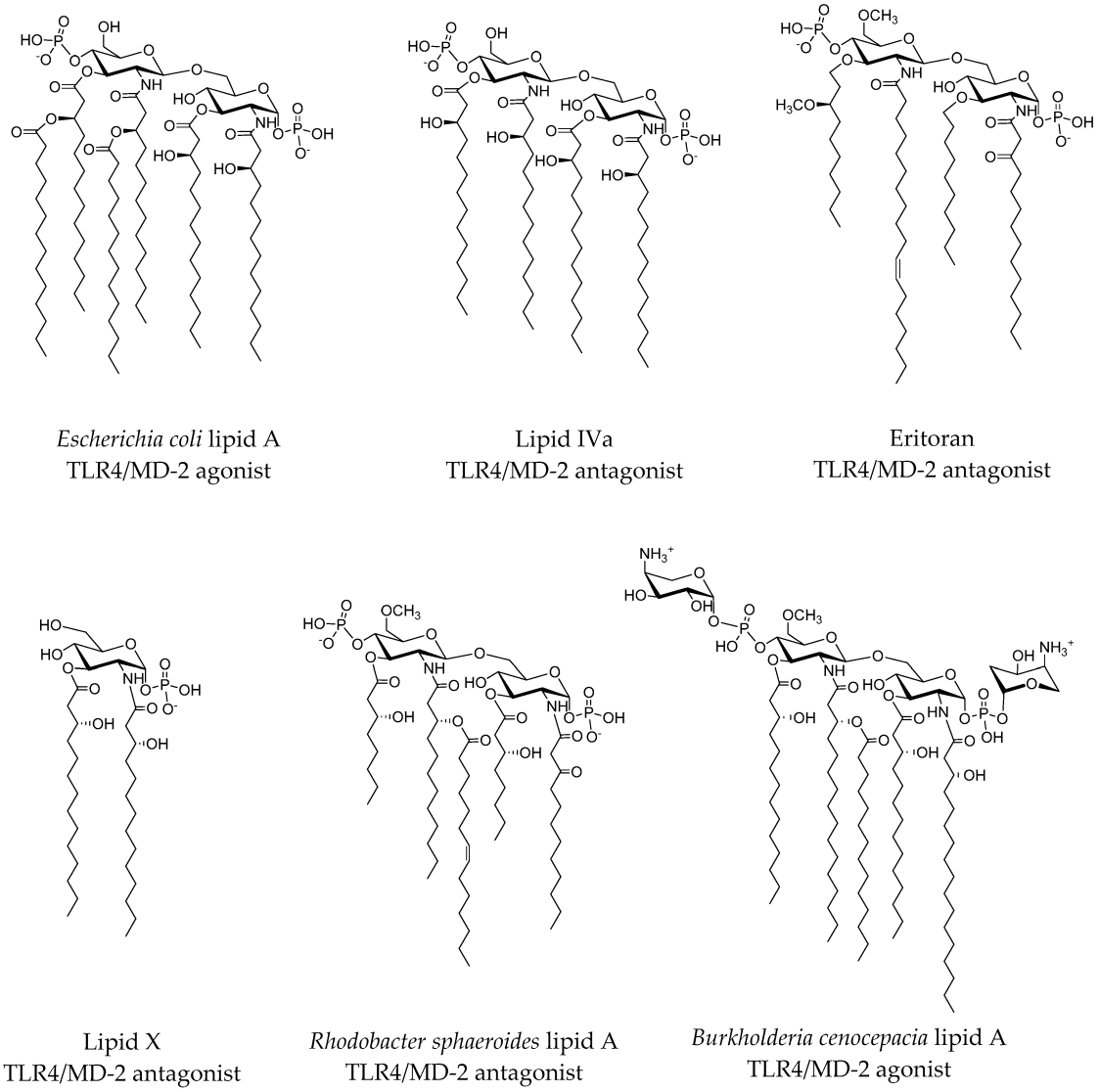


Figure 1.7. Lipid A and synthetic lipid A analogues with activity as TLR4 modulators. Activity is referred to hTMR4/MD-2.

In another study by Cighetti *et al.* new modulators were proposed based on a diphosphorylazed lipid X (Figure 1.7), a biosynthetic precursor of lipid A, scaffold leading to Compound **1** (Figure 1.8).[47] This molecule turned out to be an antagonist of both human and mouse TLR4 and was also shown to stimulate CD14 internalization in bone-marrow-derived murine macrophages. These findings indicate that Compound **1** targets CD14 in a TLR4-independent manner. The authors proposed models of interactions for Compound **1** with both CD14 and MD-2, derived from docking calculation. These models were backed-up by NMR experiments that clearly showed FA chain-protein interactions. Due to its favorable solubility properties and its lack of toxicity, according to MTT tests, Compound **1** was described as a promising TLR4 modulator.

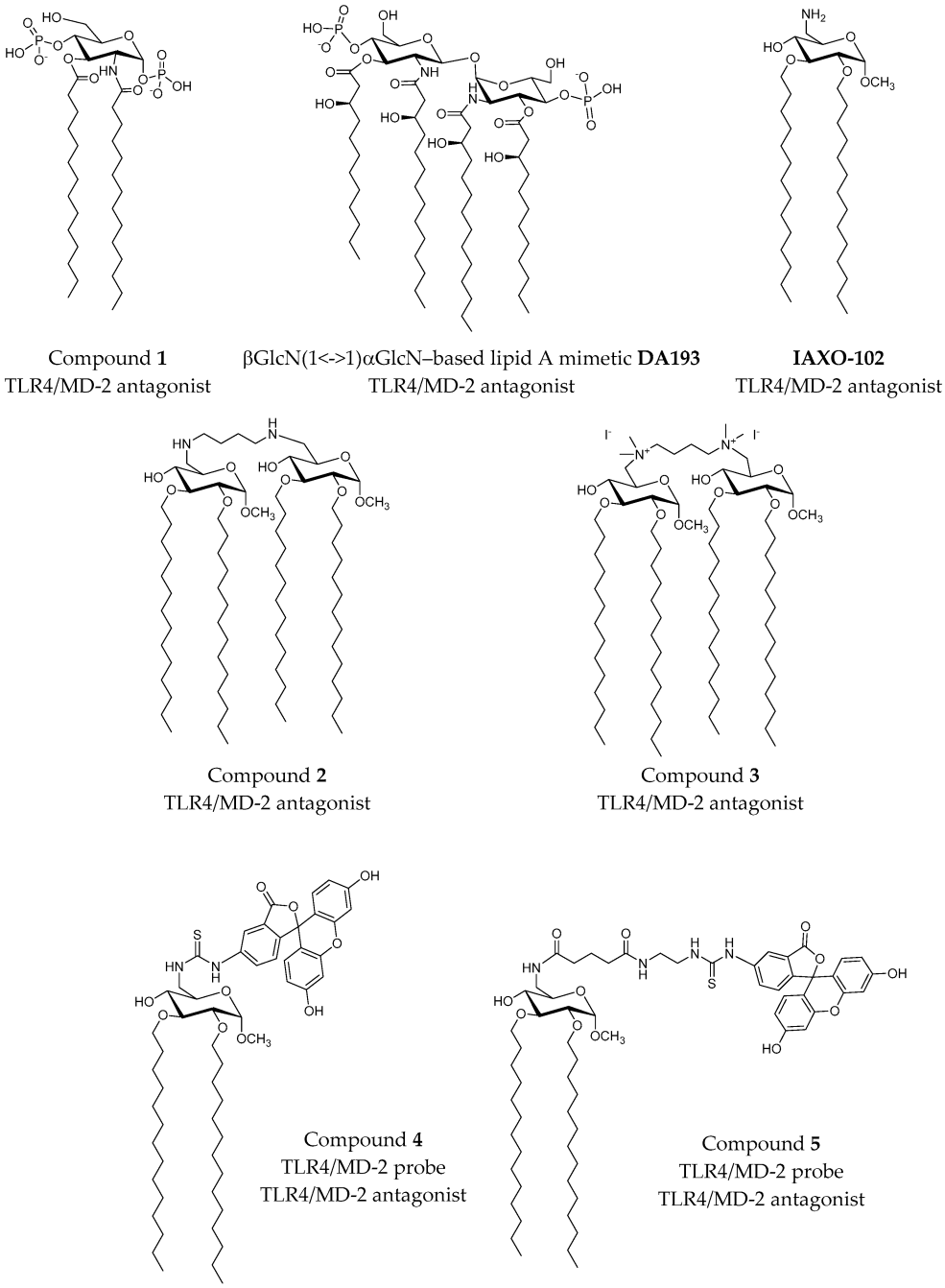


Figure 1.8. Synthetic LPS mimetics studied by computational approaches.

Another approach to developing lipid A-based modulator consisted in replacing the flexible three-bond β(1→6) diglucosamine linkage by a two-bond β,α(1↔1) glycosidic linkage conferring rigidity to the molecule.[48] Compound **DA193** (Figure 1.8), designed out of this new βGlcN(1↔1)αGlcN scaffold, proved to be a dose-dependent antagonist of human and mouse TLR4, according to assays performed in HEK293 cells transiently transfected with membrane CD14 (mCD14)/hMD-2TLR4, HEK293 cells transfected with hMD-2/TLR4 only and assays on human macrophage-like cell line (THP-1). To

propose an atomistic understanding of the interactions between the ligand and the receptor, MD simulations of 11 ns were run starting from two possible binding orientations of the ligand into the MD-2 protein: one with the α-GlcN ring facing the Phe126 loop and a second one with the β-GlcN ring facing the Phe126 loop with an energy difference similar to the one found for orientations of *E. coli* lipid A in the binding site of hMD-2. Dissociation constants of the MD-2/DA193 complex, calculated from MD simulations, gave a binding to MD-2 twenty-fold and three-fold stronger than *E. coli* lipid A and lipid IVa, respectively. It was concluded that the conformational rigidity of the βα(1↔1) diglucosamine backbone of the tetraacylated lipid A mimetics ensures strong binding to MD-2, in two possible binding poses.

The commercial TLR4 antagonist IAXO-102 (Figure 1.8) also inspired the rational design of TLR4 modulators and probes.[49] In a study reported by Ciaramelli *et al.*[50] the design was based on a previous docked binding mode of IAXO-102 into MD-2 that revealed its capacity to host two ligands simultaneously. A dimeric scaffold with two glycolipid units was designed by connecting both units through C4 di-amino and di-ammonium linkers (Figure 1.8; Compounds **2** and **3**). Both compounds were confirmed to inhibit TLR4 activation and signaling, in a concentration-dependent manner, in HEK-Blue™ cells expressing hTLR4. However, these compounds were reported to have a very poor solubility in aqueous solution.

The same IAXO-102 scaffold was used to design fluorescent probes (Figure 1.8; Compounds **4** and **5**).[50] The fluorescein moiety was chosen as the chromophore, and two thiourea-based linkers with different lengths attached to the C6 position of the glucose moiety were considered. Following a normal mode analysis, Compounds **4** and **5** were docked in three different conformations of CD14. Calculations predicted binding poses in which the fatty acid chains are buried inside both human and mouse CD14 binding site with the sugar located outside. The thiourea linker and the fluorescein moiety establish polar interactions with the hydrophilic rim but no preferred pose was reported. The best complexes were submitted to MD simulations and MM-GBSA analysis. A hybrid hTLR4/MD-2 model featuring MD-2 in its antagonist conformation was built to perform docking calculation. Predicted binding poses were similar to those found for CD-14 as the FA chains were inserted in the hydrophobic pocket of MD-2, additionally, the fluorescein moiety reached TLR4 in the case of the longer probe.

Calculations of the solvent-accessible surface area (SASA) with CASTp[51] in both CD14 and MD-2 showed that both pockets have similar topologies and volumes. The presence of a lower number of polar residues at the rim of the CD-14 pocket allows for recognition of a wide range of microbial and cellular molecular components, e.g. lipopeptides to be transferred to TLR2. In contrast, the selectivity of MD-2 towards LPS arises from the polarity of its rim.

1.2.3 Non-LPS-Like TLR4 Modulators

The species-specific discrimination of TLR4 ligands by MD-2 is exemplified by taxanes, in particular paclitaxel (PTX; Figure 1.9), a proinflammatory murine TLR4/MD-2 ligand, which activates the subsequent inflammatory cytokine response.[52-54] Zimmer *et al.* HEK293 transfected cells to demonstrate that the activation of TLR4 by PTX requires the presence mMD-2 and either hTLR4 or mTLR4.[55] They explained the need for mMD-2 to be due to the electrostatic potential surfaces, the hydrophobicity, the binding pocket size and the conformation of the loop formed by amino acids 123 to 130. hMD-2 and mMD-2 have a very large cavity volume that in principle allows either lipid IVa, PTX or Eritoran to fit inside. The study of the electrostatic surfaces of mMD-2 (PDB ID 2Z64) and hMD-2 (PDB ID 2Z65) with the SYBYL software[56] shows that the cavities for both structures are close to electroneutral, with mMD-2 being more electronegative than hMD-2, especially at the Cys95-Cys105 loop critical for MD-2/TLR4 interaction. The electrostatic surface of hMD-2 has three electropositive patches corresponding to Lys58, Lys122 and Lys125, which are absent of the mMD-2 surface. Docking studies were performed with Glide[57] using the crystal structure of hMD-2 and mMD-2. In the best predicted MD-2/PTX binding poses, the benzamido group of PTX is very close to Phe126, suggesting that a π-stacking interaction may exist between the two aromatic groups. In addition, Lys125 side chain establishes hydrophobic interaction with the phenyl ring. Another key interaction, cation-π, is formed between the phenyl group of PTX and Lys122, which is the only different amino acid in the MD-2 species-conserved sequence Phe119-Gly123. In MD-2, multiple interactions attract the Gly123-Lys130 loop that thus forms a concave surface facing the docked PTX. The same loop in the mouse protein is oriented in the opposite direction. The presence of a glutamic acid in lieu of Lys122 in mMD-2 leads to a completely different binding pose, possibly due to the absence of the cation-π interaction.[55] Another work by Resman *et al.* suggests a similar binding mode for

paclitaxel and the analogue docetaxel, on the basis of docking performed with Autodock in hMD-2 (PDB ID 2E59).[54] In this case also, the most favorable docked binding poses of both taxanes had the benzoyl group towards the nearby region formed by Ile61, Phe76, Leu78, Phe119 and Phe151 of hMD-2.

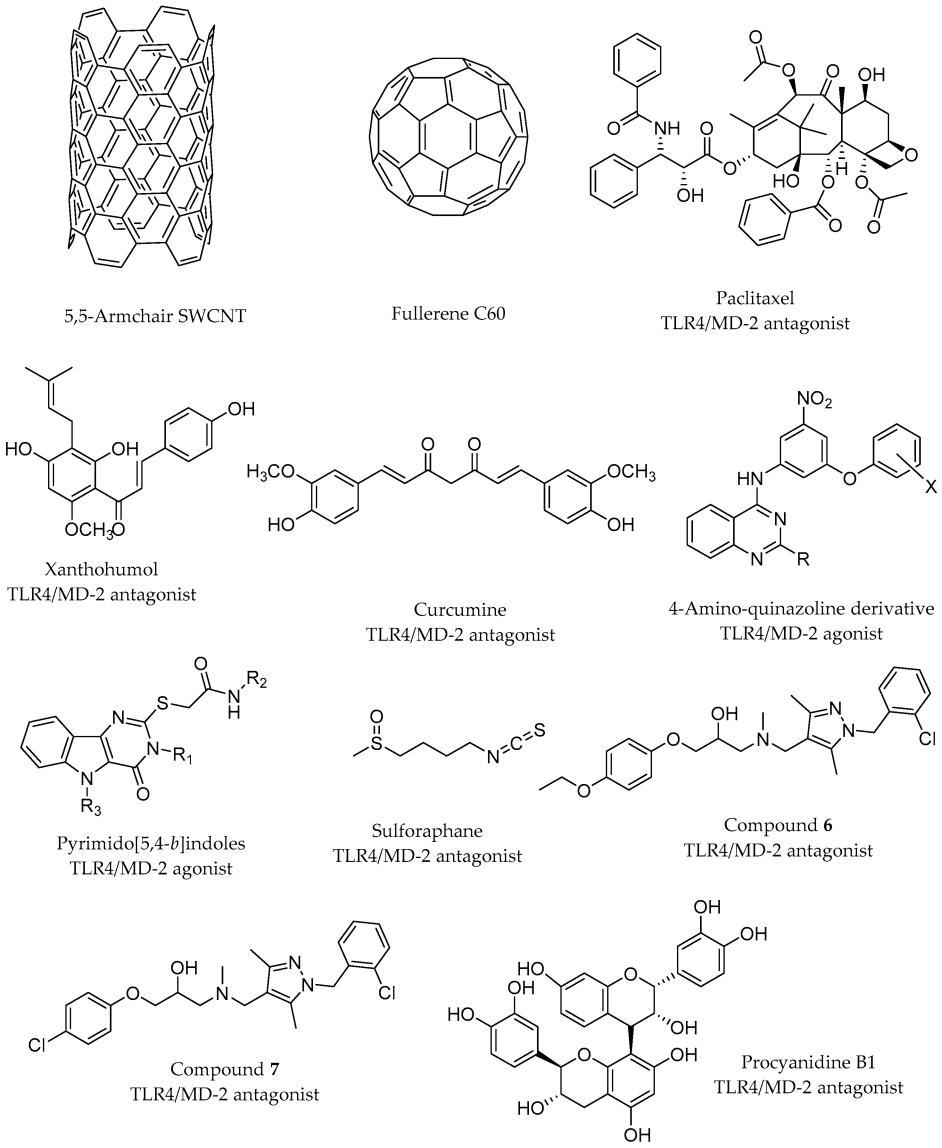


Figure 1.9. Non-LPS-like TLR4/MD-2 modulators studied by computational approaches.

A Glide-calculated docked binding mode for a prenylated chalcone-type (Xanthohumol; Figure 1.9) into the antagonist conformation of hMD-2 (PDB ID 2E59) was reported by Fu *et al.*[58] The results highlighted the importance of the H-bonds between OH groups from xanthohumol and residues Arg90 and Tyr102. Another H-bond between the OH group of the phenolic group and Glu92 was identified. However, this interaction rapidly broke during the 50-ns subsequent MD simulation, leading to a final

MD-2/xanthohumol complex stabilized by the above-mentioned interactions. An analog behavior was found for curcumin (Figure 1.9) from Autodock docking calculations in the same crystal structure of hMD-2. The hMD-2/curcumin complex obtained from the docking was subjected to MD simulations leading to a pose stabilized by interactions with Arg90 and Tyr102. Accordingly, experimental studies of R90A and Y102A MD-2 mutants pointed toward a direct binding of curcumin to MD-2 in the same binding site as LPS. In their model, curcumin occupies a large part of the hydrophobic pocket and forms H-bonds with residues Arg90 and Tyr102, stable along the simulation trajectory. Once again, the H-bond with Gly92 was broken during the simulation. MD simulations revealed that the presence of the ligand stabilizes the complex. Indeed, apo MD-2 suffers an important conformational change that reduces the volume of its hydrophobic cavity entrance. This finding is in agreement with other similar MD simulations performed on apo MD-2.[59-60] On the contrary, bond MD-2 displayed good stability along the simulations.[61]

Novel TLR4 selective ligands and potent NF-κB activators were identified through cell-based high throughput screening (HTS), namely substituted pyrimido[5-4-b]indole derivatives[62] and 4-amino-quinazolines[10]. From the former family, one hit compound was selected (Figure 1.9; R1 = phenyl, R2 = cyclohexyl, R3 = H). A series of pyrimido[5,4-b]indole rings with carboxamides substituted with various alkyl, cycloalkyl, aromatic and heteroaromatic groups were synthesized and biologically tested in order to perform SAR analysis. One of the most active compounds (Figure 1.9; R1 = phenyl, R2 = 3,3-dimethylbutyl, R3 = H) was docked in mouse TLR4/MD-2. The ligand was predicted to bind within the LPS-binding pocket forming H-bonds with residues TLR4 Glu439 and MD-2 Arg90, and multiple hydrophobic interactions with the side chains of Leu87, Phe126, Ile124, Phe121, and Phe119 from MD-2.

A 4-amino-quinazoline (Figure 1.9; R = COOEt, X = H) was identified from the second HTS with selective agonist activity on human over mouse TLR4/MD-2.[10] The docking calculations of this 4-amino-quinazoline with TLR4/MD-2 showed that it establishes hydrophobic interactions with Phe119, 121, 126 and Leu87 and makes H-bonds with residues Gln436 and Glu439 of TLR4 and Arg90 of MD-2. Two polar nitro oxygens from the compound were reported to interact with the nitrogen backbone of Ile124 and Lys122 of MD-2. The results from the computational study underlined the importance

of Lys122, which is a glutamic acid in mouse that could produce electrostatic repulsion with the nitro group. This could account for the decreased activity measured in mTLR4/MD-2.[10] Several analogs were synthesized for a SAR analysis, to confirm the relevant role of the nitro group for the binding and to guide further optimization of the lead compound.

Liquid chromatography-mass spectrometry analysis permitted to unveil that sulforaphane (SFN; Figure 1.9) forms a covalent bond with residue Cys133 of hMD-2. Covalent docking methods were applied in an attempt to explain the propensity of SFN to impair LPS engagement with the MD-2 hydrophobic pocket. The authors proposed a model in which SFN, once covalently linked to Cys133, occupies the same position as the R3" lipid chain of LPS (cf. PDB ID 3FXI; Figure 1.4b) and XA2 lipid chain of lipid IVa (cf. PDB ID 2E59). In their model, SFN is found in close proximity with residues Ile46, Phe76, Phe147, Phe151, Val135 and Leu149 of MD-2. This model suggests that SFN sterically prevents other LPS/lipid A from approaching and settling inside MD-2 hydrophobic pocket.[63] The same mechanism was reported for caffeic acid phenethyl ester through experimental methods.[64]

A series of compounds built by functionalizing pyrazole rings was reported by Bevan *et al.* to inhibit TLR4 activation.[65] Experimental studies promoted compounds **6** and **7** (Figure 1.9) to lead inhibitors. These compounds were used for docking studies against TLR4 (using the 3D coordinates extracted from the PDB ID 2Z65). The results indicate that both compounds independently bind at the surface of TLR4 where a protruding loop of MD-2 is normally found in the crystal structure. These predicted binding modes suggest that these compounds compete with MD-2 for binding TLR4, thus preventing or impairing the formation of the TLR4/MD-2 complex, resulting in a TLR4 unable to carry out its innate immunity role.

Polyphenol procyanidin B1 (Figure 1.9) is able to regulate innate and adaptive immunity by, inter alia, impairing LPS-induced inflammatory responses in human monocytes.[66-68] An attempt to explain its mode of action, combining experimental and docking studies, was reported.[69] The predicted binding pose presents similar interactions between the ligand and TLR4/MD-2 that the ones LPS establishes with TLR4/MD-2 in the crystal structure (PDB ID 3FXI). For instance, one of the phosphate groups of LPS forms a H-bond with Ser118 of MD-2 where procyanidin B1 is predicted

to form a hydrogen bond with Ser120, which is in close proximity of Ser118. In turn, the binding mode proposed by the authors suggests that procyanidin B1 impairs TLR4 signaling by successfully competing with LPS at binding inside the hydrophobic pocket of MD-2.

1.3 Computational studies of TLR4 mechanism

1.3.1 Computational studies of the TLR4/MD-2 ectodomain

Several studies in the literature are focused on the extracellular domain of TLR4 in complex with MD-2, investigating both its ability to recognize lipid A and lipid IVa and its dimerization/activation mechanism.

Garate *et al.* reported short MD simulations of apo-MD-2, TLR4/MD-2 dimer, MD-2/lipid A, MD-2/lipid IVa and TLR4/MD-2/lipid A complexes.[70] Their results highlight the hydrophobicity of the MD-2 pocket and its ability to close promptly in an aqueous environment. The H1 region, the helix connecting MD-2 with TLR4, was reported as a major actor of this event. From these results, the authors suggest a possible equilibrium between the open and the closed states of MD-2. Additionally, they noted that although the presence of TLR4 reduces the fluctuation of MD-2 it does not prevent its hydrophobic pocket to close. Charged phosphates were shown to play a key role in the early recognition of lipids impacting the formation of the heterotetramer. The MD simulations performed on the TLR4/MD-2/lipid A complex also showed that the presence of the ligand energetically stabilizes the complex, indicating cooperation in the binding process.

Evidence of the plasticity of MD-2 has also been observed by DeMarco *et al.* over several MD simulations performed in complex with variably-acylated lipid A molecules from *E. coli* and *N. meningitides* gram-negative bacteria.[71] The information gained from these 50-ns simulations led to the conclusion that the level of acylation of these ligands greatly influences the final architecture of the dimerization interface, shaping the conformation of the TLR4/MD-2 system into an agonist or an antagonist.

Paramo *et al.* performed MD simulations of at least 100 ns on the TLR4/MD-2 system in complex with different ligands, in which Phe126 was observed to transition from a closed (agonist/active conformation) to an open (antagonist/inactive conformation) state in the presence of lipid IVa, Eritoran and in the apo-form of the receptor.[59] The dimerization interface between the two heterodimers (TLR4/MD-2/TLR4*/MD-2*) was destabilized in agonist-free systems, due to the opening of Phe126, which disrupts the arrangement of nearby side chains containing Val82, Met85 and Leu87 of MD-2. These

simulations are in agreement with NMR studies pointing at the re-orientation of the Phe126 aromatic side chain induced by the binding of hexa-acylated endotoxin.[72]

MD simulations of TLR4 alone, MD-2 alone, TLR4/MD-2 complex and TLR4/MD-2/TLR4*/MD-2* complex were reported by de Aguiar *et al.*[73] The simulations of the TLR4 ectodomain revealed pronounced conformation and structural alterations in the N- and C-terminal domains, showing higher RMSD values compared to the overall protein RMSD values. Furthermore, over 100 ns of MD simulation, the distance between the N-terminal and the C-terminal regions increased from 5.7 Å to -10.9 Å, suggesting a straightening of the TLR4 curvature. In the MD simulations of the TLR4/MD-2 complex, these fluctuations and deformations were lessened, indicating a stabilizing role of MD-2. MD simulations of MD-2 alone showed high mobility of the loops, especially the one containing Lys109 and the region comprising residues Lys55 and Lys58. Interestingly, the Lys55-Lys58 region does not interact directly with TLR4, as one can observed in the crystal structure of the TLR4/MD-2 complex (PDB ID 3FXI). However, throughout the MD simulation of the TLR4/MD-2/TLR4*/MD-2* complex, MD-2 underwent structural rearrangements and interacted with TLR4 and TLR4*, reinforcing the idea of a stabilizing role of MD-2 for the TLR4 complexation.

As already reported in section *1.2.1 Natural LPSs*, Anwar *et al.* performed computational studies of the TLR4 signaling mechanism by studying the species-specific behavior of TLR4/MD-2 in the recognition of *Rs*LA (Figure 1.7).[40] In addition to the docking, the authors also reported 25-ns MD simulations of the docked complexes. Over the simulation, they monitored the local and global mobility, the surface accessible solvent area of the ligand and the surface charge distributions of TLR4 and MD-2. The GlcN1-GlcN2 backbone was shown to adopt an agonist-like conformation in horse and hamster TLR4/MD-2 and an antagonist-like conformation in human and murine TLR4/MD-2. Additionally, the Phe126 MD-2 loop, from residue 123 to residue 129, containing the on/off switch Phe126, proved to be less stable in the human and the murine complex, than in the horse and the hamster ones. The RMSD of the MD-2 loop from residue 81 to residue 89, which interact with TLR4* thus mediating the dimerization event, showed greater variations in humans and mice than in horses and hamsters. These data suggest a relationship between the flexibility of both loops (residues 81-89 and residues 123-129) and the agonist/antagonist activity of the

ligand and provide a plausible explanation for the species-specific behavior of *Rs*LA regarding TLR4 activation.

Computational strategies were also applied to study TLR4 and MD-2 mutants. In a 2009 study, Slivka *et al.* used the Rosetta software[74] to compare the binding energy of a truncated MD-2 with the original one.[75] MD-2 was truncated (termed MD-2-I) to keep only the residues identified as playing a major role in maintaining the TLR4/MD-2 heterodimer stability. The docking experiment was performed targeting both a partial human TLR4 retrieved from the Protein Data Bank (PDB ID 2Z65) and a full-length TLR4 humanized model built by mutating the residues at the TLR4/MD-2 heterodimer interface in the mouse crystal structure (PDB ID 2Z64) into their human counterparts (TLR4: F160L, G234N, K263R, D264N, T290A; MD-2: H96R, H98R). In the first case, the affinity of MD-2-I was found higher than the one of the full-length MD-2. When docked against the human TLR4 model, MD-2-I exhibited a lower affinity than the full-length MD-2. Altogether, these results indicate that MD-2-I is theoretically able to bind TLR4 and might even compete with the full-length MD-2. This was confirmed by cell assay experiments showing that the addition of MD-2-I abolishes cell responsiveness to LPS stimulation. Flow cytometry analyses on HEK293 cells transfected with all proteins involved in the TLR4 activation pathway incubated with LPS covalently linked to fluorescein isothiocyanate indicate that MD-2-I impedes TLR4/MD-2 dimerization. The SEAP assay shows that MD-2-I also alters downstream signaling.

Recently, the critical role of residue Val135 of MD-2, located deeply inside the hydrophobic pocket, was reported by Vasl *et al.*[60] hMD-2 has the ability to bind LPS in the absence of TLR4, while mMD-2 is responsive to LPS only when engaged in a complex with TLR4. Site-directed mutagenesis was applied to hMD-2 to mutate Val135 to its murine alanine counterpart. This single point mutation led to a mutant V135A hMD-2 lacking the ability to bind LPS. A series of 50-ns MD simulations of the WT hMD-2 and the V135A mutant hMD-2 in solution and in complex with TLR4 was performed to study the conformational changes. In the case of the WT hMD-2, the authors reported an abrupt decrease of the SASA and volume in the first nanoseconds of the simulation, describing it as a hydrophobic collapse. This phenomenon was not observed in the V135A systems, suggesting that Val135 is primordial to confer

plasticity to MD-2. This tendency was confirmed by another simulation of MD-2 in complex with three myristic acids (as reported in PDB ID 2E26). The V135A mutant hMD-2 needed a much longer simulation time to adapt its shape to the three myristic acids than the wild type. The authors concluded that this loss of plasticity could incapacitate hMD-2 from binding LPS.[60]

1.3.2 Computational studies on the intracellular domain of TLR4

The intracellular domain of the TLR4 transmembrane protein contains a TIR homology domain, which is a common feature of all adaptors involved in the initiation of TLR4 signaling, mediating protein-protein interactions between the TLR4 and the signal transduction components. TLR4 has two distinguished signaling pathways involving primarily four TIR-domain-containing adaptors. In the first pathway, the MyD88 adapter-like (Mal) acts as a "sorting" adaptor by recruiting the myeloid differentiation primary response gene 88 (MyD88), the "signaling" adaptor, to the plasma membrane. In the second pathway, the TRIF-related adaptor molecule (TRAM) plays the role of "sorting" adaptor, which recruits the TIR-domain-containing adapter-inducing interferon-β (TRIF), the "signaling" adaptor, to the membrane to initiate the signal. As a major component of these adaptors, the TIR domain is believed to play a central role in the recruitment processes.[76-77]

The crystal structures of human TLR1 (PDB ID 1FYV) and TLR2 (PDB ID 1FYW) revealed the structural basis of the TIR domain[78] followed by the crystal structure of TLR10 TIR domain (PDB ID 2J67)[79] and the solution structure of MyD88 TIR domain resolved by NMR (PDB ID 2JS7 and 2Z5V).[80] Prior to that release, two homology models of the TIR domain of MyD88 were reported. Both were built based on the TLR2 TIR domain crystal structure (PDB ID 1FYW) resolved by X-ray crystallography.[81-82] In 2012, the crystal structure of Mal was also resolved by X-ray crystallography (PDB ID 3UB2).[83]

The lack of structural information for the TIR domain of TLR4 has driven the creation of models to clarify the recruitment of adaptors from a structural perspective. Dunne *et al.* built monomer models of TLR4, Mal and MyD88 using comparative modeling and loop refining techniques.[84] They noted differences in the electrostatic surface potentials suggesting that adaptor binding is driven by electrostatic complementarity. This point was also emphasized in a study by Kubarenko *et al.* in which they compared the surface

charges of TIR domains of the crystal structure of hTLR2 and of the models of hTLR3 and hTLR4 and noted that the surface charge distribution of the BB loop and the αC-helix (Figure 1.10) present similarities in TLR2 and TLR4 and differ between TLR3 and TLR4.[85] The authors considered that these findings could explain why TLR2 and TLR4 recruit MyD88, whereas TLR3 does not. In their computational study Gong *et al.* highlighted that, whereas the BB-loop is highly conserved among TIR-domains, the APBS electrostatic surfaces differ.[86] The authors hypothesized that this finding might explain the specificity and selectivity of adaptors recruitment. An experimental study showed that a single point mutation in the TIR domain of murine TLR4 (P712H) renders the system hypo-responsive to LPS stimulation. The authors noted that their data does not suggest a direct role for this residue.

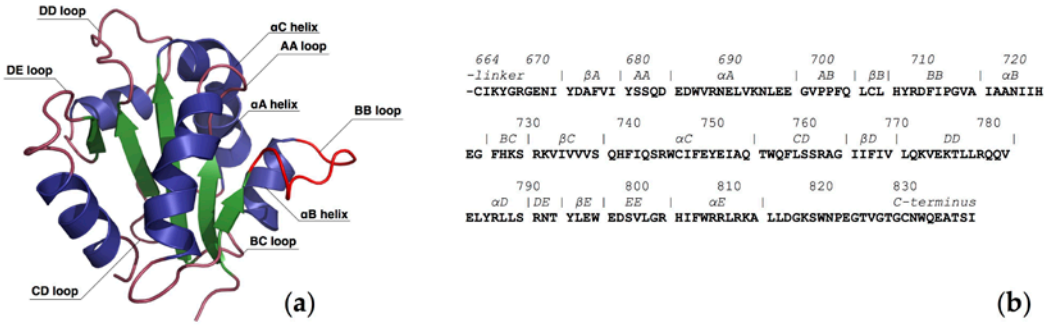


Figure 1.10. Intracellular TIR domain of TLR4. (**a**) 3D representation of the homology model with details on its structural composition; (**b**) FASTA sequence divided by its secondary structure elements.

Dunne *et al.* used a docking procedure based on hydrophobicity and geometry.[84] Their results suggest that Mal and MyD88 bind at two distinct, non-overlapping, binding sites: the DD- and DE-loops of Mal forming interactions with the BB-loop and αC helix of TLR4-TIR domain and the AA- and DD-loop of MyD88 with the CD-loop of TLR4 (Figure 1.10). The biological relevance of this binding mode was later questioned, as it was discovered that TLR4 activation required homodimerization. In line with that, in 2007, Miguel *et al.* reported the first 3D model of the dimer of the TIR domain of TLR4; a dimer composed of two identical subunits, arranged in a two-fold axis of symmetry (Figure 1.11a).[87] Despite the observation that some loops are differently oriented, the overall monomeric fold and the secondary structure of each subunit are very similar to the monomer model reviewed above.[84] This dimer model outlines significant interactions between the BB-loops of each monomer; for instance, residues Phe712 are engaged in homotypic aromatic interactions. A flat, but slightly curved

surface, was observed and attributed to the side facing the membrane. The authors also reported a docking study of TRAM and Mal with the TLR4 dimeric model in which the two adaptors bind at either sides of the dimer interface formed by the union of the two TLR4-TIR domains, which are identical due to the symmetry. They noted that both adaptors are forming strong interactions with TLR4 Trp757. Mal is also interacting with His728, Arg763 and Lys819, whereas TRAM interacts with Glu684, Arg780 and Glu824. The residues of the adaptors found at the TLR4 interface are mostly located on the BB-loop suggesting that the BB-loop of all three TIR-containing structures is of critical importance for binding specificity and selectivity.

Gong *et al.* performed a docking study based on the geometry, hydrophobicity and electrostatic complementarity of the molecular surface reporting a dimeric model different from the model described above.[86] The interface is formed by residues Pro714 to Ala717 from the BB-loop of one monomer protruding into a groove formed by residues Cys747 to Ile748 from the αC of the other monomer, and vice versa (Figure 1.11b). In another study Basith *et al.* used in silico approaches (homology modeling, protein-protein docking and 5.5-ns MD simulations) to investigate the inhibitory effect of ST2L toward TLR4 activation. ST2L (IL-33r) is a member of the Toll-like/IL-1 receptor superfamily known to negatively regulate MyD88-dependent signaling pathway.[88] The authors reported a TLR4-TLR4 homodimer model very similar to the first one reported here[87] (as shown in Figure 1.11a), and their docking study also gave a similar binding mode for Mal (at each side of the dimer). Their results indicate that MyD88 is recruited by Mal, and that ST2L prevents the recruitment of MyD88 by binding at the Mal interface. Thus, according to these results, ST2L successfully competes with MyD88 to bind at the Mal interface.

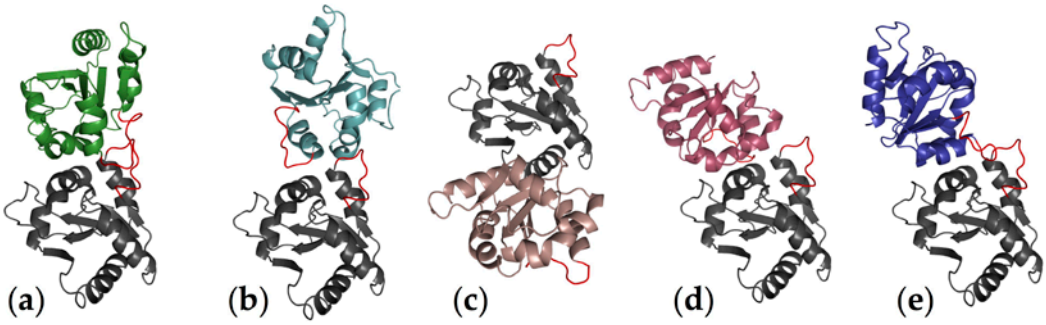


Figure 1.11. Representation of the different ways the dimer is proposed by published computational strategies to be assembled in the literature by computational strategies. (**a**) First reported by Miguel *et al.*[87]; (**b**) reported by Gong *et al.*[86]; (**c**-**e**) reported by Guven-Maiorov *et al.*[56]. The monomer has been built by homology modeling, and the secondary structure representation has been altered to resemble the other models. The dimers have been assembled manually, fitting as best as possible the schemes present in each paper, to provide an overview of the variety of binding poses reported so far. The dimmers shown do not have the pretention of being as precise as those shown in the original papers and should be considered schematic.

In a later study, Bovijn *et al.* reported a homology model constructed based on the crystal structure of the dimeric TLR10 TIR domain.[57] This model is also in agreement with the first model reported.[87] The authors proposed that Mal and TRAM adaptors are competing for binding an extended site formed by the reunion of two TLR4 intracellular domains. An experimental mutation study showed that all mutations that impaired Mal binding also impaired TRAM binding, strengthening the idea that Mal and TRAM bind to the same molecular surface. They define the TLR4/TLR4* dimer interface as binding site II, composed of residues from the BB-loop, DD-loop and αC (Figure 1.10). Then, they describe that the binding site for TRAM and Mal is formed by the reunion of two sites I (as defined in the study: residues from αA αB BB and BC), which is in disagreement with the binding site proposed by Miguel *et al.*[87]. The authors thus argue that their model is supported by experimental data and residue conservation analysis. The binding site III is defined as being located at the opposite direction of the binding site I and might be implicated in the interferon regulatory factor 3 (IRF-3) activation.

Singh *et al.* studied the importance of the highly conserved β-sheets among TLRs' TIR domain and revealed their primordial implications in the communication network. MD simulations of 100 ns of models based on sequence similarity were performed.[89] MD simulations were used to study the long-range interactions between residues separated by at least 20 residues in the sequence. They reported interactions between the backbone atoms of the first β-sheet with the BB-loop and the third β-sheet. The authors

identified four interacting hubs mainly constituted of hydrophobic residues. Among them, three are in the β-sheets just before the BB-loop, the αC helix and the DD-loops, stressing their role in TIR/TIR interaction. This hypothesis was further supported by analyzing the mutations known to completely abrogate signaling. They show that mutants IFI767-769AAA and L815A disturb the interacting network, thus explaining the impaired TIR domain homodimerization capacity. In a very recent paper by Guven-Maiorov *et al.*, the authors used computational techniques to describe the architecture of the signalosome of TLR4.[56] They built three models of the intracellular part of the TLR4 protein (Figure 1.11c-e). These three dimer models are all unprecedented despite that the secondary structure of the monomer is in great agreement with all of the published models. Furthermore, the authors used two of their models (Figure 1.11c-d) to propose different binding modes for Mal.

1.4 Objectives

TLR4 attracted lots of attention for the finding of new modulators with important applications in biomedicine. Several new compounds modulating TLR4 are undergoing preclinical and clinical evaluation, for the treatment of sepsis, inflammatory diseases, rheumatoid arthritis, and as vaccines and cancer immunotherapeutic. However, a TLR4 modulator to effectively treat septic shock is still yet to be discovered and approved for commercialization. Also, TLR4 agonists are sought to develop co-adjuvants for antitumoral treatments. Elucidating the molecular determinants that make a given molecule to be an agonist or an antagonist of TLR4, and understanding the mechanism of the TLR4/MD-2 system, would greatly help the design of new TLR4 modulators.

The relatively recent elucidation of the X-ray crystallographic structure of the extracellular domain of TLR4/MD-2 has opened new perspectives for the research around this challenging receptor. Towards that end, this thesis can be divided into three major parts.

The first one is to assist, by computational techniques, the design of synthetic LPS-like and non LPS-like TLR4 modulators by fine-tuning their relative agonist or antagonist potency through subtle molecular changes. An important effort is made to predict their effect on TLR4 and to assess their mode of action.

A second part is to computationally explain the effect of natural compounds (LPSs) by understanding how they interact selectively with some component of the TLR4 activation pathways, most relevantly MD-2 in complex with TLR4, MD-2 alone, and CD14. To unravel atomic details about the molecular recognition mechanism of the receptor and about the ligand-receptor interactions of these natural modulators by applying molecular modeling and computational chemistry techniques.

A third part is dedicated to gaining a deeper understanding on the molecular aspects of TLR4 activation and signaling by computational approaches. To clarify how minute molecular rearrangements on the ectodomain of TLR4 in complex with MD-2 is translated into intracellular signaling. This part includes the transmembrane domain of TLR4 and its intracellular domain as they both play an important role in the signal transmission. Apart from the fundamental knowledge they provide, these findings can guide the future development of novel agonists and antagonists of the TLR4/MD-2

system with promising biomedical applications in sepsis, inflammation, vaccines and cancer immunotherapy, among others.

As a side note, this work took place within an interdisciplinary European consortium (http://www.tollerant.eu) of 13 laboratories and industrial partners spanning 6 countries and cumulating a broad expertise in chemistry, biology, biophysics, biochemistry, and pharmacology.

Bibliography

1. Tobias, P. S.; Soldau, K.; Ulevitch, R. J. Isolation of a lipopolysaccharide-binding acute phase reactant from rabbit serum. *J. Exp. Med.* **1986,** *164* (3), 777-793.
2. Tobias, P. S.; Soldau, K.; Ulevitch, R. J. Identification of a lipid A binding site in the acute phase reactant lipopolysaccharide binding protein. *J. Biol. Chem.* **1989,** *264* (18), 10867-10871.
3. Schumann, R. R.; Flaggs, G.; Gray, P.; Wright, S.; Mathison, J.; Tobias, P.; Ulevitch, R. Structure and function of lipopolysaccharide binding protein. *Science.* **1990,** *249* (4975), 1429-1431.
4. Eckert, J. K.; Kim, Y. J.; Kim, J. I.; Gürtler, K.; Oh, D.-Y.; Sur, S.; Lundvall, L.; Hamann, L.; van der Ploeg, A.; Pickkers, P. The crystal structure of lipopolysaccharide binding protein reveals the location of a frequent mutation that impairs innate immunity. *Immunity.* **2013,** *39* (4), 647-660.
5. Setoguchi, M.; Nasu, N.; Yoshida, S.; Higuchi, Y.; Akizuki, S. i.; Yamamoto, S. Mouse and human CD14 (myeloid cell-specific leucine-rich glycoprotein) primary structure deduced from cDNA clones. *Biochim. Biophys. Acta.* **1989,** *1008* (2), 213-222.
6. Kim, J.-I.; Lee, C. J.; Jin, M. S.; Lee, C.-H.; Paik, S.-G.; Lee, H.; Lee, J.-O. Crystal structure of CD14 and its implications for lipopolysaccharide signaling. *J. Biol. Chem.* **2005,** *280* (12), 11347-11351.
7. Kelley, S. L.; Lukk, T.; Nair, S. K.; Tapping, R. I. The crystal structure of human soluble CD14 reveals a bent solenoid with a hydrophobic amino-terminal pocket. *J. Immunol.* **2013,** *190* (3), 1304-1311.
8. Lemaitre, B.; Nicolas, E.; Michaut, L.; Reichhart, J.-M.; Hoffmann, J. A. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. *Cell.* **1996,** *86* (6), 973-983.
9. Poltorak, A.; He, X.; Smirnova, I.; Liu, M.-Y.; Van Huffel, C.; Du, X.; Birdwell, D.; Alejos, E.; Silva, M.; Galanos, C. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science.* **1998,** *282* (5396), 2085-2088.
10. Nour, A.; Hayashi, T.; Chan, M.; Yao, S.; Tawatao, R. I.; Crain, B.; Tsigelny, I. F.; Kouznetsova, V. L.; Ahmadiiveli, A.; Messer, K. Discovery of substituted 4-aminoquinazolines as selective Toll-like receptor 4 ligands. *Bioorg. Med. Chem. Lett.* **2014,** *24* (21), 4931-4938.
11. Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature.* **2009,** *458* (7242), 1191.
12. O'neill, L. A.; Golenbock, D.; Bowie, A. G. The history of Toll-like receptors—redefining innate immunity. *Nat. Rev. Immunol.* **2013,** *13* (6), 453.
13. Akira, S.; Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **2004,** *4* (7), 499.
14. Beutler, á. TLR4 as the mammalian endotoxin sensor. In *Toll-like receptor family members and their ligands*, Springer: 2002; pp 109-120.
15. Klett, J.; Reeves, J.; Oberhauser, N.; Perez-Regidor, L.; Martin-Santamaria, S. Modulation of toll-like receptor 4. Insights from x-ray crystallography and molecular modeling. *Curr. Top. Med. Chem.* **2014,** *14* (23), 2672-2683.
16. Muta, T.; Takeshige, K. Essential roles of CD14 and lipopolysaccharide-binding protein for activation of toll-like receptor (TLR) 2 as well as TLR4. *FEBS J.* **2001,** *268* (16), 4580-4589.

17. Wright, S. D.; Ramos, R. A.; Tobias, P. S.; Ulevitch, R. J.; Mathison, J. C. CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein. *Science.* **1990,** *249* (4975), 1431-1433.
18. Akira, S.; Uematsu, S.; Takeuchi, O. Pathogen recognition and innate immunity. *Cell.* **2006,** *124* (4), 783-801.
19. Kawai, T.; Akira, S. Signaling to NF-κB by Toll-like receptors. *Trends Mol. Med.* **2007,** *13* (11), 460-469.
20. Lu, Y.-C.; Yeh, W.-C.; Ohashi, P. S. LPS/TLR4 signal transduction pathway. *Cytokine.* **2008,** *42* (2), 145-151.
21. Liu, Y.; Yin, H.; Zhao, M.; Lu, Q. TLR2 and TLR4 in autoimmune diseases: a comprehensive review. *Clin. Rev. Allergy Immunol.* **2014,** *47* (2), 136-147.
22. Alexander, C.; Rietschel, E. T. Invited review: bacterial lipopolysaccharides and innate immunity. *J. Endotoxin Res.* **2001,** *7* (3), 167-202.
23. Raetz, C. R.; Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **2002,** *71* (1), 635-700.
24. Hiruma, R.; Yamaguchi, A.; Sawai, T. The effect of lipopolysaccharide on lipid bilayer permeability of β-lactam antibiotics. *FEBS Lett.* **1984,** *170* (2), 268-272.
25. Raetz, C. R. Biochemistry of endotoxins. *Annu. Rev. Biochem.* **1990,** *59* (1), 129-170.
26. Molinaro, A.; Holst, O.; Di Lorenzo, F.; Callaghan, M.; Nurisso, A.; D'Errico, G.; Zamyatina, A.; Peri, F.; Berisio, R.; Jerala, R. Chemistry of lipid A: at the heart of innate immunity. *Chem. Eur. J.* **2015,** *21* (2), 500-519.
27. Billod, J.-M.; Lacetera, A.; Guzmán-Caldentey, J.; Martín-Santamaría, S. Computational approaches to toll-like receptor 4 modulation. *Molecules.* **2016,** *21* (8), 994.
28. Pérez-Regidor, L.; Zarioh, M.; Ortega, L.; Martín-Santamaría, S. Virtual screening approaches towards the discovery of Toll-Like receptor modulators. *Int. J. Mol. Sci.* **2016,** *17* (9), 1508.
29. Meng, J.; Lien, E.; Golenbock, D. T. MD-2-mediated ionic interactions between lipid A and TLR4 are essential for receptor activation. *J. Biol. Chem.* **2010,** *285* (12), 8695-8702.
30. Di Lorenzo, F.; Kubik, Ł.; Oblak, A.; Lorè, N. I.; Cigana, C.; Lanzetta, R.; Parrilli, M.; Hamad, M. A.; De Soyza, A.; Silipo, A. Activation of human Toll-like receptor 4 (TLR4)· myeloid differentiation factor 2 (MD-2) by hypoacylated lipopolysaccharide from a clinical isolate of burkholderia cenocepacia. *J. Biol. Chem.* **2015,** *290* (35), 21305-21319.
31. Maeshima, N.; Evans-Atkinson, T.; Hajjar, A. M.; Fernandez, R. C. Bordetella pertussis lipid A recognition by Toll-like receptor 4 and MD-2 is dependent on distinct charged and uncharged interfaces. *J. Biol. Chem.* **2015,** *290* (21), 13440-13453.
32. 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.* **2012,** *109* (19), 7421-7426.
33. Resman, N.; Vašl, J.; Oblak, A.; Pristovšek, P.; Gioannini, T. L.; Weiss, J. P.; Jerala, R. Essential roles of hydrophobic residues in both MD-2 and toll-like receptor 4 in activation by endotoxin. *J. Biol. Chem.* **2009,** *284* (22), 15052-15060.
34. Ohto, U.; Fukase, K.; Miyake, K.; Satow, Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science.* **2007,** *316* (5831), 1632-1634.
35. Scior, T.; Alexander, C.; Zaehringer, U. Reviewing and identifying amino acids of human, murine, canine and equine TLR4/MD-2 receptor complexes conferring

endotoxic innate immunity activation by LPS/lipid A, or antagonistic effects by Eritoran, in contrast to species-dependent modulation by lipid IVa. *Comput. Struct. Biotechnol. J.* **2013,** *5* (6), e201302012.
36. Kim, H. M.; Park, B. S.; Kim, J.-I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell.* **2007,** *130* (5), 906-917.
37. Scior, T.; Lozano-Aponte, J.; Figueroa-Vazquez, V.; Yunes-Rojas, J. A.; Zähringer, U.; Alexander, C. Three-dimensional mapping of differential amino acids of human, murine, canine and equine TLR4/MD-2 receptor complexes conferring endotoxic activation by lipid A, antagonism by Eritoran and species-dependent activities of Lipid IVA in the mammalian LPS sensor system. *Comput. Struct. Biotechnol. J.* **2013,** *7* (9), e201305003.
38. Ohto, U.; Yamakawa, N.; Akashi-Takamura, S.; Miyake, K.; Shimizu, T. Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *J. Biol. Chem.* **2012,** *287* (48), 40611-40617.
39. Han, J.; Kim, H. J.; Lee, S.-C.; Hong, S.; Park, K.; Jeon, Y. H.; Kim, D.; Cheong, H.-K.; Kim, H.-S. Structure-based rational design of a Toll-like receptor 4 (TLR4) decoy receptor with high binding affinity for a target protein. *PLoS One.* **2012,** *7* (2), e30929.
40. Anwar, M. A.; Panneerselvam, S.; Shah, M.; Choi, S. Insights into the species-specific TLR4 signaling mechanism in response to Rhodobacter sphaeroides lipid A detection. *Sci. Rep.* **2015,** *5*, 7657.
41. Webb, B.; Sali, A. Protein structure modeling with MODELLER. In *Protein Structure Prediction*, Springer: 2014; pp 1-15.
42. Walsh, C.; Gangloff, M.; Monie, T.; Smyth, T.; Wei, B.; McKinley, T. J.; Maskell, D.; Gay, N.; Bryant, C. Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. *J. Immunol.* **2008,** *181* (2), 1245-1254.
43. Irvine, K. L.; Gangloff, M.; Walsh, C. M.; Spring, D. R.; Gay, N. J.; Bryant, C. E. Identification of key residues that confer Rhodobacter sphaeroides LPS activity at horse TLR4/MD-2. *PLoS One.* **2014,** *9* (5), e98776.
44. Klett, J.; Núñez-Salgado, A.; Dos Santos, H. G.; Cortés-Cabrera, A. l.; Perona, A.; Gil-Redondo, R. n.; Abia, D.; Gago, F.; Morreale, A. MM-ISMSA: an ultrafast and accurate scoring function for protein–protein docking. *J. Chem. Theory Comput.* **2012,** *8* (9), 3395-3408.
45. Opal, S. M.; Laterre, P.-F.; Francois, B.; LaRosa, S. P.; Angus, D. C.; Mira, J.-P.; Wittebole, X.; Dugernier, T.; Perrotin, D.; Tidswell, M. Effect of eritoran, an antagonist of MD2-TLR4, on mortality in patients with severe sepsis: the ACCESS randomized trial. *JAMA.* **2013,** *309* (11), 1154-1162.
46. Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins.* **2009,** *77* (4), 778-795.
47. Cighetti, R.; Ciaramelli, C.; Sestito, S. E.; Zanoni, I.; Kubik, Ł.; Ardá-Freire, A.; Calabrese, V.; Granucci, F.; Jerala, R.; Martín-Santamaría, S. Modulation of CD14 and TLR4· MD-2 Activities by a Synthetic Lipid A Mimetic. *Chembiochem.* **2014,** *15* (2), 250-258.
48. Artner, D.; Oblak, A.; Ittig, S.; Garate, J. A.; Horvat, S.; Arrieumerlou, C. c.; Hofinger, A.; Oostenbrink, C.; Jerala, R.; Kosma, P. Conformationally constrained Lipid A mimetics for exploration of structural basis of TLR4/MD-2 activation by lipopolysaccharide. *ACS Chem. Biol.* **2013,** *8* (11), 2423-2432.

49. Piazza, M.; Calabrese, V.; Baruffa, C.; Gioannini, T.; Weiss, J.; Peri, F. The cationic amphiphile 3, 4-bis (tetradecyloxy) benzylamine inhibits LPS signaling by competing with endotoxin for CD14 binding. *Biochem. Pharmacol.* **2010,** *80* (12), 2050-2056.
50. Ciaramelli, C.; Calabrese, V.; Sestito, S. E.; Pérez-Regidor, L.; Klett, J.; Oblak, A.; Jerala, R.; Piazza, M.; Martín-Santamaría, S.; Peri, F. Glycolipid-based TLR4 Modulators and Fluorescent Probes: Rational Design, Synthesis, and Biological Properties. *Chem. Biol. Drug Des.* **2016,** *88* (2), 217-229.
51. Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **2006,** *34* (suppl_2), W116-W118.
52. Kawasaki, K.; Akashi, S.; Shimazu, R.; Yoshida, T.; Miyake, K.; Nishijima, M. Mouse toll-like receptor 4· MD-2 complex mediates lipopolysaccharide-mimetic signal transduction by Taxol. *J. Biol. Chem.* **2000,** *275* (4), 2251-2254.
53. Manthey, C.; Qureshi, N.; Stütz, P.; Vogel, S. Lipopolysaccharide antagonists block taxol-induced signaling in murine macrophages. *J. Exp. Med.* **1993,** *178* (2), 695-702.
54. Resman, N.; Gradišar, H.; Vašl, J.; Keber, M. M.; Pristovšek, P.; Jerala, R. Taxanes inhibit human TLR4 signaling by binding to MD-2. *FEBS Lett.* **2008,** *582* (28), 3929-3934.
55. Zimmer, S. M.; Liu, J.; Clayton, J. L.; Stephens, D. S.; Snyder, J. P. Paclitaxel binding to human and murine MD-2. *J. Biol. Chem.* **2008,** *283* (41), 27916-27926.
56. Guven-Maiorov, E.; Keskin, O.; Gursoy, A.; VanWaes, C.; Chen, Z.; Tsai, C. J.; Nussinov, R. The Architecture of the TIR Domain Signalosome in the Toll-like Receptor-4 Signaling Pathway. *Sci. Rep.* **2015,** *5*, 13128.
57. Bovijn, C.; Ulrichts, P.; De Smet, A.-S.; Catteeuw, D.; Beyaert, R.; Tavernier, J.; Peelman, F. Identification of interaction sites for dimerization and adapter recruitment in Toll/interleukin-1 receptor (TIR) domain of Toll-like receptor 4. *J. Biol. Chem.* **2012,** *287* (6), 4088-4098.
58. Fu, W.; Chen, L.; Wang, Z.; Zhao, C.; Chen, G.; Liu, X.; Dai, Y.; Cai, Y.; Li, C.; Zhou, J. Determination of the binding mode for anti-inflammatory natural product xanthohumol with myeloid differentiation protein 2. *Drug Des. Devel. Ther.* **2016,** *10*, 455.
59. Paramo, T.; Piggot, T. J.; Bryant, C. E.; Bond, P. J. The structural basis for endotoxin-induced allosteric regulation of the Toll-like receptor 4 (TLR4) innate immune receptor. *J. Biol. Chem.* **2013,** *288* (51), 36215-36225.
60. Vašl, J.; Oblak, A.; Peternelj, T. T.; Klett, J.; Martín-Santamaría, S.; Gioannini, T. L.; Weiss, J. P.; Jerala, R. Molecular basis of the functional differences between soluble human versus murine MD-2: Role of Val135 in transfer of lipopolysaccharide from CD14 to MD-2. *J. Immunol.* **2016,** *196* (5), 2309-2318.
61. Wang, Z.; Chen, G.; Chen, L.; Liu, X.; Fu, W.; Zhang, Y.; Li, C.; Liang, G.; Cai, Y. Insights into the binding mode of curcumin to MD-2: studies from molecular docking, molecular dynamics simulations and experimental assessments. *Mol. Biosyst.* **2015,** *11* (7), 1933-1938.
62. Chan, M.; Hayashi, T.; Mathewson, R. D.; Nour, A.; Hayashi, Y.; Yao, S.; Tawatao, R. I.; Crain, B.; Tsigelny, I. F.; Kouznetsova, V. L. Identification of substituted pyrimido [5, 4-b] indoles as selective Toll-like receptor 4 ligands. *J. Med. Chem.* **2013,** *56* (11), 4206-4223.

63. Koo, J. E.; Park, Z.-Y.; Kim, N. D.; Lee, J. Y. Sulforaphane inhibits the engagement of LPS with TLR4/MD2 complex by preferential binding to Cys133 in MD2. *Biochem. Biophys. Res. Commun.* **2013,** *434* (3), 600-605.
64. Kim, S. Y.; Koo, J. E.; Seo, Y. J.; Tyagi, N.; Jeong, E.; Choi, J.; Lim, K. M.; Park, Z. Y.; Lee, J. Y. Suppression of Toll-like receptor 4 activation by caffeic acid phenethyl ester is mediated by interference of LPS binding to MD2. *Br. J. Pharmacol.* **2013,** *168* (8), 1933-1945.
65. Bevan, D. E.; Martinko, A. J.; Loram, L. C.; Stahl, J. A.; Taylor, F. R.; Joshee, S.; Watkins, L. R.; Yin, H. Selection, preparation, and evaluation of small-molecule inhibitors of Toll-like receptor 4. *ACS Med. Chem. Lett.* **2010,** *1* (5), 194-198.
66. Byun, E.-B.; Sung, N.-Y.; Byun, E.-H.; Song, D.-S.; Kim, J.-K.; Park, J.-H.; Song, B.-S.; Park, S.-H.; Lee, J.-W.; Byun, M.-W. The procyanidin trimer C1 inhibits LPS-induced MAPK and NF-κB signaling through TLR4 in macrophages. *Int. Immunopharmacol.* **2013,** *15* (2), 450-456.
67. Terra, X.; Palozza, P.; Fernandez-Larrea, J.; Ardevol, A.; Blade, C.; Pujadas, G.; Salvado, J.; Arola, L.; Blay, M. T. Procyanidin dimer B1 and trimer C1 impair inflammatory response signalling in human monocytes. *Free Radical Res.* **2011,** *45* (5), 611-619.
68. Jung, M.; Triebel, S.; Anke, T.; Richling, E.; Erkel, G. Influence of apple polyphenols on inflammatory gene expression. *Mol. Nutr. Food Res.* **2009,** *53* (10), 1263-1280.
69. Xing, J.; Li, R.; Li, N.; Zhang, J.; Li, Y.; Gong, P.; Gao, D.; Liu, H.; Zhang, Y. Anti-inflammatory effect of procyanidin B1 on LPS-treated THP1 cells via interaction with the TLR4–MD-2 heterodimer and p38 MAPK and NF-κB signaling. *Mol. Cell. Biochem.* **2015,** *407* (1-2), 89-95.
70. Garate, J. A.; Oostenbrink, C. Lipid A from lipopolysaccharide recognition: structure, dynamics and cooperativity by molecular dynamics simulations. *Proteins.* **2013,** *81* (4), 658-674.
71. DeMarco, M. L.; Woods, R. J. From agonist to antagonist: Structure and dynamics of innate immune glycoprotein MD-2 upon recognition of variably acylated bacterial endotoxins. *Mol. Immunol.* **2011,** *49* (1-2), 124-133.
72. Yu, L.; Phillips, R. L.; Zhang, D.; Teghanemt, A.; Weiss, J. P.; Gioannini, T. L. NMR studies of hexaacylated endotoxin bound to wild-type and F126A mutant MD-2 and MD-2· TLR4 ectodomain complexes. *J. Biol. Chem.* **2012,** *287* (20), 16346-16355.
73. de Aguiar, C.; Costa, M. G.; Verli, H. Dynamics on human Toll-like receptor 4 complexation to MD-2: The coreceptor stabilizing function. *Proteins.* **2015,** *83* (2), 373-382.
74. Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D. Protein structure prediction using Rosetta. In *Methods Enzymol.*, Elsevier: 2004; Vol. 383, pp 66-93.
75. Slivka, P. F.; Shridhar, M.; Lee, G. i.; Sammond, D. W.; Hutchinson, M. R.; Martinko, A. J.; Buchanan, M. M.; Sholar, P. W.; Kearney, J. J.; Harrison, J. A. A peptide antagonist of the TLR4–MD2 interaction. *Chembiochem.* **2009,** *10* (4), 645-649.
76. O'Neill, L. A.; Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.* **2007,** *7* (5), 353.
77. Narayanan, K. B.; Park, H. H. Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis.* **2015,** *20* (2), 196-209.
78. Xu, Y.; Tao, X.; Shen, B.; Horng, T.; Medzhitov, R.; Manley, J. L.; Tong, L. Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature.* **2000,** *408* (6808), 111.

79. Hasan, U.; Chaffois, C.; Gaillard, C.; Saulnier, V.; Merck, E.; Tancredi, S.; Guiet, C.; Brière, F.; Vlach, J.; Lebecque, S. Human TLR10 is a functional receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. *J. Immunol.* **2005,** *174* (5), 2942-2950.
80. Ohnishi, H.; Tochio, H.; Kato, Z.; Orii, K. E.; Li, A.; Kimura, T.; Hiroaki, H.; Kondo, N.; Shirakawa, M. Structural basis for the multiple interactions of the MyD88 TIR domain in TLR4 signaling. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (25), 10260-10265.
81. Loiarro, M.; Sette, C.; Gallo, G.; Ciacci, A.; Fantò, N.; Mastroianni, D.; Carminati, P.; Ruggiero, V. Peptide-mediated interference of TIR domain dimerization in MyD88 inhibits interleukin-1-dependent activation of NF-κB. *J. Biol. Chem.* **2005,** *280* (16), 15809-15814.
82. Jiang, Z.; Georgel, P.; Li, C.; Choe, J.; Crozat, K.; Rutschmann, S.; Du, X.; Bigby, T.; Mudd, S.; Sovath, S. Details of Toll-like receptor: adapter interaction revealed by germ-line mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103* (29), 10961-10966.
83. Lin, Z.; Lu, J.; Zhou, W.; Shen, Y. Structural insights into TIR domain specificity of the bridging adaptor Mal in TLR4 signaling. *PLoS One.* **2012,** *7* (4), e34202.
84. Dunne, A.; Ejdebäck, M.; Ludidi, P. L.; O'Neill, L. A.; Gay, N. J. Structural complementarity of Toll/interleukin-1 receptor domains in Toll-like receptors and the adaptors Mal and MyD88. *J. Biol. Chem.* **2003,** *278* (42), 41443-41451.
85. Kubarenko, A.; Frank, M.; Weber, A. Structure–function relationships of Toll-like receptor domains through homology modelling and molecular dynamics. Portland Press Limited: 2007.
86. Gong, J.; Wei, T.; Stark, R. W.; Jamitzky, F.; Heckl, W. M.; Anders, H. J.; Lech, M.; Rössle, S. C. Inhibition of Toll-like receptors TLR4 and 7 signaling pathways by SIGIRR: a computational approach. *J. Struct. Biol.* **2010,** *169* (3), 323-330.
87. Miguel, R. N.; Wong, J.; Westoll, J. F.; Brooks, H. J.; O'Neill, L. A.; Gay, N. J.; Bryant, C. E.; Monie, T. P. A dimer of the Toll-like receptor 4 cytoplasmic domain provides a specific scaffold for the recruitment of signalling adaptor proteins. *PLoS One.* **2007,** *2* (8), e788.
88. Basith, S.; Manavalan, B.; Govindaraj, R. G.; Choi, S. In silico approach to inhibition of signaling pathways of Toll-like receptors 2 and 4 by ST2L. *PLoS One.* **2011,** *6* (8), e23989.
89. Singh, S.; Pandey, K.; Rathore, Y. S.; Sagar, A.; Pattnaik, U. B. K.; Ashish A communication network within the cytoplasmic domain of toll-like receptors has remained conserved during evolution. *J. Biomol. Struct. Dyn.* **2014,** *32* (5), 694-700.

CHAPTER 2

Computational methodology

2.1 Simulation of biomolecules

2.1.1 General introduction about computation of biomolecules.

Predicting and understanding biological processes and complex chemical reactions was one of the great challenges in the 1970s. While classical Newtonian physics was limited to analyzing molecules in their ground state, quantum physics was able to simulate the excited states of molecules in chemical reactions. However, the limitation of the latter was the absence of powerful computers capable of integrating the vast amount of data any larger protein would require.[1] Advances in this field, which resulted in the development of novel theoretical methods and more accurate algorithms, along with the increase of computer power, allowed accelerating the time-scale accessible by simulation techniques. Computational techniques can now provide an atomistic description of the molecular mechanisms for ligand recognition by biomolecules, enzymatic processes, solvation, folding events, conformational and allosteric transitions, among others.[2] Many of these algorithms combine classical and quantum physics principles to get a proper description of the (bio)chemical problem. These efforts have been acknowledged by the awarding of the Nobel Prize in Chemistry in 1998 to Kohn (for his development of the density-functional theory) and Pople (for his development of computational methods in quantum chemistry),[3] and in 2013 to Karplus, Levitt and Warshel for the development of multiscale models for complex chemical systems.[1, 4] In particular, the development of multiscale models of macromolecules has allowed the study of a wide variety of biological problems, such as protein folding and packing, prediction of macromolecular structures, protein-ligand interactions, protein energetics and theories of enzymatic mechanisms.[1] Various computational techniques are currently available to describe and predict the behavior of molecular biosystems on a wide distribution of length and time scales (Figure 2.1). Each method is appropriate for a particular range of length and time scales.

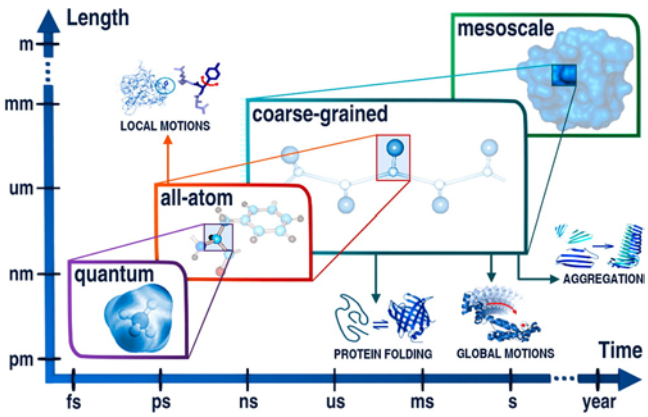


Figure 2.1. Different computational techniques for a variety of length and time scales (illustration retrieved from reference[5]).

2.1.2 Quantum mechanics methods: Hartree–Fock and DFT

Quantum mechanics (QM) is the fundamental theory of physics explaining the behavior of atoms and subatomic particles. Mathematically, the changes over time of such a system are described by the Schrödinger equations (eq. **[2-1]**).

$$i\hbar\frac{\partial}{\partial t}|\psi(\mathbf{r},t)\rangle = \hat{H}|\psi(\mathbf{r},t)\rangle \qquad \textbf{[2-1]}$$

In this equation, i is the imaginary unit, ħ is the reduced Planck constant, $\frac{\partial}{\partial t}$ indicates a partial derivative with respect to t, the time, Ψ is the wave function of the quantum system, $\mathbf{r}$ is the position vector, and $\hat{H}$ is the Hamiltonian operator, it characterizes the total energy of the system.

Hartree–Fock (HF) method is a method of approximation for the determination of the wave function and the energy of a quantum many-body system in a stationary state. The Schrödinger equation (eq. **[2-1]**) describes the quantum state in an exact way.

In analogy to classical systems, the Hamiltonian, $\hat{H}$, can be seen as the sum of the kinetic and the potential energy operators (eq. **[2-2]**).

$$\hat{H} = \hat{T} + \hat{V} \qquad \textbf{[2-2]}$$

The hardest part in electronic structure calculations is to deal with electron correlation i.e. the repulsion between pairs of electrons. In the HF method, the electron correlation is not treated exactly but in an average way. The Hamiltonian and the wave function are thus divided into one electron contribution and the electron interaction is then added

with an integral. This method is simpler but does not provide exact solutions to the Schrödinger equation.

Density functional theory (DFT) as developed by Kohn and Sham in 1965[6-7], is a computational QM modeling method to investigate the electronic structure principally the ground state of atomic systems. In the Kohn-Sham formulation, DFT can be viewed as a variant of HF theory in which the fundamental variable is the ground state electron density rather than the molecular orbitals. The density is then decomposed into Kohn-Sham orbitals which the only restriction is to provide a density that when integrated over all space should generate the appropriate number of electrons. For most practical purposes, the same basis set as in *ab initio* theory can be applied. From the DFT equations, we essentially obtain the electronic energy of the system (as in HF theory) corrected for the correlated interaction between electrons. This ingredient is missing at the HF level. In order to describe electronic interactions corrected for electron correlation, an extra potential term arises in DFT compared to HF, referred to as the exchange-correlation potential. The exact form of this is not known, but many derivations exist based on theoretical physics, chemical parametrization, and other approaches. The mathematical entities are termed exchange-correlation functionals, and a plethora of different DFT functionals of varying complexity and accuracy exist. An important choice is thus which functional to use. When it comes to the study of organic molecules, the B3LYP[8] functional is the most used. There is an extension of DFT to describe properties and dynamics of particles in presence of time-dependent potentials (termed TDDFT)[9], but that is beyond the scope of this thesis.

2.1.3 Molecular mechanics and molecular dynamics simulations

Atomic knowledge of biological structure can be obtained through a number of methods among which X-ray crystallography is the prevalent one. A crystallographic structure of a given protein brings excellent information about the spatial organization of each of the atoms (hydrogen excluded) of the amino acids sequence composing this protein at a given crystallized state.[10] From this three-dimensional atomic description, one can derive structural features such as the secondary structure of its different subdomains, their spatial relation one to another, known as the tertiary structure of the protein, and the possible arrangement of multiple folded proteins, known as the quaternary structure. In addition, a binding sites along with ligand binding mode can be revealed. Other experimental methods to gain insights into the spatial organization of the atoms

composing a protein are nuclear magnetic resonance spectroscopy (NMR) and cryogenic electron microscopy (Cryo-EM).[11-12] Atomic information can also be predicted through computational techniques such as fold prediction and comparative modeling. However, the majority of these methods only give a static picture of a protein, highly dependent on the experimental conditions, which often fails to thoroughly explain the function of the protein and the way this function is carried out. Nowadays, to overcome this limitation, internal motions and conformational transitions, key for understanding protein functional mechanism, can be modeled by molecular dynamics methods.

At the atomic scale, particle's behavior can be accurately described by the law of quantum chemistry considering molecular orbitals and the electron occupying them, from which one can derive important chemical properties (cf. *2.1.1 Quantum mechanics methods: Hartree–Fock and DFT*). However, such a precise description of an atomic system is very costly to be computed in systems the size of a protein. Interestingly, the apparent motion of the atoms governed by forces that arise from the quantum world can be rather accurately described through classical physics combining mechanical tools (springs, tensors, rotators) and electrostatics (Coulomb's law). The use of classical mechanics to model molecular system is called molecular mechanics. In this classical description, the atoms are represented as charged spheres, which size is usually proportional to the Van der Waals radius of the atom they describe. These spheres are connected both by direct linkage, representing chemical bonds, and by non-bonded interactions, comprising van der Waals and electrostatics interactions.

$$V(r) = \sum_{bonds} k_l\,(l - l_o)^2 + \sum_{angles} k_\theta\,(\theta - \theta_o)^2 + \sum_{torsions} \frac{1}{2} V_n\,[1 + \cos(n\omega - \gamma)] + \sum_{\substack{nonbond \\ pairs}} \left[\frac{A_{ij}}{{r_{ij}}^{12}} - \frac{C_{ij}}{{r_{ij}}^{6}} + \frac{q_j q_i}{4\pi\varepsilon_0 r_{ij}}\right] \quad \textbf{[2-3]}$$

In molecular mechanics, the potential energy of the system is given by equation **[2-3]**. This equation takes into account two main contributions, the covalent components, comprising bonds, angles, and torsions, and the noncovalent components containing the electrostatic and the van der Waals interactions, sometimes an extra term is added for the hydrogen bonds. In more details, the first term of Equation **[2-3]** addresses bond

stretching (i.e. intramolecular motion between two covalently bonded atoms), in which k_l is the force constant enclosing the energy cost relative to the displacement from the equilibrium value, l is the instantaneous bond length and l_0 the bond length at equilibrium. The second term describes bond angle vibrations (i.e. geometric distortions between three covalently bonded atoms A-B-C) written as a harmonic potential, with k_θ the force constant, θ the instantaneous angle and θ_0 the equilibrium angle. The third term represents the dihedral angle potential (i.e. the rotation around the central bond B-C in a covalently bonded sequence A-B-C-D). In case rotation needs to be restricted, e.g. to ensure planarity or to maintain the chirality of a certain group, improper torsion can be introduced, which is defined between atoms not connected in sequence. V_n is the dihedral constant, n the periodicity parameter, ω the instantaneous dihedral angle and γ the phase. The fourth term represents non-bonded or 'through-space' interactions between atom pairs, which can be decomposed into Lennard-Jones and Coulomb interactions, q_i and q_j being the respective charges of atoms i and j, r_{ij} the distance between the two atoms and *Aij* and *Bij* parameters for the repulsive and attractive components of the Lennard-Jones potential (Figure 2.2). A graphical representation of these components is given in Figure 2.3.

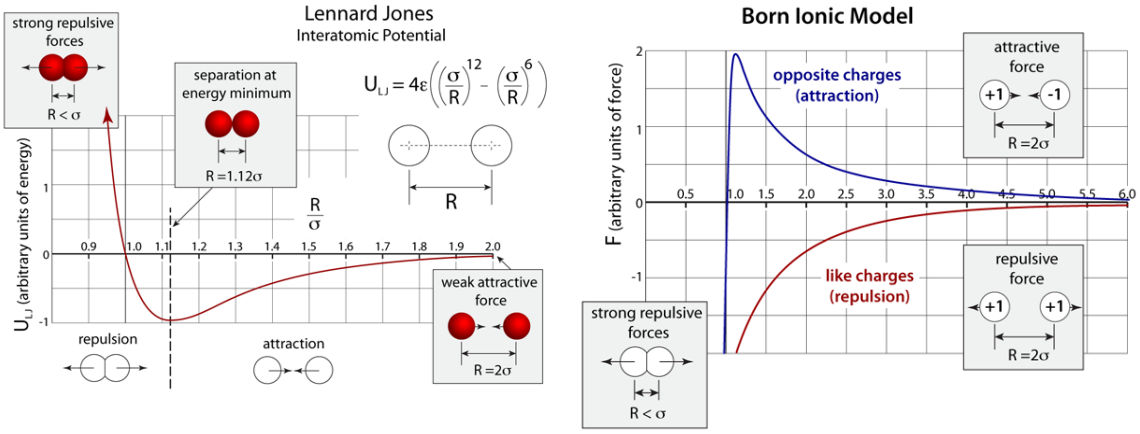


Figure 2.2. On the left: graph of the Lennard-Jones potential function in which regions of repulsion, on the left side of the graph, and attraction, on the right side of the graph are explicitly described. At lowest temperature distances between atoms tend toward the energy minimum. On the right: the born ionic-model describing the energy between two non-bonded charged partners (e.g. ions) in function of the distance separating them. Both the attraction of opposite charges and repulsion of like charges are shown (illustrations retrieved here[13]).

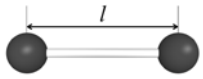
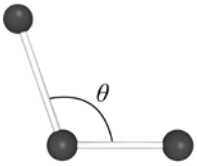
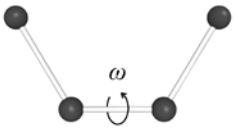
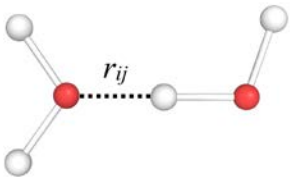
bond stretching	*angle bending*	*torsional rotation*	*non-bonded interactions*
			
$\sum_{bonds} k_l\,(l - l_o)^2$	$\sum_{angles} k_\theta\,(\theta - \theta_o)^2$	$\sum_{torsions} \frac{1}{2} V_n\,[1 + \cos(n\omega - \gamma)]$	$\sum_{\substack{nonbond \\ pairs}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} + \frac{q_j q_i}{4\pi\varepsilon_0 r_{ij}}\right]$

Figure 2.3. Schematic representation of the molecular mechanics potential function components. Atoms are symbolized by black, white, and red spheres and covalent bonds by white sticks.

In molecular dynamics, successive configurations of the system are generated by integrating Newton's equation of motion (equation **[2-4]**). The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

$$\mathbf{F} = m\mathbf{a} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2} \qquad \textbf{[2-4]}$$

The force **F** acted upon an atom is equal to the mass *m* of that atom multiplied by the acceleration **a** of the atom. The acceleration is also the first derivative of the velocity with respect to time $\left(\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right)$ and the second derivative of the position with respect to time $\left(\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2}\right)$.

In a molecular dynamics simulation, these forces are calculated at a given time over a very short period, called the time step (Δt), and the atoms are moved accordingly, then the forces are calculated for those new coordinates and so on until reaching the desired simulated time. Δt is computed using a simple Taylor expansion (equation **[2-5]**).

$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \frac{\mathrm{d}\mathbf{q}(\mathrm{t})}{\mathrm{d}t}\Delta\mathrm{t} + \frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2}\frac{\Delta t^2}{2} + \cdots \qquad [2\text{-}5]$$

One can see that the position $\mathbf{q}(t)$, velocity $\frac{\mathrm{d}\mathbf{q}(\mathrm{t})}{\mathrm{d}t}$ and acceleration $\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2}$, are sufficient for the propagation of the molecular system. The acceleration can be computed from equation **[2-4]** in which the force F is obtained by differentiating the energy of the system.[14]

An MD simulation is set up by assigning initial velocities and positions to all atoms in the system. The velocities are usually randomly assigned, whereas the positions are

typically resolved by one of the methods mentioned above. Thereafter, the force acting on each atom is calculated, giving the direction of movement. The atoms are moved in this direction, giving news forces on each atom, and the procedure is then repeated. Practically, this integration of motion can be treated by several methods such as leapfrog, Verlet or velocity-Verlet.[14-16]

A major limitation to an efficient sampling with MD simulations is the discrete time step, Δt. It is desirable to choose a longer time step, which would give longer simulations with less computational resources. However, Δt is limited by the fastest motion in the simulated system. For an all-atom system, the fastest motion is the bond vibration between a hydrogen and a carbon atom, which limits Δt to about 1fs. Therefore, these bonds are typically constrained in the simulations, allowing a 2 fs time step.[14]

Calculating long-range interactions is very costly and was usually stopped after a given cut-off distance introducing important approximation in the calculations. This problem is now overcome by the introduction of Ewald summation and particle mesh Ewald (PME) methods rendering long-range electrostatic interactions significantly more accurate.[17]

The temperature of a simulated system is controlled by implementing a thermostat within the equation of motions that creates modifications, the common ones being modifying velocities (e.g. weak-coupling[18]), introducing fictitious particles in an extended system (e.g. Nosé-Hoover[19]) and introducing friction (e.g. Langevin dynamics). Similarly, the pressure can be controlled by the introduction of a barostat which also introduces modifications such as scaling the box dimension (e.g. weak-coupling), introducing fictitious particles (e.g. Parrinello-Rahman[20-21]) and introducing a piston.[14] In addition, the pressure regulation of a simulation needs to be handled accordingly to the system under simulation. In a system comprising a solute dissolved in a solvent, an isotropic pressure coupling is usually the most representative of the reality it aims to describe. In the case of simulation comprising a membrane, anisotropic pressure scaling is required to account for the non-isotropic nature of such a system due to the fact that the surface tension is not the same in all the directions.

The LEaP software, designed to construct simulations of molecules in solution, will overlay a pre-equilibrated solvent mask over the (biomolecular) solute, tile that mask

throughout the simulation cell, and then prune solvent residues which clash with the solute. The result of this procedure is a system which will likely contract under constant pressure dynamics as the pruning process has left vacuum bubbles at the solute:solvent interface.[22] We address this issue by using a MD multi-step protocol that takes care of relaxing the solvent before simulating the system (cf. *2.2.2 Molecular dynamics simulations protocols*).

2.1.4 Molecular dynamics force fields

Molecular mechanics can be a great tool to understand the behavior of biological molecules at atomic scale at the one condition that the force field it uses is accurate toward the residues it describes the motion. In summary, the trajectory of a molecular dynamics simulation is as representative of the reality as the input parameters are accurate. All force fields used throughout this thesis are briefly described in this section. An overview and the history of the development of the Amber force fields, the CHARMM force fields, the OPLS force fields and other protein force fields can be found at reference [23].

Amber ff14SB is the reference Amber force field for proteins. The Amber ff99SB force field[24] improved protein secondary structure balance and dynamics from earlier force fields like ff99,[25] but weaknesses in side chain rotamer and backbone secondary structure preferences have been identified. For the ff14SB force field, the authors performed a complete refit of all amino acid side chain dihedral parameters, which had been carried over from ff94.[26] The training set of conformations included multidimensional dihedral scans designed to improve transferability of the parameters. Improvement in all amino acids was obtained as compared to ff99SB. Parameters were also generated for alternate protonation states of ionizable side chains. Average errors in relative energies of pairs of conformations were under 1.0 kcal/mol as compared to QM, reduced 35% from ff99SB.[27] Additionally, empirical adjustments were made to the protein backbone dihedral parameters as compared to ff99SB. Multiple small adjustments of φ and ψ parameters were tested against NMR scalar coupling data and secondary structure content for short peptides. The best results were obtained from a physically motivated adjustment to the φ rotational profile that compensates for lack of ff99SB QM training data in the β-ppII transition region. Together, these backbone and side chain modifications not only better reproduced their benchmarks, but also

improved secondary structure content in small peptides and reproduction of NMR χ1 scalar coupling measurements for proteins in solution.[27]

Lipid14, the Amber lipid force field. In 2014, Lipid14[28] was released as the latest Amber lipid force field. Lipid14 represents a major advancement over the previous Amber compatible lipid force fields for lipid bilayer simulations in the NPT ensemble without the need for an artificial constant surface tension term. Lipid14 combines the modular framework of Lipid11[29] with a number of refinements inspired by GAFFlipid[30]. The modular nature of the force field allows for many combinations of lipid head groups and tail groups as well as rapid parameterization of further lipid types. In summary, several van der Waals and dihedral angle parameters have been refined to fit experimental data and quantum energies and new partial charges have been derived for the head and tail groups. The force field was validated on six principle lipid bilayer types for a total of 0.5 microseconds each without applying a surface tension or constant area term. The lipid bilayer structural features compare favorably with experimental measures such as area per lipid, bilayer thickness, NMR order parameters, scattering data, and lipid lateral diffusion. In addition, further validation of the Lipid14 parameters has been provided through extensive self-assembly simulations[31-32].[22] Furthermore, Lipid14 was recently expanded to include cholesterol parameters[33]. Lipid14 has been designed to be fully compatible with the other pairwise-additive protein, nucleic acid, carbohydrate, and small molecule Amber force fields[22], such as **ff14SB** mentioned above.

GLYCAM06, the Amber carbohydrates force field. It was originally developed as a set of parameters for MD simulation of carbohydrates in addition to the AMBER force field, later the force field was extended to other classes of molecules and its AMBER-dependency removed.[34] GLYCAM06 is a consistent and transferable parameter set for modeling carbohydrates,[34] and glycoconjugates.[35-36] When combining GLYCAM06 with AMBER parameters for other biomolecules, parameter orthogonality is ensured by assigning unique atom types for GLYCAM. In order to facilitate combining GLYCAM06 with other AMBER parameter sets for other biomolecules, a variation on the GLYCAM atom types has been introduced in which the new name consists of an uppercase letter followed by a second character, either a number or lowercase letter.[22] The GLYCAM force field family, especially, GLYCAM06, has been extensively employed in simulations of biomolecules by the larger scientific community.[37-40] The

updated GLYCAM parameters and documentation are available for download at the GLYCAM website (*www.glycam.org*). Also available on the website are tools for simplifying the generation of structure and topology files for performing simulations of oligosaccharides, glycoconjugates and glycoproteins.[22]

GAFF is a general Amber force field for organic molecules. GAFF is designed to be compatible with existing Amber force fields for proteins and nucleic acids, and has parameters for most organic and pharmaceutical molecules that are composed of H, C, N, O, S, P, and halogens. It uses a simple functional form and a limited number of atom types, but incorporates both empirical and heuristic models to estimate force constants and partial atomic charges. The performance of GAFF in test cases was considered encouraging with data comparable to results from Parm99/RESP. GAFF can be applied to wide range of molecules in an automatic fashion, making it suitable for rational drug design and database searching.[41]

2.2.4 Coarse-grained modeling and simulations

In the two previous sections emphasize was given on MD simulations at the atomic scale (cf. *2.1.2 Molecular mechanics and molecular dynamics simulations* and *2.1.3 Molecular dynamics force fields*), which has proven to be a powerful tool to study the structure and dynamics of model biological systems. However, studying a system with that degree of precision on the individual components has a high computational cost that limits the time and length scales for which such a system can be simulated dictated by computational power capabilities. Thus, some systems of biological relevance, such as protein folding, ion channel gating, signal transduction, and membrane remodeling, are difficult to investigate using atomistic simulations. To overcome this limitation, less detailed systems have been developed, among which coarse-graining (CG) is a very popular one.[42]

A coarse-grained model is a model in which a certain number of atoms are grouped together into a so-called grain. The computational cost of calculations is reduced as the number of degrees of freedom is lessened, permitting simulations of larger systems for longer times. The model is as accurate as the physical properties carried by the grains are representative of the properties of the atoms in regroups. The apparent geometry of the molecule is kept, as much as possible, by appropriately choosing which atoms to regroup. This CG approach has the huge advantage of greatly reducing the number of

forces that need to be computed and thus the computational power required for a simulation. Indeed a coarse-grained model contains fewer particles that its all-atom counterparts greatly reducing the computational load at every iteration resulting in faster simulation allowing a greater exploration of the energy landscape of a system in a shorter time (Figure 2.4). Several coarse-grained models were developed for simulation of proteins, the Martini CG force field being one of the most widely used. In the Martini force field[43-46], one grain usually accounts for four atoms and is parameterized based on experimentally measured values, *ab initio* calculations and molecular dynamics simulations derived values. An in-depth description of the Martini CG FF is given in section *2.2.3 Procedures and tools for Martini CG simulations*.

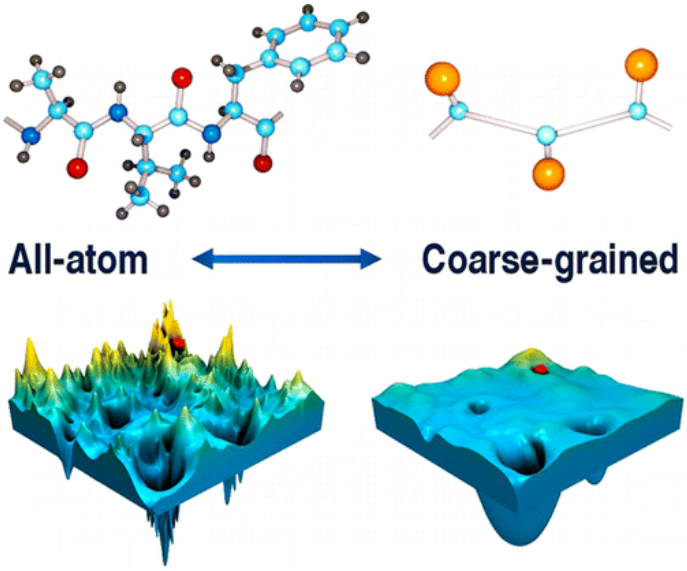


Figure 2.4. Schematic comparison of the structures and the energy landscape between a AA and a CG system (illustration retrieved here[47]).

2.2 Protocols and analysis

2.2.1 Basic data analysis

Here are presented some tools that are often used to perform a primary analysis of molecular dynamics trajectories. These tools are used to ensure the correctness of a simulation and to understand the evolution of the overall simulated system, as well as to catch subtler molecular behavior.

Root-mean-square deviation (RMSD) of atomic positions is the measure of the average distance of the same atom between any time step of a simulation and a similar reference structure or between two different simulation time steps (one being the reference) of a simulation, as given in equation **[2-6]**, where *N* represents the number of atoms and $\mathbf{r}_i(t)$ the position of atom *i* at time *t*.

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{N}(\mathbf{r}_i(t_0) - \mathbf{r}_i(t))^2} \qquad \textbf{[2-6]}$$

Usually, to obtain more meaningful and easily interpretable results, the overall translation and rotation are first removed by performing a rigid superposition of the structure to its reference that minimizes the RMSD. Although it is a straightforward analysis and gives an indication of local equilibrium, it is a far too simple method to assess the global convergence of the simulation.[14]

Whereas the RMSD provides an overall estimate for the entire protein, an approach to assess the degree of motion of individual residues is to compute the **root mean squared fluctuation (RMSF)**, which is simply the variance of the position of an atom.[14]

$$\mathrm{RMSF} = \sqrt{\frac{1}{T}\sum_{T}(\mathbf{r}(t) - \bar{\mathbf{r}})^2} \qquad \textbf{[2-7]}$$

RMSF is calculated by equation **[2-7]** in which T is the number of frames to be considered and $\bar{\mathbf{r}}$ the average position. The analysis can also be done on a per-residue basis, where all the atoms of a residue are included in the average and can for instance be used to assess the movement of side chains. Alternatively one can include only Cα atoms in the analysis to assess the backbone movement.[14]

Distances are straightforward to be calculated since all the atoms are defined in three-dimensional coordinates along with their trajectories. However, a special care needs to be addressed toward imaging since it is important to ensure that the two atoms, between

which the distance is measured, are staying in the same simulation box along the analyzed time.

Angles between specific atoms in MD trajectories are calculated by first defining two vectors between these sets of atoms, and then the angle can straightforwardly be retrieved, as the dot product of these two vectors is equal to the cosine of the angle between the vectors.

In the context of simulating membranes, it can be interesting to evaluate the average **area per lipid (APL)**. This is achieved by first printing out the two dimensions of the simulation box parallel to the membrane plane over time. Then these two lengths are multiplied to obtain the total membrane area. This area is then divided by the number of lipids occupying the layer from which one wants to know the APL of the lipids populating it.

2.2.2 Molecular dynamics simulations protocols

The simulations reported here are mainly performed following two different protocols that are chosen based on the type of system being simulated. The two types of system investigated throughout this thesis are: a system composed of at least one protein, alone or in complex, with at least one ligand, named *protein protocol* and a system composed of a bilayer and eventually one or more proteins and/or ligands, dubbed *membrane protocol*.

Protein protocol. Before being submitted to the production run, the system undergoes a height steps preparation. The first one consists of 1000 steps of steepest descent algorithm followed by 7000 steps of conjugate gradient algorithm under a 100 $kcal.mol^{-1}.A^{-2}$ harmonic potential constraint applied on the non-solvent component of the system. The conjugate gradient algorithm minimization continues while the harmonic potential is progressively lowered to 10, 5, 2.5 and 0 $kcal.mol^{-1}A^{-1}$ every 600 steps. The system is then heated from 0 K to 100 K using the Langevin thermostat in the canonical ensemble (NVT, number of particles, volume and temperature, respectively) while a 20 $kcal.mol^{-1}.A^{-2}$ harmonic potential restraint is applied on the protein. Finally, the system is heated up, from 100 K to 300 K, in the isothermal–isobaric ensemble (NPT, number of particles, pressure and temperature, respectively) under the same restraint conditions than the previous step, followed by a simulation for 100 ps under no harmonic restraint.

At this point, the system is ready for the production run, which is performed using the Langevin thermostat under NPT ensemble, at a 2 fs time step.

Membrane protocol. Steepest descent gradient algorithm is iterated for 5000 steps followed by 5000 iterations of conjugate gradient algorithm under no constraint. The system is then heated from 0 to 100K for 2500 steps in the NVT ensemble while the proteins and the lipids are held by a 10 $kcal.mol^{-1}A^{-1}$ harmonic potential. In the subsequent step the system is heated from 100K to 303K for 50000 steps. In membrane system the dimension of the box can change considerably during the first nanoseconds of simulation (due to intermolecular empty spaces left by the membrane building process), thus, to allow the program to recalculate them frequently, the first 10 steps of the production run are performed for a maximum of 500 ps. In all the steps the temperature is controlled by a Langevin thermostat. The warming up phase and the production run are performed under an anisotropic NPT ensemble to account for different physical properties along the dimensions tangential to the membrane than the one normal to it.

2.2.3 Procedures and tools for Martini CG simulations

MARTINIZE.py is a script to create Coarse Grain Martini input files of proteins, ready for use in the molecular dynamics simulations package Gromacs. More information can be found in the force field webpage and in related papers.[48-50]

Martinate and **Gromit** are auxiliary tools for automated atomistic and coarse-grained molecular dynamics simulations using Gromacs. Molecular dynamics simulations have complex workflows, including the generation of a model, setting up the environment, relaxation of the system and finally the production simulation. Despite the intrinsic complexity, the steps of the process are well-defined. For simulations of protein and/or DNA in solution, with or without ligand and with or without ions standard protocols are available. Gromit and martinate are versatile wrappers providing such protocols for atomistic (gromit) and coarse-grain (martinate) simulations, using the molecular simulation package Gromacs and, for martinate, the coarse grain Martini force field.[51]

Backward. The conversion of coarse-grained to atomistic models is an important step in obtaining insight about atomistic scale processes from coarse-grained simulations. Backward is a method, consisting of geometric projection and subsequent force field based relaxation. The method is designed to be simple and flexible, and offers a generic

solution for resolution transformation. For simple systems, the conversion only requires a list of particle correspondences on the two levels of resolution. For special cases, such as non-default protonation states of amino acids and virtual sites, a target particle list can be specified. The mapping uses simple building blocks, which list the particles on the different levels of resolution. For conversion to higher resolution, the initial model is relaxed with several short cycles of energy minimization and position-restrained MD. The reconstruction of an atomistic backbone from a coarse-grained model is done using a new dedicated algorithm. The method is generic and can be used to map between any two particle based representations, provided that a mapping can be written.[48]

DAFT (Docking Assay For Transmembrane components).[49] One example of the use of the coarse grain scale is to study protein-protein interactions. Transmembrane protein-protein interactions are particularly challenging to study experimentally and computationally. Experimentally the lipid environment is difficult to reproduce and once reproduced, difficult to resolve at atomic resolution. Computationally the size of these systems and the simulated time needed to obtain relevant information are usually the limiting factors. Coarse grain simulations have alleviated the later issue, but the slow movement through the bilayer, coupled to the long life times of non-optimal dimers, still stands in the way of characterizing binding distributions. DAFT, was developed to identify preferred binding orientations. The key feature of DAFT is the setup of starting structures, for which optimal periodic boundary conditions are devised. The purpose of DAFT is to perform a large number of simulations with different components, starting from unbiased non-interacting initial states, such that the simulations evolve collectively, in a manner reflecting the underlying energy landscape of interaction.[49] DAFT and its auxiliary programs are available from http://cgmartini.nl, together with a working example.

2.3 Protein-ligand docking

Molecular docking is a computational procedure that aims to predict the preferred orientation of a ligand with a macromolecular target (the receptor) in which they are bound to each other to form a stable complex. Docking is a handy tool to predict bioactive conformations, identify binding sites in a given receptor, unveil essential ligand-receptor interactions, or to screen vast databases of potential ligands. These applications are useful in the context of hit identification and lead optimization.[52-53]

The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, or lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism or antagonism). Therefore, docking is useful for predicting both the strength and type of signal produced.

Several docking programs are available in the context of molecular recognition and drug design, we can cite AutoDock[54], AutoDock Vina[55], DOCK[56], FlexX[57], GLIDE[58], ICM[59], PhDOCK[60], and Surflex[61]). They have been extensively tested and compared.[62-63] Although each docking program operates slightly differently, protein-ligand docking calculations are usually performed in two steps: conformational sampling and scoring. Conformational sampling consists in generating a database of conformers of the ligand to be used in ligand docking. Conformational search can be performed as a separate step before docking or it can be implemented as an integrated part of the docking process. Scoring functions widely vary across the different programs, but often fold under one of the following categories: force field-based, empirical, knowledge-based, clustering and entropy-based, or consensus scoring methods.[64]

Docking calculations are widely used throughout this thesis with the main goal of predicting plausible binding modes to understand how a given, naturally occurring or not, ligand interacts with a receptor. The main receptors studied in this thesis are CD14 and the TLR4/MD-2 dimer (cf. chapter I section *1.1 Scientific background about TLR4*). Docking was often performed independently with AutoDock 4.2 and AutoDock Vina, both are described hereafter.

2.3.1 AutoDock 4.2

AutoDock[51, 54] is an automated procedure for predicting the interaction of ligands with bio-macromolecular targets. It works with AutoGrid, an accessory program that pre-

calculates grid maps of interaction energies for various atom types with the receptor. These maps are used by AutoDock during the docking calculation to estimate the total energy of binding between the ligand with the macromolecule. This pre-calculation greatly reduce the time needed for the calculation as it increases the efficiency of the program from a $O(N^2)$ complexity to a O(N) (big O notation), with N the number of atoms interacting.[51]

AutoDock 4.2 uses a semi-empirical free energy force field to evaluate conformations during docking simulations. The force field was parameterized using a large number of protein-inhibitor complexes for which both structure and inhibition constants, or Ki, are known.[65] The force field evaluates binding in two steps. The ligand and protein start in an unbound conformation. In the first step, the intramolecular energetics are estimated for the transition from these unbound states to the conformation of the ligand and protein in the bound state. The second step then evaluates the intermolecular energetics of combining the ligand and protein in their bound conformation. The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}\right) \quad \textbf{[2.4]}$$

where L refers to the ligand and P refers to the protein in a ligand-protein docking calculation.[65]

Each of the pair-wise energetic terms includes evaluations for dispersion/repulsion, hydrogen bonding, electrostatics, and desolvation:

$$V = W_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec} \sum_{i,j} \frac{q_i q_j}{e(r_{ij}) r_{ij}} + W_{sol} \sum_{i,j} \left(S_i V_j + S_j V_i\right) e^{\left(-\frac{r_{ij}^2}{2\sigma^2}\right)} \quad \textbf{2.5]}$$

The weighting constants W have been optimized to calibrate the empirical free energy based on a set of experimentally determined binding constants. The first term is a typical 6/12 potential for dispersion/repulsion interactions. The parameters are based on the Amber force field. The second term is a directional H-bond term based on a 10/12 potential. The parameters C and D are assigned to give a maximal well depth of 5 kcal/mol at 1.9 Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5 Å for hydrogen bonds with sulfur. The function E(t) provides

directionality based on the angle t from ideal H-bonding geometry. The third term is a screened Coulomb potential for electrostatics. The final term is a desolvation potential based on the volume of atoms (V) that surround a given atom and shelters it from the solvent, weighted by a solvation parameter (S) and an exponential term with distance-weighting factor σ=3.5 Å.[65]

By default AutoDock 4.2 estimates the contribution of the unbound state by assuming that the unbound form of the ligand (V_{bound}^{L-L} in the equation **[2-6]**) is the same as the final docked conformation of the ligand ($V_{unbound}^{L-L}$), yielding a final contribution $V_{bound}^{L-L} - V_{unbound}^{L-L} = 0$.[65]

In all the docking performed with AutoDock 4.2 throughout this thesis, the Lamarckian evolutionary algorithm was chosen and all parameters were kept default except for the number of genetic algorithm (GA) runs which was set to 200 to enhance the sampling. AutoDockTools 1.5.6 was used to assign the Gasteiger-Marsili empirical atomic partial charges to the atoms of both the ligands and the receptors. The structure of the receptors was always kept rigid, whereas the structure of the ligand was set partially flexible by providing freedom to some appropriately selected dihedral angles.

2.3.2 AutoDock Vina

AutoDock Vina is an open-source program for doing molecular docking. It was designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute.[51, 55] The authors describe the Vina scoring function as more of "machine learning" than directly physics-based in its nature, which they justify by its performance on test problems rather than by theoretical considerations following some, possibly too strong, approximating assumptions.[55]

Vina scoring function was mostly inspired by X-score, and, like X-score, was tuned using the PDBbind.[66-68] However, some terms are different from X-score, and, in tuning the scoring function, Vina's developers went beyond linear regression.[55] As optimization algorithm, Vina uses the Iterated Local Search global optimizer[69-70] similar to that by Abagyan *et al.*[59] In this algorithm, a succession of steps consisting of a mutation and a local optimization are taken, with each step being accepted according to the Metropolis criterion. Vina uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[71] method for the local optimization, which is described as an efficient quasi-Newton method.[55]

BFGS, like other quasi-Newton optimization methods, uses not only the value of the scoring function but also its gradient, i.e., the derivatives of the scoring function with respect to its arguments. The arguments, in Vina's case, are the position and orientation of the ligand, as well as the values of the torsions for the active rotatable bonds in the ligand and flexible residues, if any. Vina can concurrently perform several runs starting from random conformations allow it to take advantage of multithreading.[55] AutoDock Vina tends to be faster than AutoDock 4 by orders of magnitude.[51]

2.4 Other computational methods and software programs

2.4.1 Homology modeling

Amongst protein structure prediction techniques, homology modeling (or comparative modeling) aims at unraveling the secondary, tertiary and eventually quaternary structure of a protein given a primary amino acids sequence by comparing it to already existing structures of proteins with similar amino acid sequences. This approach is justified by the evolutionary driven fact that similar protein sequences encode similar three-dimensional structures.[72] A usual homology modeling workflow goes through template selection and sequence alignment, target-template alignment, model construction, and finally model assessment. An in-depth explanation of these steps can be found elsewhere[14]. Ultimately, protein structure prediction aims at bridging primary amino acids sequence to the function of the protein it represents.

In this thesis, protein prediction needs were addressed with the homology modeling feature of YASARA,[73] a program for molecular visualizing, modeling, and dynamics. First, the program collects several templates, based on sequence similarities, then the algorithm attempts to construct three-dimensional models that it ranks according to alignment score and structural quality. The loops are optimized by conformational sampling with side-chains re-optimization. Ultimately, the program seeks to improve the final result by building a hybrid model in which bad regions of the top-ranked model are iteratively swapped with corresponding fragments from other models.

2.4.2 PyMOL

In the context of parameters derivation PyMOL molecular graphics and modeling package[74] was used to construct and modify ligands. Additionally, PyMOL was used for visual analysis of the docking outputs, of molecular dynamics trajectories, and to render all molecules pictures present in this thesis.

2.4.3 Maestro

Maestro[75] is a molecular modeling and visualizing program for drug design and materials science. Here it was mainly used for structure visualization and restrained minimization procedure under the OPLS3 force field. Additionally, the software was used for AlogP calculation when evaluating ligand solubility.

2.4.4 Antechamber

Antechamber[76] regroups a set of auxiliary programs for MD simulation. Its main application in this thesis is to greatly facilitate the creation of new force field entry of novel molecules or residues of interest and to support their parameterization. Parameters for molecular dynamics simulations were set up with the standard Antechamber procedure. Briefly, charges were calculated with Gaussian at the Hartree-Fock level (HF/6-31G* Pop=MK iop(6/33=2) iop(6/42=6)) from the solvated DFT B3LYP optimized structure (cf. *2.1.1 Quantum mechanics methods: Hartree–Fock and DFT*), then derived and formatted for Ambertools15 and Amber 14 with Antechamber assigning the general AMBER force field (GAFF)[41] atom types. Subsequent modifications to the atom types were made when thought necessary, e.g. to take advantage of the GLYCAM06[34] force field should a saccharide fragment of a bigger molecule be already described by the force field.

2.4.5 LEaP

LEaP[77], under both its command line format, known as teLeap (run by the *tleap* shell script), and its X-windows graphical user interface enhanced format, known as xaLeap (*xleap*) is the primary program to create a new system in Amber, or to modify existing systems. It combines the functionality of prep, link, edit and parm from much earlier versions of Amber.[22] LEaP serves the major purpose of connecting a coordinate file, which contains a spatial description of all the atoms contained in a system, with a desired force field, to create a new amber-compatible coordinate file and a topology file. The force field file contains all the parameters required by the potential function, such as molecule and residue information, atom names, atom types, atomic charges, atomics connectivities, atomic coordinates, atomic masses, bonded parameters (bond, angle, dihedral) and non-bonded parameters (electrostatic and van der Waals).

LEaP is also a powerful tool for force field modifications and adaptations. For instance, when using more than one force field within a unique molecule, such as GAFF and GLYCAM06 in the case of many LPS-like ligands presented in this thesis, one needs to define the parameters necessary to describe the interface between the two force fields. Toward that end, modifications were introduced within GAFF, GLYCAM06, and AMBERff14. LEaP was used to create force field template for new residues. All modifications used in the work reported herein have been attached in annex within their

relevant sections. Additionally, LEaP, in a classical pdb to MD simulations workflow, is used to define simulation box, solvate the system and add ions, among other things.

Bibliography

1. Hodak, H. The Nobel Prize in chemistry 2013 for the development of multiscale models of complex chemical systems: a tribute to Martin Karplus, Michael Levitt and Arieh Warshel. *J. Mol. Biol.* **2014,** *426* (1), 1-3.
2. *Computational Tools for Chemical Biology*. The royal society of chemistry: 2018.
3. https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1998/ (accessed February 2018).
4. https://www.nobelprize.org/nobel_prizes/chemistry/laureates/2013/ (accessed February 2018).
5. http://biocomp.chem.uw.edu.pl/research/coarse-grained-modeling (accessed February 2018).
6. Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Phys. Rev.* **1964,** *136* (3B), B864-B871.
7. Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140* (4A), A1133-A1138.
8. Stephens, P.; Devlin, F.; Ashvar, C.; Chabalowski, C.; Frisch, M. Theoretical calculation of vibrational circular dichroism spectra. *Faraday Discuss.* **1994,** *99*, 103-119.
9. Runge, E.; Gross, E. K. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984,** *52* (12), 997-1000.
10. Drenth, J. *Principles of protein X-ray crystallography*. Springer Science & Business Media: 2007.
11. Bai, X.-C.; McMullan, G.; Scheres, S. H. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **2015,** *40* (1), 49-57.
12. Raman, S.; Lange, O. F.; Rossi, P.; Tyka, M.; Wang, X.; Aramini, J.; Liu, G.; Ramelot, T. A.; Eletsky, A.; Szyperski, T. NMR structure determination for larger proteins using backbone-only data. *Science.* **2010,** *327* (5968), 1014-1018.
13. http://atomsinmotion.com/book/chapter5/md (accessed February 2018).
14. Genheden, S.; Reymer, A.; Saenz-Mendez, P.; Eriksson, L. A. Chapter 1 Computational Chemistry and Molecular Modelling Basics. In *Computational Tools for Chemical Biology*, The Royal Society of Chemistry: 2018; pp 1-38.
15. Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* **1982,** *76* (1), 637-649.
16. Verlet, L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **1967,** *159* (1), 98-103.
17. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
18. Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.
19. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984,** *81* (1), 511-519.
20. Parrinello, M.; Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **1980,** *45* (14), 1196-1199.
21. Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.

22. Case, D.; Betz, R.; Botello-Smith, W.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A. Amber 2016.
23. Ponder, J. W.; Case, D. A. Force fields for protein simulations. In *Adv. Protein Chem.*, Elsevier: 2003; Vol. 66, pp 27-85.
24. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* **2010,** *78* (8), 1950-1958.
25. Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21* (12), 1049-1074.
26. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117* (19), 5179-5197.
27. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
28. Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: the amber lipid force field. *J. Chem. Theory Comput.* **2014,** *10* (2), 865-879.
29. Skjevik, Å. A.; Madej, B. D.; Walker, R. C.; Teigen, K. LIPID11: a modular framework for lipid simulations using amber. *J. Phys. Chem. B.* **2012,** *116* (36), 11124-11136.
30. Dickson, C. J.; Rosso, L.; Betz, R. M.; Walker, R. C.; Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter.* **2012,** *8* (37), 9617-9627.
31. Skjevik, Å. A.; Madej, B. D.; Dickson, C. J.; Teigen, K.; Walker, R. C.; Gould, I. R. All-atom lipid bilayer self-assembly with the AMBER and CHARMM lipid force fields. *Chem. Commun.* **2015,** *51* (21), 4402-4405.
32. Skjevik, Å. A.; Madej, B. D.; Dickson, C. J.; Lin, C.; Teigen, K.; Walker, R. C.; Gould, I. R. Simulation of lipid bilayer self-assembly using all-atom lipid force fields. *PCCP.* **2016,** *18* (15), 10573-10584.
33. Madej, B. D.; Gould, I. R.; Walker, R. C. A parameterization of cholesterol for mixed lipid bilayer simulation within the Amber Lipid14 force field. *J. Phys. Chem. B.* **2015,** *119* (38), 12424-12435.
34. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
35. DeMarco, M. L.; Woods, R. J. Atomic-resolution conformational analysis of the GM3 ganglioside in a lipid bilayer and its implications for ganglioside–protein recognition at membrane surfaces. *Glycobiology.* **2008,** *19* (4), 344-355.
36. DeMarco, M. L.; Woods, R. J.; Prestegard, J. H.; Tian, F. Presentation of membrane-anchored glycosphingolipids determined from molecular dynamics simulations and NMR paramagnetic relaxation rate enhancement. *J. Am. Chem. Soc.* **2010,** *132* (4), 1334-1338.
37. Kadirvelraj, R.; Grant, O. C.; Goldstein, I. J.; Winter, H. C.; Tateno, H.; Fadda, E.; Woods, R. J. Structure and binding analysis of Polyporus squamosus lectin in complex with the Neu5Acα2-6Galβ1-4GlcNAc human-type influenza receptor. *Glycobiology.* **2011,** *21* (7), 973-984.

38. DeMarco, M. L.; Woods, R. J. From agonist to antagonist: Structure and dynamics of innate immune glycoprotein MD-2 upon recognition of variably acylated bacterial endotoxins. *Mol. Immunol.* **2011,** *49* (1-2), 124-133.
39. Foley, B. L.; Tessier, M. B.; Woods, R. J. Carbohydrate force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012,** *2* (4), 652-697.
40. Ficko-Blean, E.; Stuart, C. P.; Suits, M. D.; Cid, M.; Tessier, M.; Woods, R. J.; Boraston, A. B. Carbohydrate recognition by an architecturally complex α-N-acetylglucosaminidase from Clostridium perfringens. *PLoS One.* **2012,** *7* (3), e33524.
41. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
42. Barnoud, J.; Monticelli, L. Coarse-grained force fields for molecular simulations. In *Molecular modeling of proteins*, Springer: 2015; pp 125-149.
43. Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B.* **2007,** *111* (27), 7812-7824.
44. Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S. J. Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.* **2010,** *6* (6), e1000810.
45. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI coarse-grained force field: extension to proteins. *J. Chem. Theory Comput.* **2008,** *4* (5), 819-834.
46. de Jong, D. H.; Singh, G.; Bennett, W. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput.* **2012,** *9* (1), 687-697.
47. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016,** *116* (14), 7898-7936.
48. Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **2014,** *10* (2), 676-690.
49. Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-throughput simulations of dimer and trimer assembly of membrane proteins. The DAFT approach. *J. Chem. Theory Comput.* **2015,** *11* (5), 2278-2291.
50. http://www.cgmartini.nl (accessed February 2018).
51. https://github.com/Tsjerk/gromit (accessed February 2018).
52. Chaudhary, K. K.; Mishra, N. A Review on Molecular Docking: Novel Tool for Drug Discovery. *JSM Chem.* **2016,** *4* (3), 1029-1033.
53. Yuriev, E.; Holien, J.; Ramsland, P. A. Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J. Mol. Recognit.* **2015,** *28* (10), 581-604.
54. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
55. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
56. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput. Aided Mol. Des.* **2006,** *20* (10-11), 601-619.

57. Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins.* **1999,** *37* (2), 228-241.
58. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-1749.
59. Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994,** *15* (5), 488-506.
60. Joseph-McCarthy, D.; Thomas, B. E.; Belmarsh, M.; Moustakas, D.; Alvarez, J. C. Pharmacophore-based molecular docking to account for ligand flexibility. *Proteins.* **2003,** *51* (2), 172-188.
61. Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003,** *46* (4), 499-511.
62. Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009,** *49* (6), 1455-1474.
63. Castro-Alvarez, A.; Costa, A. M.; Vilarrasa, J. The Performance of Several Docking Programs at Reproducing Protein–Macrolide-Like Crystal Structures. *Molecules.* **2017,** *22* (1), 136.
64. Elokely, K. M.; Doerksen, R. J. Docking challenge: protein sampling and molecular docking performance. *J. Chem. Inf. Model.* **2013,** *53* (8), 1934-1945.
65. Forli, W. E. H.; Halliday, S.; Belew, R.; Olson, A. J. AutoDock Version 4.2.
66. Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.* **2002,** *16* (1), 11-26.
67. Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind database: Collection of binding affinities for protein−ligand complexes with known three-dimensional structures. *J. Med. Chem.* **2004,** *47* (12), 2977-2980.
68. Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind database: methodologies and updates. *J. Med. Chem.* **2005,** *48* (12), 4111-4119.
69. Baxter, J. Local optima avoidance in depot location. *J. Oper. Res. Soc.* **1981,** *32* (9), 815-819.
70. Blum, C.; Roli, A. Hybrid metaheuristics: an introduction. In *Hybrid Metaheuristics*, Springer: 2008; pp 1-30.
71. Nocedal, J.; Wright, s. Numerical Optimization, Chapter 10, Nonlinear Least-Squares Problems, eds. Glynn, P. & Robinson, sM. springer, New York: 1999.
72. Kaczanowski, S.; Zielenkiewicz, P. Why similar protein sequences encode similar three-dimensional structures? *Theor. Chem. Acc.* **2010,** *125* (3-6), 643-650.
73. Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins.* **2004,** *57* (4), 678-683.
74. Schrödinger, L. PyMOL(TM) Molecular Graphics System, Version 1.6.0.0.
75. https://schrodinger.com/maestro (accessed February 2018).
76. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006,** *25* (2), 247-260.
77. Case, D.; Berryman, J.; Betz, R.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A. AMBER 2015. *University of California, San Francisco.* **2015**.

CHAPTER 3

TLR4 Modulators

3.1 Naturally occurring modulators

3.1.1 *Bradyrhizobium* LPS

Rhizobia are Gram-negative bacteria able to establish symbiotic relationship with legumes and to reduce atmospheric nitrogen into ammonium, thus providing nitrogen nutrition for the host plants.[1-2] Bacteria belonging to the *Bradyrhizobium* genus promote nitrogen-fixing nodules development on roots and stems of both wild-growing and cultivated *Aeschynomene* legumes.[3-4] It was previously demonstrated that the lipopolysaccharide (LPS) macromolecule in Rhizobia plays a key role throughout the symbiotic process and that its structural features are altered in response to plant signals.[5-7]

Our collaborators within the MSCA-ITN TOLLerant project, Profs. Molinaro and Silipo from the University of Naples Federico II, Italy, recently elucidated the structure of the lipid A from *Bradyrhizobium* strains (Figure 3.1). These bacteria produce a mixture of lipid A species differing by the number, length and nature of the acyl chains.[8-10] All are made up of a pentasaccharide sugar backbone formed by a skeleton of β(1→6) linked 2,3-diamino-2,3-dideoxy-glucose (DAG) substituted by an α-galacturonic acid on the vicinal DAG and by an α-mannose disaccharide linked to the distal β-DAG unit. They detected the presence of very long-chain fatty acids (VLCFA), which are known to be pivotal in the bacterium adaptation to the intracellular life.[8, 11-13] More intriguingly, a hopanoid molecule linked to the VLCFA, was also present in a non-stoichiometric fashion (Figure 3.1). Such a highly heterogeneous lipid A was also identified in other *Bradyrhizobium* strains, as ORS278 and ORS285 strains[10].

It is widely accepted that the LPS structure and, in particular, the lipid A part, predisposes the macromolecule to act as a elicitor of the host innate immune system.[14-16] Rhizobial lipid A remarkably differ from enterobacterial analogs in the fatty acid pattern, carbohydrate backbone and phosphate content. This prompted us to investigate a potential antagonistic action of rhizobial lipid A compounds that might be applied as therapeutic agents for the prevention of Gram-negative-induced sepsis. Indeed, several studies have reported the weak endotoxic activity of rhizobial lipid A, as well as the antagonistic properties towards the toxic effects of enterobacterial LPS[17-20] with the single exception of *S. meliloti*[21]. Knowing that Bradyrhizobia strains express a very

unusual lipid A structure prompted us to evaluate the impact of *Bradyrhizobium* LPS/lipid A on the innate immune system.

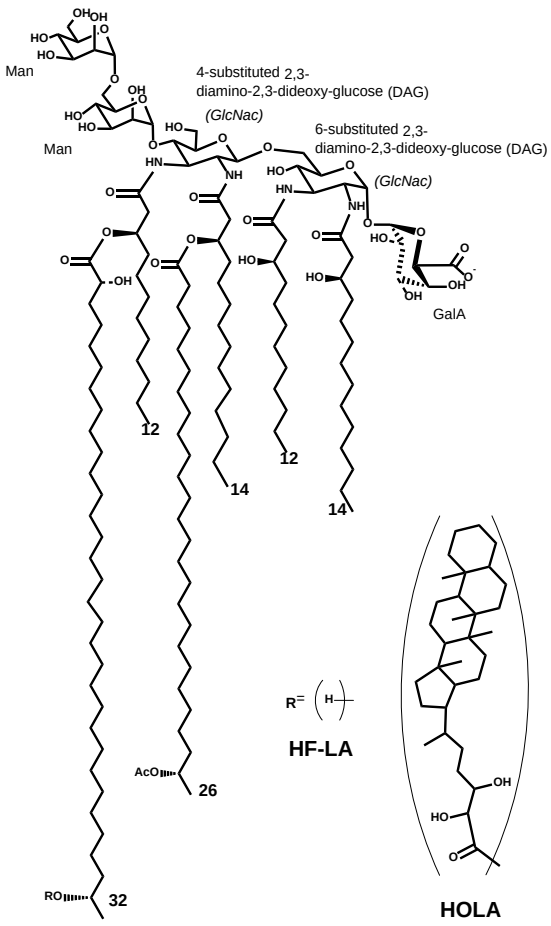


Figure 3.1. Representation of the HOLA and HF-LA structures used as representative species of the sample for the computational studies.

Experimental studies, including cell assays on both murine and human bone marrow-derived macrophages and HEK 293-TLR4/MD-2/CD14 cells, performed by Prof. Silipo and collaborators, revealed an extremely low capability to elicit an immune response. More intriguingly, a potent antagonistic activity towards the toxic *E. coli* LPS was observed (data not shown).[22] Only the computational study is reported herein.

Molecular Modeling of *Bradyhrizobium* lipid A binding to MD-2/TLR4

We performed computational studies to discover the possible binding modes and understand the dynamic behavior of *Bradyhrizobium* lipid A (HOLA, Figure 3.1) in complex with human MD-2/TLR4. Since the experimental samples contain different derivatives, we also studied hopanoid-free *Bradyrhizobium* lipid A (HF-LA, Figure 3.1). Docking results were evaluated based on the predicted binding score and on the apparent degree of similarity with *E. coli* LPS and lipid IVa as known from their

crystallographic structures accessible under the accession codes 3FXI and 2E59, respectively, considering both the insertion of the fatty acid chains into the MD-2 pocket and the positioning of the disaccharide core. In addition, we systematically discarded the poses in which at least one of the saccharide bearing acyl chains was rotated such as that the amide groups connecting the saccharide to the lipid chains were facing the opposite direction of the binding pocket. This orientation causes a large portion of the lipid chains to be exposed to the solvent, which we consider unlikely.

We started by carrying out docking calculations of HF-LA and HOLA in MD-2. Plausible binding modes were obtained with most of the FA chains inserted inside the MD-2 pocket, while the sugar moieties interacted at its rim. HOLA and HF-LA were first docked against MD-2 alone, in both its agonist and antagonist conformations. The predicted binding energy for HOLA in the agonist conformation ranges from -5.9 to -4.2 kcal mol^{-1} and from -5.5 to -4.3 kcal mol^{-1} in the antagonist conformation. In the case of HF-LA the energy ranges from -6.0 to -4.3 kcal mol^{-1} in the agonist conformation and from -5.6 to -4.5 kcal mol^{-1} in the antagonist conformation of MD-2. These results, being of similar magnitude, do not permit to determine a preferential binding toward the agonist or the antagonist conformation of MD-2.

As for the interactions, in the case of HF-LA, the two VLCFA are often fully accommodated inside the MD-2 cavity where they are surrounded with hydrophobic residues such as Val24, Ile32, 46, 63, 94, 117, 153 and Leu61, 78, leaving space for only two shorter lipid chains to enter the pocket. One of the two remaining shorter chain is often directed toward Phe126. The other one is placed in a small corridor pointing at Ser103 (Figure 3.2).

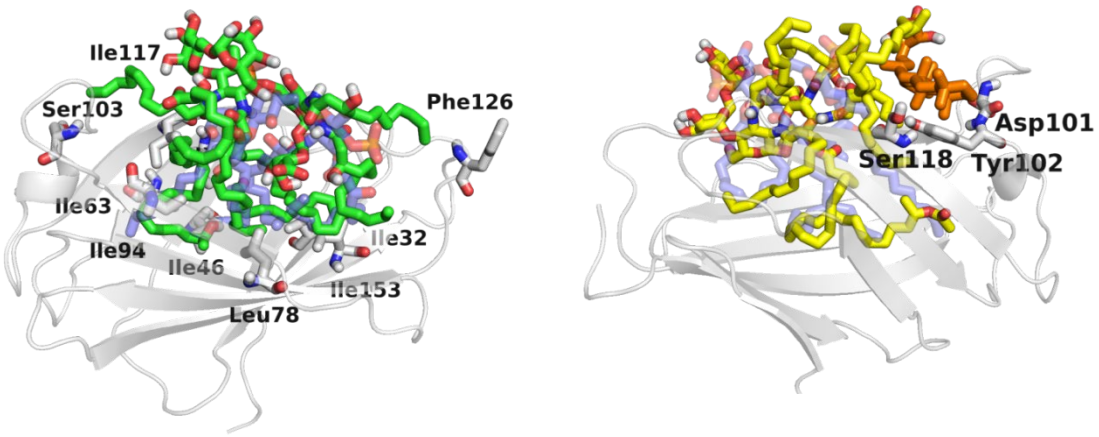


Figure 3.2. Best poses for HF-LA, represented in green sticks (on the left), and HOLA, in yellow sticks with its hopanoid residue in orange (on the right) in MD-2 only. Lipid IVa, added for comparison purposes, is depicted in violet carbon CPK colored semi-transparent sticks. MD-2 is in semi-transparent cartoon. The residues mentioned in the text are in sticks with their corresponding individual labelling.

In the case of HOLA the hopanoid moiety lies at the rim of the MD-2 pocket close to the residues Asp101, Tyr102 and Ser118 (Figure 3.2) or at a completely distinct location close to residue Lys125 and Phe126. In many docked poses, the hopanoid moiety remains on a surface outside or at the rim of the MD-2 pocket. Some low-score docking poses feature the hopanoid moiety in the hydrophobic pocket, proving that in theory it can sterically be accommodated inside MD-2. However, in these poses, the penta-saccharide backbone is accommodated further away from the binding pocket, on a loop followed by a β-sheet formed by residues 87 to 91 (Figure 3.3). This is likely due to the steric constraints inherent to the hopanoid moiety being inserted in the pocket. Additionally, it occupies a consequent volume and seems to obstruct the passage for the lipids chains, resulting in poses in which at least three acyls chains are left outside the pocket. These results suggest that the hopanoid moiety may not play an important role in the effective binding of HOLA to MD-2, and thus its presence might not be necessary for *Bradyrhizobium* to exert its antagonist activity.

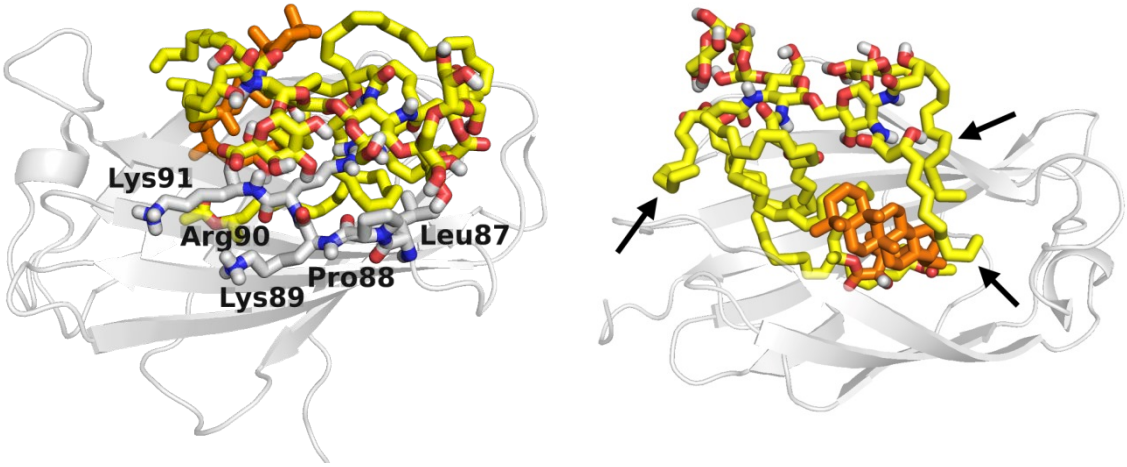


Figure 3.3. Example of a docked pose in which the hopanoid moiety (in orange sticks) is inserted inside the hydrophobic pocket. MD-2 (from PDB ID 2E59) is represented in semi-transparent grey cartoon and HOLA in yellow carbon and CPK colored sticks. On the left: the residues of the loop and β-sheet mentioned in the text are shown in sticks and are individually labelled. On the right: the three shorter acyl chains exposed to the solvent are indicated by black arrows.

In a second approach, docking calculations were performed in the TLR4/MD-2 system in the antagonist conformation. We used a computational model optimized by us (see section *3.3.1 Amphiphilic Guanidinocalixarenes Inhibit Lipopolysaccharide (LPS)- and Lectin-Stimulated Toll-like Receptor 4 (TLR4) Signaling* for details).[23] Interestingly, when compared with the agonist structure of the TLR4/MD-2 complex (PDB ID 3FXI), among the non-bonded interaction between the two proteins, a loop of TLR4, composed of amino acids 263 to 266, protrudes into a MD-2 channel (Figure 3.4), located approximately between Asp161 and Tyr118. This protrusion is further amplified by the side chain of Arg264 that goes as far as to hover over the MD-2 hydrophobic pocket. This impingement of TLR4 over MD-2 diminishes the space available for ligand interactions in the TLR4/MD-2 complex, both agonist and antagonist conformations, compared with MD-2 alone. We superimposed TLR4, based on the crystal structure, to the docking results from the MD-2-only study, and noted a steric incompatibility between the TLR4 protruding loop, and either, a lipid chain, or the hopanoid moiety from the docked ligand (Figure 3.4). This observation could point to the fact that HOLA and HF-LA carry on their antagonist activities by preventing or impairing the formation of a proper TLR4/MD-2 dimer essential for TLR4 activation.

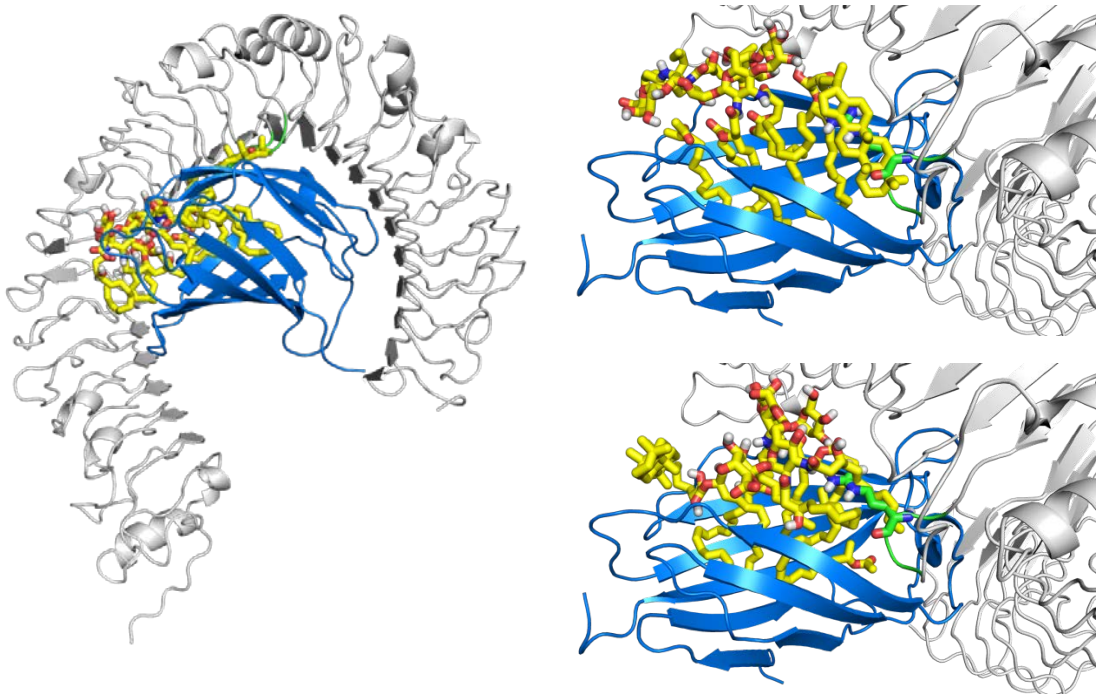


Figure 3.4. Representation of the steric clash observed over the addition of TLR4 to the docked poses calculated in MD-2 only. TLR4, MD-2 and HOPA are respectively represented in grey cartoon, blue cartoon and yellow sticks. The TLR4 protruding loop mentioned in the text is colored in green and Arg264 is in green sticks. A general view of the TLR4/MD-2/HOPA complex (on the left) and two examples of steric clashes are given (on the right). These involve Arg264 and, respectively, the hopanoid moiety (on the top) and one of the short acyl chains of HOPA (on the bottom).

The score attributed to the poses by AutoDockVina ranges from -6.8 to -4.9 kcal mol^{-1} for HOLA and from -6.6 to -5.1 kcal mol^{-1} for HF-LA in the TLR4/MD-2 system. The hydrophobic interactions taking place inside the hydrophobic pocket were essentially the same as the one described in the case of MD-2 alone. However, in the case of the docking in the TLR4/MD-2 system, the presence of TLR4, reducing the space available for ligand binding, resulted in very few poses that abide the sugar orientation criterion mentioned above. Two poses of HOLA, in good agreement with lipid IVa, were selected for MD simulations to further investigate interactions with the receptor and overall stability of the TLR4/MD-2/ligand complex. These two poses are 180^{o} rotated one to the other: in the first one, the HOLA lipid A is oriented as lipid IVa (PDB ID 3E59) and, in the second one, it is oriented as *E. coli* LPS (PDB ID 3FXI, cf. Figure 1.6 in *Introduction* chapter).

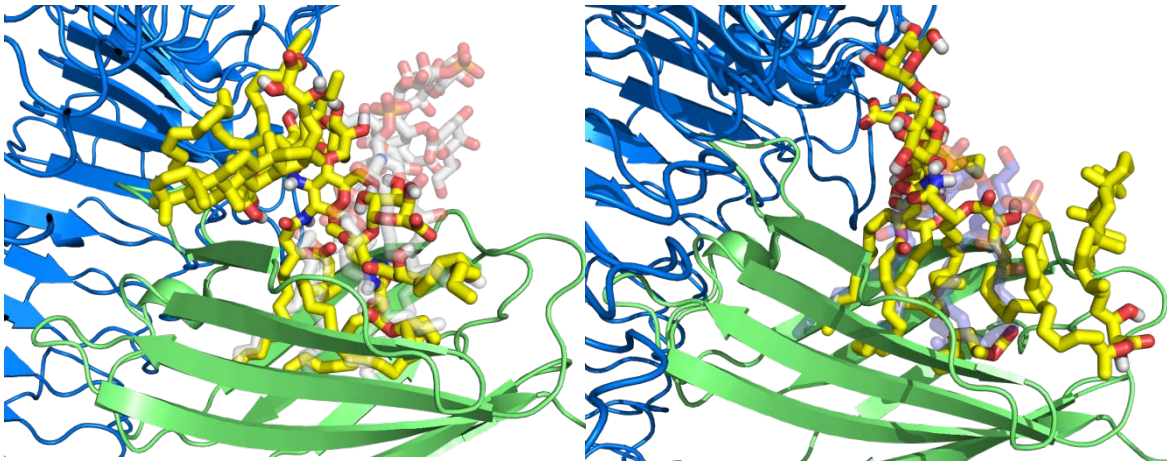


Figure 3.5. Representation of the two poses, *E. coli* lipid A-like (on the left) and lipid IVa-like (on the right), selected for MD simulations.

In the *E. coli* lipid A-like pose the hopanoid moiety, the VLCFA linked to it, and the C_{12} acyl chain from the same ramification, are folded in a narrow region between a TLR4 loop formed by residues from 381 to 386 and a MD-2 β-sheet formed by residues 90 to 95 (Figure 3.5). The second VLCFA and two shorter acyl chains are accommodated inside the MD-2 pocket and the last shorter chain follow the same corridor toward Phe126 as one acyl chain of *E. coli* LPS. The two mannoses are forming polar interactions with TLR4 residues Arg264, Asn339 and Lys362. In the lipid IVa-like pose, the hopanoid moiety is in close proximity of Phe126 and its VLCFA is partially inserted in the MD-2 pocket along with the other VLCFA and 3 shorter acyl chains. The remaining acyl chain lies outside the pocket next to the TLR4 protruding loop mentioned earlier. As for the saccharides, the galacturonic acid establishes polar interactions with TLR4 residues Arg264, Asn339 and Lys362. The two mannoses are packed at the same position where the hopanoid moiety is found in the *E. coli* LPS-like pose (Figure 3.5).

Both docked complexes of HOLA/TLR4/MD-2 (with the lipid IVa-like and the *E. coli* lipid A-like poses) were submitted to MD simulations. TLR4 presents important deviation compared to the crystal structure in relation with MD-2, as shown in the RMSD plot (Figure 3.6, left panel). TLR4 displays a tendency to break apart from MD-2 indicating that the presence of the ligand destabilizes the TLR4/MD-2 complex (Figure 3.7). It distorts the relative TLR4 vs MD-2 disposition in disagreement with the agonist geometry. This observation backs-up the hypothesis that HOPA and HF-LA act as antagonists by either preventing complex formation (cf. protruding loop mentioned in the docking study above) or by disturbing the complex stability. In addition, Phe126

remains in its open conformation all along the simulation (Figure 3.6, right panel), this stability was previously associated with antagonist ligands[23].

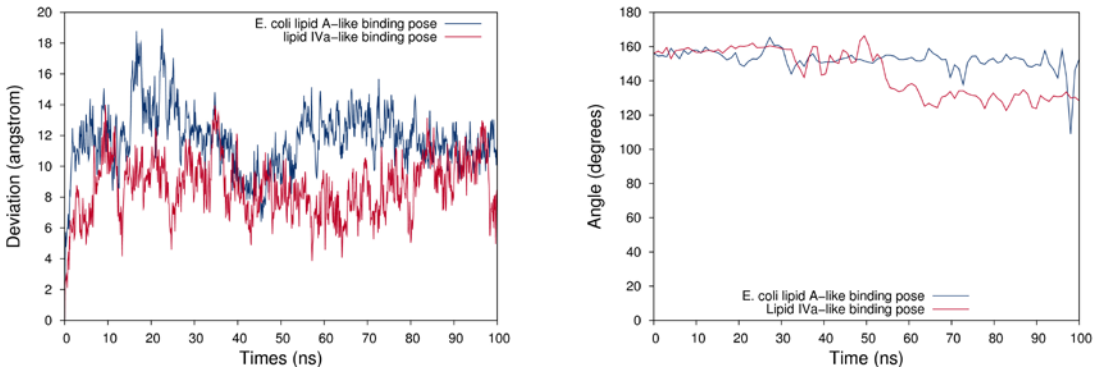


Figure 3.6. On the left: RMSD of TLR4 in relation to the first frame in comparison with MD-2 as the minimum fit of the system is performed on the backbone of MD-2. On the right: angle over simulation time between two arbitrarily selected vectors starting both from the α-carbon of residue Phe126 to, respectively, the zeta-carbon of the same residue and the α-carbon of residue Ser21 (Figure 3.14A). The angle plotted over time shows the stability of residue Phe126 during the MD simulation associated with antagonist activity of the ligand[23].

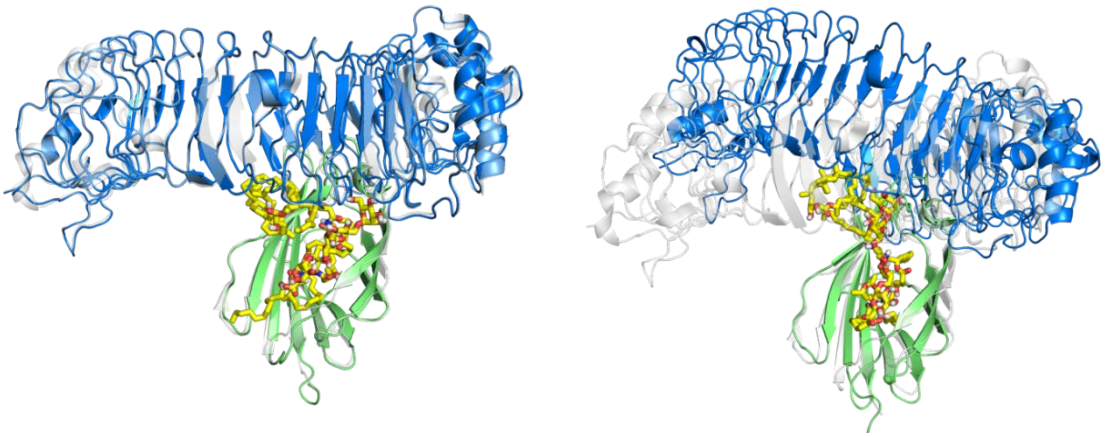


Figure 3.7. Evolution of the TLR4/MD-2/HOPA complex over the MD simulation. TLR4/MD-2 from PDB ID 3FXI, represented in semi-transparent cartoon, was superimposed for comparison purposes. TLR4, MD-2 and HOLA from the simulation are depicted in blue cartoon, green cartoon and yellow sticks, respectively. On the left: system at t=0 ns, from the docking calculation. On the right: system at t=100 ns of MD simulation.

Conclusion

Our computational studies have allowed the proposal of plausible binding modes of the recently characterized Bradyrhizobium LPS to the TLR4/MD-2 system. These binding modes account for the potent activity antagonizing the binding of *E. coli* LPS to the MD-2/TLR4 complex thus inhibiting its toxic effects.[22] It is likely that the TLR4

signaling modulation occurs by direct interaction with the TLR4/MD-2 complex, both in its hopanoid-containing and hopanoid-free forms. Our studies do not point toward a primary role of the hopanoid moiety in the biological activity regarding TLR4 signaling.

Materials and methods

Structure construction. The 3D structures of both the hopanoid-containing and HOLA and HF-LA were built with PyMOL molecular graphics and modeling package[24] based on the saccharide backbone of *E. coli* LPS retrieved from the PDB ID 3FXI. Atoms were modified and added accordingly and bond type and length were carefully selected and revised. The geometry of these two structures was further optimized with Maestro under the OPLS3 force field.[25] The antagonist conformation of the human TLR4/MD-2 complex was assembled by merging the ectodomain of TLR4 from RCSB (www.rcsb.org) PDB ID 3FXI and MD-2 from PDB ID 2E59. The latter was aligned to the spatial coordinates of the MD-2 present in 3FXI and solvent, ligands, and ions were removed.

Structure optimization. Hydrogen atoms were added to the X-ray structures using the preprocessing tool of the Protein Preparation Wizard of the Maestro package, and then the structures went through a restrained minimization under the OPLS3 force field with a convergence parameter to RMSD for heavy atoms kept default at 0.3Å.

Docking procedure. Gasteiger charges were computed and assigned with AutoDockTools 1.5.6 to both the proteins and the ligands. Both HOLA and HF-LA were left flexible by allowing some appropriately selected dihedral angles to rotate whereas the receptor was always kept completely rigid. The docking was performed with AutoDock Vina[26]. A cubic docking box of 60 Å in size and 1 Å in spacing was defined. The box was centered equidistant to the geometric center of residues Arg90 (MD-2), Arg96 (MD-2), and Arg264 (TLR4).

Parametrization. HOPA and HF-LA structures were split into residues to facilitate and homogenize the parametrization process. The partial charges and atom types of the 4-substituted and the 6-substituted 2,3-diamino-2,3-dideoxy-glucose monosaccharides composing the oligosaccharide backbone were established based respectively on residues 4YB (4-substituted GlcNac) and 6YA (6-substituted GlcNac) of the GLYCAM force field[27]. The partial charges and parameters for the two mannose (Man) residues and the galacturonic acid (GalA) were retrieved from the GLYCAM force field

respectively under the name 0MA, 6MA and 1OA. Partial charges for the primary and secondary acylation as well as for the hopanoid residue were derived, with the help of antechamber[28], following the standard GAFF procedure described in the AMBER manual and the parameters were assigned by the GAFF force field.

MD simulations. All MD simulations were performed with Amber14[29], the protein was described by the ff14SB all-atom force field[30], the pentasaccharide backbone of the BTAi1 lipid A by the GLYCAM_06j-1 force field[27] and the other constituents of the lipid A (the lipid chains and the hopanoid moiety) were parametrized with the GAFF force field.[31] The simulation box was designed such as the edges are distant of at least 10 Å of any atoms. The system was solvated with the TIP3P water molecules model. One Na+ ion was added to counterbalance the negative charge of the galacturonate group. All the simulations were performed with the same equilibration and production protocol. First, the system was submitted to 1000 steps of steepest descent algorithm followed by 7000 steps of conjugate gradient algorithm. A 100 kcal mol^{-1} A^{-2} harmonic potential constraint was applied on the proteins and the ligand. In the subsequent steps, the harmonic potential was progressively lowered (respectively to 10, 5, 2.5 and 0 kcal mol^{-1} A^{-2}) for 600 steps of conjugate gradient algorithm each time. Next, the system was heated from 0 K to 100 K by a Langevin thermostat in the canonical ensemble (NVT) under a 20 kcal mol^{-1} A^{-2} harmonic potential restraint on the proteins and the ligand. Finally, the system was heated up from 100 K to 300 K in the Isothermal-isobaric ensemble (NPT) under the same restraint condition than the previous step, followed by a simulation of 100 ps in which all harmonic restraints were removed. At this point the system was ready for the production run, which was performed using the Langevin thermostat in the NPT ensemble, at a 2 fs time step.

3.2 LPS-like synthetic modulators

3.2.1 Structure-activity relationship (SAR) in monosaccharide-based Toll-like receptor 4 (TLR4) antagonists

The content of this subsection was reported in a multidisciplinary article published in the *Journal of Medicinal Chemistry*.[32] The content was restructured to give the emphasis on the computational approach, relevant to this thesis. This work was performed in collaboration with Prof. Peri (University Milano-Bicocca, Italy), Prof. Beyaert (VIB-UGent, Belgium) and Prof. Jerala (National Institute of Chemistry, Slovenia) within the MSCA-ITN TOLLerant project.

Our group developed the lipid X mimetic FP7,[33] a glucosamine derivative with two phosphate groups and two myristic (C_{14}) FA chains, whose design was inspired by other glucosamine-based TLR4 modulators (Figure 3.8).[34-35] FP7 is active in inhibiting in a dose-dependent way human[33] and murine[36] TLR4 activation by LPS. Some preliminary observations from NMR experiments suggest that FP7 interact with MD-2, probably inserting FA chains into the hydrophobic binding cavity.[33] This direct competition with LPS for MD-2 binding is probably reinforced by the capacity of FP7 to induce endocytosis of CD14, thus causing the absence of this receptor on the plasma membrane.[33] FP7 is active in blocking PR8 virus lethality that is mainly due to TLR4 over-stimulation by endogenous DAMPs (mainly oxidized phospholipids and HMGB-1 protein) derived from viral damage to lung tissue.[36] In a proof-of-concept experiment in support of the proposed activation mechanism, FP7 was shown to inhibit HMGB-1 activation of dendritic cells.[36] Other monosaccharide-based TLR4 modulators were developed and structure-activity relationship (SAR) studies showed that the length of FA chain is a critical factor determining the potency of TLR4 antagonism or agonism.[34, 37] The biological activity and the agonist/antagonist behavior on TLR4 of lipid A variants and other amphiphilic glycolipids including FP7 is not only determined by the interaction with MD-2 but also by the aggregation state in solution. As LPS and lipid A, FP7 is an anionic amphiphile with a low value of CMC (9 μM).[33] Even though the CMC value of FP7 is higher than its IC50 (about 2 μM in HEK cells assays), equilibrium between aggregates and single molecules in solution is present in the concentration range in which FP7 is active.

It has been proposed for lipid A derivatives that the size and the 3D shape of aggregates influences the TLR4 activity, with lamellar aggregates being associated to antagonism and aggregates with non-lamellar cubic symmetry to agonism.[38-39] While the last step of ligand presentation to TLR4 and formation of the activated heterodimer $(TLR4/MD\text{-}2/ligand)_2$ are dominated by single molecule interactions between the ligand and CD14 and MD-2 receptors[40], the early phases of endotoxin (ligand) recognition by LPS-binding proteins are very likely influenced by the aggregation state of the ligand.

We performed the structure-activity relationship (SAR) study on synthetic FP7 variants differing only for FA chains lengths (10, 12, 14 and 16 carbon atoms, Figure 3.8). In this study we took into account both the interaction with MD-2 and the aggregation properties of the molecules. Additionally, we showed the relationship between the chemical structure of FP7 variants with different fatty acid chains lengths and their effect on functional activity of TLR4 in different *in vitro* cell models.

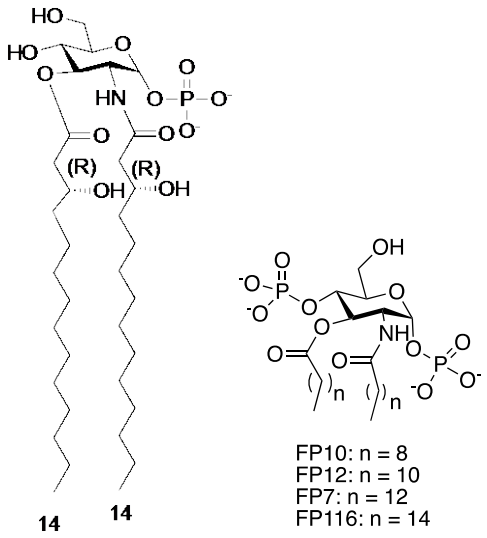


Figure 3.8. Chemical structures of lipid X and FP7 variants

Computational design of FP7 variants as ligands of human MD-2 and CD14.

Given our previous studies on the lipid X mimetic FP7 as ligand of TLR4/MD-2 and CD14 proteins, with TLR4/MD-2 antagonist activity,[33] we were prompted to investigate the influence of acyl chain length on the antagonist activity. To address this point, we designed three new FP7 derivatives with different FA lengths: FP10 (C_{10}), FP12 (C_{12}), FP7 (C_{14}), and FP116 (C_{16}). The ability of these ligands to bind to the TLR4/MD-2 complex and to CD14, compared with FP7, was initially assessed through various computational techniques.

We first docked the ligands in the binding site of CD14 using AutoDock Vina. For all the four ligands, docked poses inside the hydrophobic pocket were found. The obtained binding poses were very similar for all the ligands (Figure 3.9A) with also very close favorable predicted binding energies for the top poses (range from -6.5 kcal mol^{-1} to -5.9 kcal mol^{-1}). Therefore, the docking calculation showed that all four ligands are theoretically able to interact with CD14 inside its hydrophobic pocket and to engage in favorable interactions. In the most populated and most favorable docked poses, one phosphate group is interacting with the NH groups of Arg72 and Val73, and with the OH group of Tyr82 (Figure 3.9B), while the other phosphate group is exposed to the solvent. The FA chains are accommodated inside the hydrophobic pocket of CD14 interacting with aliphatic residues, mainly Ala, Val, Leu, and Ile, and aromatic Phe49 (details are depicted in Figure 3.9C). The results were in agreement with previous docking studies of FP7 reported by us.[41]

We performed the docking calculations of ligands FP7, FP10, FP12 and FP116 inside the TLR4/MD-2 complex in the antagonist conformation (Figure 3.10). For all the compounds, favorable docked poses were found, with predicted binding energies, for the best ones, ranging from -7.8 to -6.5 kcal mol^{-1}. The polar head groups are placed at the rim of MD-2 and the FA chains deep inside the hydrophobic pocket interacting with many hydrophobic residues, namely Val24, Ala30, Ile32, Ile44, Ile46, Val48, Ile52, Leu54, Leu61, Ile63, Tyr65, Phe76, Leu78, Ile80, Phe104, Val113, Ile117, Phe119, Phe121, Ile124, Tyr131, Val135, Phe147, Leu149, Phe151, and Ile153 (Figure 3.10B).

In addition, we observed more diversity in the predicted binding poses in TLR4/MD-2 than in CD14. Results for FP7 were in agreement with those previously reported in MD-2 protein.[33] In many poses, one of the phosphate groups was close to the hydroxyl group of MD-2 Tyr102 where it establishes hydrogen bonds, and the other one was often close to MD-2 Arg90 establishing hydrogen bonds and electrostatic interactions (Figure 3.10C). In some docked poses, the phosphate groups were observed to interact with the backbone of residues Phe119, Ser120, and Phe121. Both phosphate groups were often placed at the rim of MD-2 where they are exposed to the solvent, in agreement with the reported X-ray crystallographic complexes of TLR4/MD-2 with glycolipids (for example, complex with eritoran, PDB ID 2Z65, or with lipid IVa, PDB ID 2E59). Two different orientations were also found: type A (antagonist-like binding mode), similar to that found for lipid IVa in PDB ID 2E59; and type B (agonist-like

binding mode), similar to that found for *E. coli* lipid A in PDB ID 3FXI (Figure 1.6 in *Introduction* chapter). It is well known that these two ligands, lipid IVa and *E. coli* lipid A, bind to TLR4/MD-2 in a different manner, one being rotated 180° compared to the other one, leading to opposed biological activities.

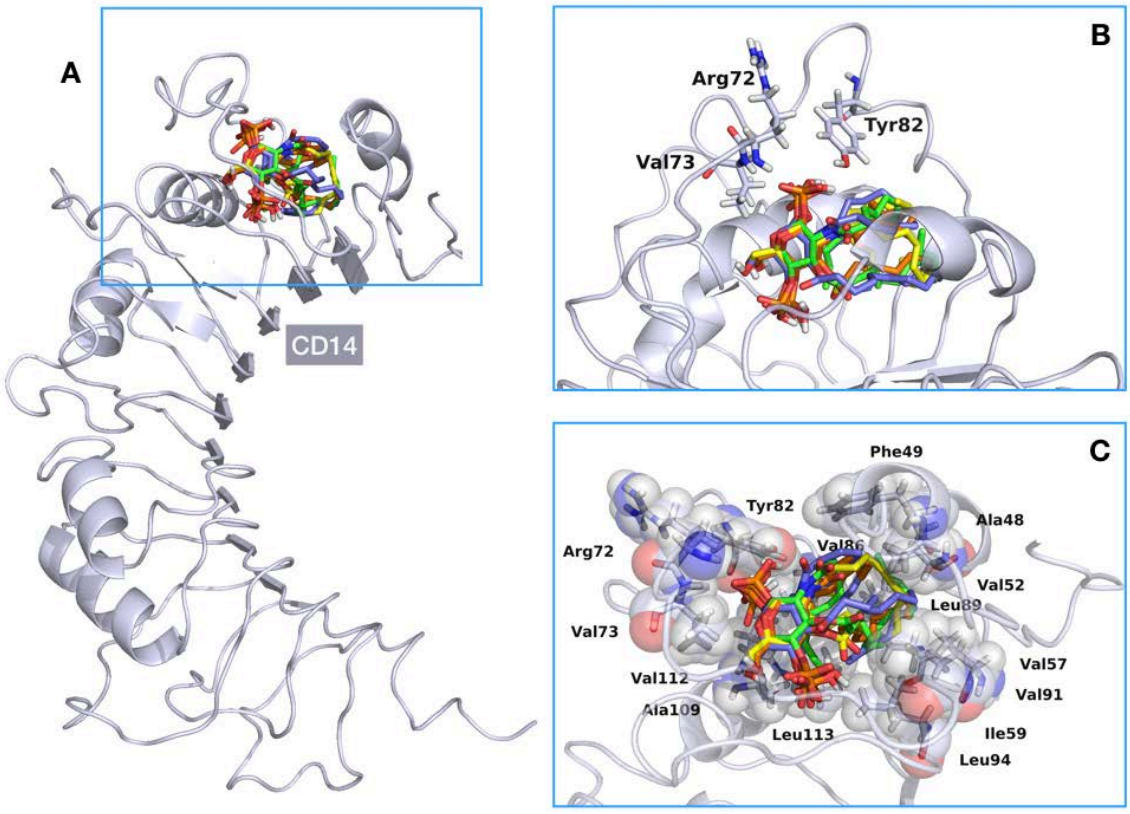


Figure 3.9. (**A**) Full view of CD14 (in light blue). Superimposed best-score docked poses of FP10 (in orange), FP12 (in yellow), FP7 (in green), and FP116 (in violet) are shown in sticks. (**B**) Close-up of the binding pocket of CD14, showing the major interactions of the head groups of the ligands docked within the protein. (**C**) Top view of the binding pocket of CD14. CD14 residues that form the hydrophobic pocket and whose side chains are close to the FA chains of the docked ligands are displayed in sticks and partially transparent spheres.

Selected binding poses were used as starting structures for re-docking with AutoDock4 resulting in predicted binding energies ranging from -4.6 kcal mol^{-1} to +4.3 kcal mol^{-1}. Among the docked solutions, the best poses (from -4.6 kcal mol^{-1} to -2.5 kcal mol^{-1}) corresponded to binding poses very similar to those obtained with AutoDock Vina (data not shown). The narrow binding energy range did not permit to rank the ligands by predicted affinity, showing that the four ligands are putative binders of the TLR4/MD-2 system. Given that the main interactions (the polar ones) are common to the four ligands, and that the MD-2 pocket is big enough to host two longer FA chains, from the

docking calculations it was not possible to clearly correlate the subtle differences in FA chain length with preferred ligand binding.

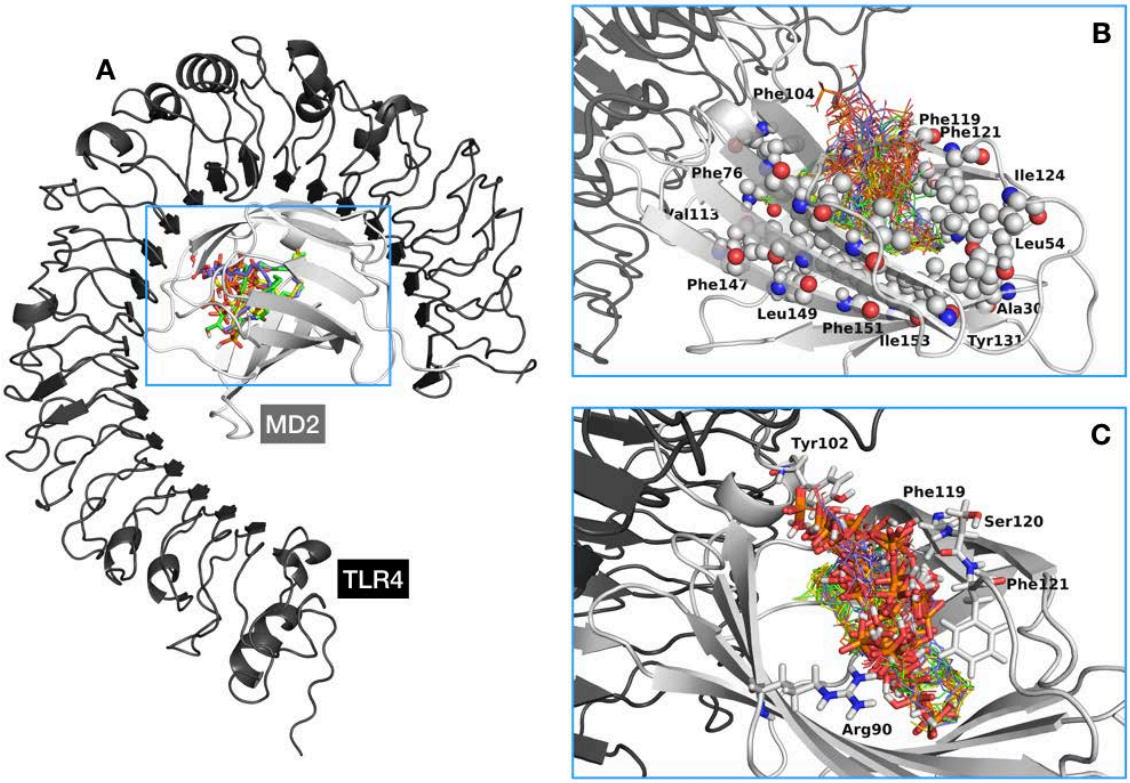


Figure 3.10. (**A**) General view of FP10 (in orange), FP12 (in yellow), FP7 (in green), and FP116 (in violet) docked inside TLR4/MD-2 (TLR4 is shown in black and MD-2 in grey). (**B**) Details of the MD-2 hydrophobic pocket occupied by all the best docked poses for each ligand (represented as lines). Hydrophobic residues mentioned in the text as interacting with the FA chains of the ligands are represented in spheres. (**C**) Details of the polar interactions of the ligands inside the TLR4/MD-2 system. Phosphate groups of the best docked poses of each ligand and the MD-2 residues with which they interact are represented in sticks.

Stability of the predicted TLR4/MD-2/ligand complexes was further studied by MD simulations. We selected two of the best binding poses for each ligand (Figure 3.11): one type A (antagonist-like binding pose), and one type B (agonist-like binding pose). Therefore, a total of eight 50-ns MD simulation were run.

A. Type A binding (antagonist-like) **B.** Type B binding (agonist-like)

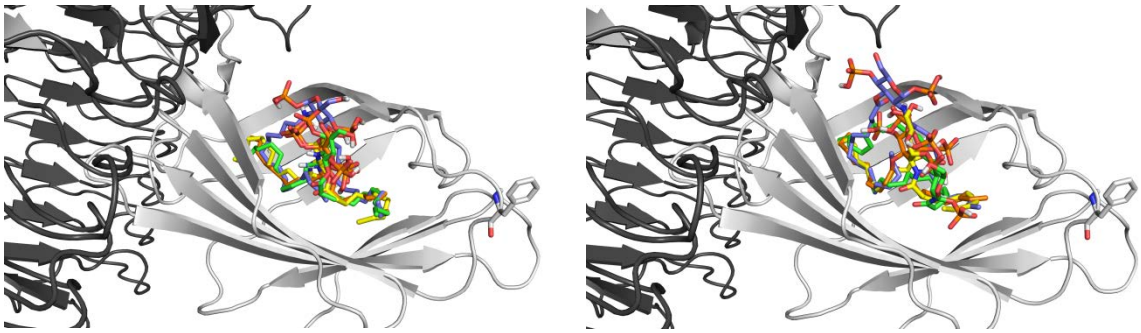


Figure 3.11. Selected docked poses of each ligand for MD simulations. (**A**) Type A antagonist-like binding pose. (**B**) Type B agonist-like binding pose. Ligands are depicted following the CPK coloring scheme, excepting the carbon atoms that are shown in orange for FP10, in yellow for FP12, in green for FP7, and in violet for FP116. TLR4 is colored dark, MD-2 is shown in light grey and Phe126 side chain is represented in sticks.

We monitored the motion of MD-2 over time and examined the RMSD and RMS fluctuation per residues, as well as the motion of Phe126 side chain over time (Figure 3.12). All the complexes showed stable ligand-receptor interactions along the MD simulation as predicted by the docking calculations. In particular, in the MD simulation of the TLR4/MD-2/FP7 complex in the type A (antagonist-like) binding pose, the Phe126 side chain moves around its initial position staying largely exposed to the solvent in a conformation in agreement with the X-ray crystallographic antagonist conformation of MD-2 (Figure 3.15B).

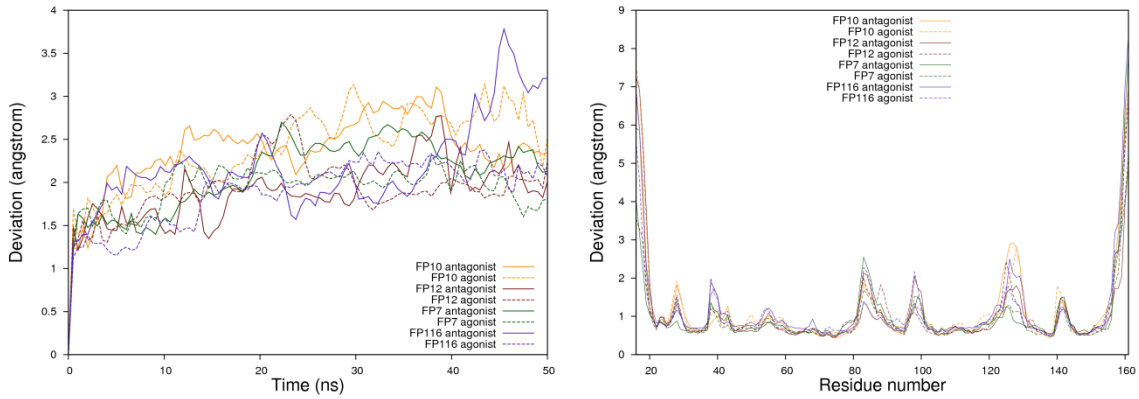


Figure 3.12. Molecular dynamics simulations of the TLR4/MD-2 system in complex with ligands FP10, FP12, FP7, and FP116. On the left: RMSD of the MD-2 backbone over time. On the right: RMS fluctuations per residues of MD-2.

A) B)

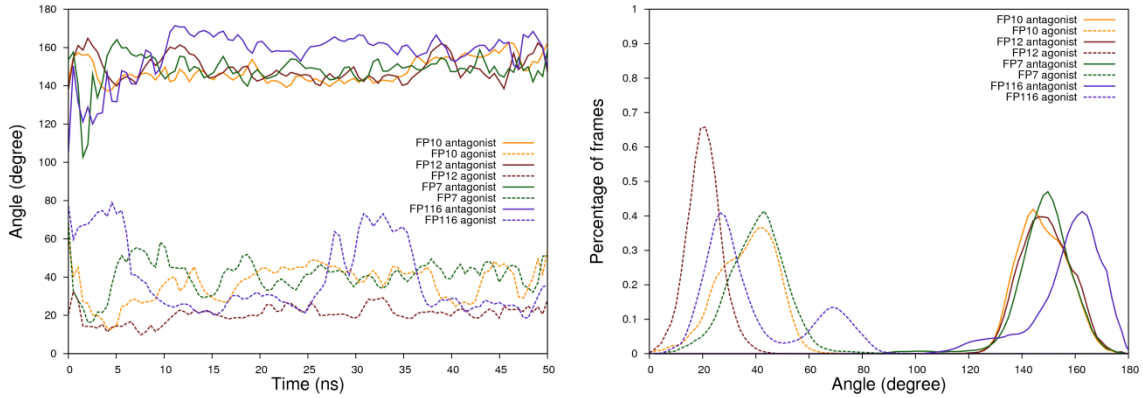


Figure 3.13. (**A**) Angle between two vectors along the MD simulation projected over time. First vector was defined between the amide α-carbon atom and the ester α-carbon atom of the ligand, and the second vector between the α-carbon atoms of residues Pro78 and Thr105 of MD-2 (Figure 3.14). Angle between 0 and 90 degrees is characteristic of the type B binding (agonist-like) as observed in the PDB ID 3FXI (TLR4/MD-2/lipid-A complex); angle between 90 and 180 degrees is characteristic of the type A binding mode (antagonist-like) as observed in the PDB ID 2E59 (TLR4/MD-2/lipid-IVa complex). (**B**) Percentage of time frames along the MD simulation projected over angle.

To evaluate the relative orientation between the ligands and MD-2, we arbitrarily defined two vectors, one from the amide α-carbon atom to the ester α-carbon atom of the ligand, and another one from the α-carbon of residues Pro78 to Thr105 of MD-2 (Figure 3.14A). The angle between these two vectors was plotted both over time and as a percentage of frames per 0.1 degree angle range (Figure 3.13). It was observed that none of the ligands undergoes orientation flip during the 50-ns simulations, all remaining in the orientation obtained from the docking process. Interestingly, only in the case of the TLR4/MD-2/FP116 complex with FP116 in the type A (antagonist-like) binding pose, the orientation of Phe126 side chain flips over (Figure 3.15C). We monitored this flipping behavior along the MD simulations, for all the ligands, by arbitrary choosing two vectors, within MD-2, both starting from the α-carbon of residue Phe126 to, respectively, the phenyl C-4 atom of the same residue and the α-carbon of residue Ser21 (data shown in Figure 3.14 and Figure 3.15). This observation could suggest that FP116 is not able to efficiently retain an antagonist conformation of MD-2, thus pointing to a poor antagonist capacity.

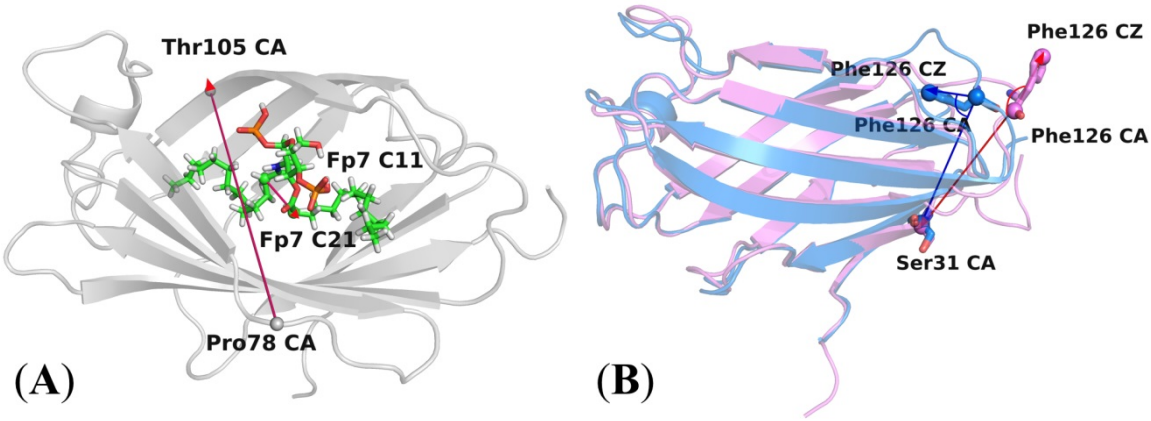


Figure 3.14. (**A**) Representation of the vector from the α-carbon (CA) of Pro78 to the α-carbon of Thr105, and the vector from amide α-carbon atom (C11) and the ester α-carbon atom (C21) of FP7, used to follow the orientation of the ligands along the MD simulations (cf. S4). FP7 is used as an example; the same applies for the other ligands. MD-2 is represented in grey an FP7 is depicted in CPK coloring at the exception of the carbon atoms that are in green. (**B**) Representation of two vectors, within MD-2, starting both from the α-carbon (CA) of residue Phe126 to, respectively, the phenyl C-4 atom (CZ) of the same residue and the α-carbon (CA) of residue Ser21. Agonist MD-2 from PDB ID 2E59 and antagonist MD-2 from PDB ID 3FXI are represented in semi-transparent blue and pink cartoons, respectively.

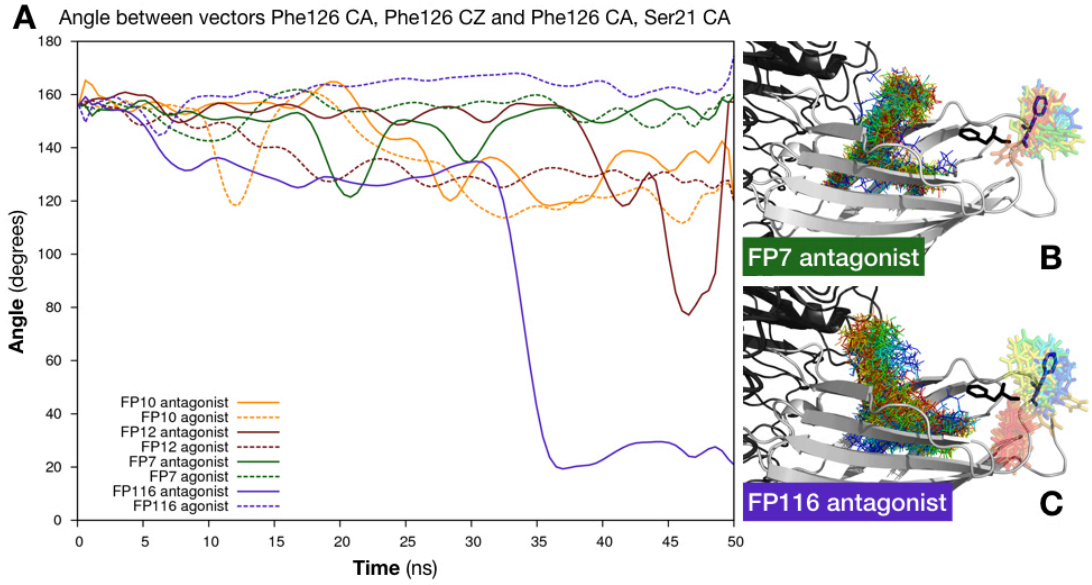


Figure 3.15. (**A**) Angle between two vectors defined in Figure 3.14. The angle between these two arbitrarily selected vectors, plotted over time, illustrates the flip that residue Phe126 undergoes during the MD simulation when MD-2 is engaged by FP116 in the type A binding mode. (**B-C**) Superimposition of snapshots (one for each simulated nanosecond), from the MD simulations, colored from blue, t=0 ns, to red, t=50 ns. From these snapshots only ligands (as lines) and residue Phe126 (as sticks) are made visible. (**B**) TLR4/MD-2/FP7 complex starting from the type A binding pose (antagonist-like). (**C**) TLR4/MD-2/FP116 complex starting from the type A binding pose (antagonist-like).

In addition, logP values of compounds FP10, FP12, FP7 and FP116 were computationally calculated, ranging from approximately 4 to 10 with a linear distribution (Figure 3.16). The highest logP value was obtained for FP116 indicating a high lipophilicity that might result in low water solubility. This was in agreement with the lower acyl chain mobility as analyzed by FT-IR spectroscopy (Figure A3.1). In any case, this did not interfere with the performance of the cell assays. Summarizing, the computational studies assessed the ability of ligands FP7, FP10, FP12, and FP116 to bind both CD14 and TLR4/MD-2, pointing to the long FP116 acyl chain (C_{16}) as the maximum length bordering good (predicted) binding properties.

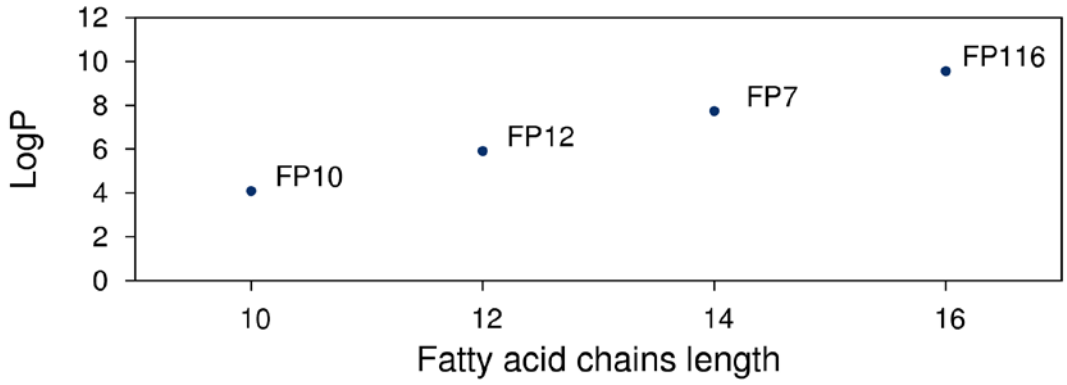


Figure 3.16. Computed logP values for compounds FP10, FP12, FP7 and FP116.

The compounds were therefore synthesized and tested. Details for the synthesis and the biological assays can be found in the article reported by us.[32] Below, the discussion based on the biological activity studies is provided.

Biological activity studies and discussion

The homologous series of FP glycolipids with fatty acid chain lengths varying from 10 to 16 carbon atoms were rationally designed as MD-2 ligands and synthesized. In a first set of *in vitro* experiments we aimed at studying the SAR of these molecules in binding experiments with functional hMD-2. For this purpose, hMD-2 expressed in yeast (*P. pastoris*), was used because it showed higher activity in responding to LPS stimulus than bacterial (*E. coli*) MD-2 and was produced with higher yields than MD-2 from mammalian (HEK) cells (Figure A3.2). Four different binding experiments between synthetic compounds and hMD-2 were carried out. These were competition (displacement) experiments in which the synthetic glycolipids compete with biotin-LPS, with the fluorescent MD-2 ligand bis-ANS and with anti-MD-2 antibody for MD-2 binding (Figure A3.3). SPR measurements allowed to analyze directly the binding

between synthetic glycolipids and MD-2. All binding experiments consistently provided the same order of affinity among hMD-2 and synthetic molecules: FP12(C_{12})>FP7(C_{14}) >FP10(C_{10})>FP116(C_{16}).

The biological activity was then assessed on cells: when provided alone, the synthetic FP compounds did not display any TLR4 agonist activity in human and murine cells. On the contrary, when administrated with LPS, the molecules with 10, 12 and 14 carbon chains (respectively, FP10, FP12 and FP7) were active in blocking LPS/TLR4 signal (antagonism) in human and murine cells, while the molecule with 16 carbons (FP116) showed very weak, or no activity. The order of activity of FP variants as TLR4 antagonists was confirmed in human HEK-TLR4 (Figure A3.4) and THP-1 (Figure A3.6 and Figure A3.7), and murine RAW macrophages cells (Figure A3.5). The molecules with 10, 12 and 14 carbon chains seem to be non-species specific TLR4 antagonists, because these compounds are active in both human (HEK and THP-1) and murine cells, with higher potency in human ones. The compound with higher biological activity was FP12, with 12 carbons, followed by FP7 and FP10 with 14 and 10 carbons, while FP116 with 16 carbons showed very weak or no activity in cell models.

The variation of compounds' functional activity was related to the number of carbon atoms of the aliphatic chains which could be described by a bell-shaped curve with a maximum at C_{12}. This is a common structure-activity trend that is found in a number of series of homologous compounds in medicinal chemistry and can be explained in terms of docking within the binding pocket of the pharmacological target (as it exists an optimal number of carbon atoms that can be accommodated into the pocket) and also in terms of variation of solubility and bioavailability (when the chain length is too long the solubility decreases and also the biological activity). Thus, the difference of TLR4 functional activity of FP monosaccharides related to FA chains length can be explained in terms of their interaction with MD-2(/TLR4) and/or by their aggregation properties in solution (Figure A3.1 and Figure A3.8).

The docking and MD simulation studies have shown that FP10, FP7 and FP12 would accomplish optimal binding properties while FP116 could be bordering the limits of the maximum length compatible with a proper MD-2 binding. Although MD-2 pocket is able to host up to five FA chains, the highly long and flexible C_{16} acyl chains present in FP116 seem to point to less efficient ability to interact with TLR4/MD-2 in an

antagonistic binding mode, given that the required exposed conformation of Phe126 side chain could be jeopardized. Additionally, calculated logP values for the FP variants point to a very high lipophilicity for FP116, maybe affecting the aggregation properties in solution (Figure 3.16). Taken together, these data strongly suggest that the mechanism of TLR4 antagonism of that class of compounds is mainly based on the competition with LPS (or other ligands, as bis-ANS) in the binding to the MD-2/TLR4 complex.

Interestingly, an identical order of activity on TLR4 has been found in a series of monosaccharide TLR4 agonists, the Gifu Lipid As (GLA), and the following order of potency in inducing the production of TNF-α in murine cells was detected: C_{12}>C_{14}>C_{10}>>C_{16}.[34] Also in the case of GLA compounds, with three FA chains and one phosphate in C-4 position, the C_{12} and C_{14} variants were the most active ones, C_{10} less active and C_{16} were inactive. Similarly to FP compounds, GLA are more active on murine than on human cells.[34] However, the authors did not provide any evidences or explanation about the link between TLR4 activity of monosaccharide and FA chain length.

Regarding the aggregation properties some important differences among FP compounds were detected by FT-IR analysis in solution (Figure A3.1). These measurements showed marked variations in acyl chain fluidity of aggregated FP compounds depending on the chemical structure. The phase transition temperature Tc exhibits a clear inverse correlation with the length of the acyl chains with Tc C_{16} >> Tc C_{14} > Tc C_{12} > Tc C_{10}. Of note, this behavior results in marked differences at the biological relevant temperature of 37 °C, where FP10, FP12, and FP7 are in a fluid membrane phase, whereas FP116 is still in a rigid membrane phase and requires much higher temperatures for acyl chain melting to occur. The occurrence of a very broad phase transition at temperatures above 37 °C and occurrence of a second phase transition at higher temperature as observed for FP116 were also found for inactive glucosamine monosaccharide GLA compounds.[40] Differences in phase behavior have also been shown for the TLR4 ligands lipid A and LPS. The antagonistic tetraacylated synthetic compound 406 is highly fluid at 37 °C, whereas the biologically active hexaacylated compound 506 and LPS Re have phase transition temperatures above 37 °C.[41] The fluidity state of the acyl chains in aggregated glycolipids is thus not an exclusive determinant of inflammatory or antagonistic activity of chemically different

compounds. It is rather a modifying parameter of biological activity by affecting aggregate properties such as hydrophobic thickness, packing density, and aggregate stability. NMR and SAXS analysis revealed striking differences in aggregate formation of FP compounds which are likely to explain differences in their biological activity (Figure A3.8). Concentration-dependent NMR analysis of the two most antagonistic compounds FP12 and FP7 revealed aggregation of FP7 (C_{14}) at much lower concentrations than FP12 (C_{12}), reflecting further differences in the biophysical state and bioavailability of these compounds (data not shown). Aggregate structures resolved by SAXS analysis provided evidence for lamellar bilayer structures for FP10 and FP12, which are associated with antagonistic activity, for FP7 a tendency to for non-lamellar structures was determined (Figure A3.8). Considering the crucial role of lipid supramolecular aggregate structure for the presentation to LPS receptor molecules, the different aggregate structures observed by SAXS might explain the slightly lower antagonistic activity of FP7 compared to FP12 in some biological systems.

Conclusion

The present study provides structural and functional biological data demonstrating the ability of novel FP variants to negatively regulate TLR4 signaling in different cell model systems. Having shown the strong potential of FP12 to modulate second messengers activation and various end points of TLR4 signaling pathways including its lack of toxicity, this study supports the idea of further drug development of FP12 as a lead compound in preclinical and clinical studies for pharmacological intervention of inflammatory-based diseases. Our computational studies were relevant in the context of the SAR study and to propose the rationale for the mechanism of binding. Our models suggest that there is an optimum length for the FA chains for an appropriate TLR4 antagonist activity related to the binding mode and to the physical-chemical properties of the FP variants.

Materials and methods

Structure construction and refinement. The 3D structures of ligands FP10, FP12, FP7, and FP116 were built with PyMOL[24] using 6YA monosaccharide found in the GLYCAM database[42] as a template. The 3D coordinates of human TLR4/MD-2 model in the antagonist conformation is reported elsewhere.[23]

Parameters Derivation. The parameters needed for MD simulations were obtained using the standard Antechamber procedure in Amber14.[29] Briefly, ligand structures, already refined at the AM1 level of theory, were optimized and their atomic partial charges were calculated with Gaussian g09/e1[43] at the Hartree-Fock level (HF/6-31G* Pop=MK iop(6/33=2) iop(6/42=6)), then the partial charges were derived and formatted for AmberTools15 and Amber14 with Antechamber, assigning the general AMBER force field (GAFF) atom types. Later, the atom types of the atom constituting the saccharide ring were changed to the GLYCAM force field atom types.[27] The GAFF parameters for the phosphate group were modified as shown in SI.

Docking calculations of ligands FP10, FP12, FP7, and FP116. The Gasteiger charges were computed within the AutoDockTools 1.5.6 program,[44] and the nonpolar hydrogens were merged for all the ligands, the human TLR4/MD-2 antagonist model and human CD14 (PDB ID 4GLP). AutoDock Vina 1.1.2 was used for the docking of the ligands and AutoDock 4.2 was used to redock the best-predicted binding poses. In AutoDock 4.2, the Lamarckian evolutionary algorithm was chosen and all parameters were kept default except for the number of genetic algorithm runs that was set to 200 to enhance the sampling. AutoDockTools 1.5.6 was used to assign the Gasteiger-Marsili empirical atomic partial charges to the atoms of both the ligands and the receptors. The structure of the receptors was always kept rigid, whereas the structure of the ligand was set partially flexible by providing freedom to some appropriately selected dihedral angles. Regarding the docking boxes, grid spacing was set to the default value of 1 Å for VINA, and 0.375 Å for AutoDock. For human CD14 structure, the size of the box was set to 33.00 Å in the x-axis, 33.75Å in the y-axis and 33.75 Å in the z-axis, and the center of the box was located equidistant to the center of mass of residues Phe69, Tyr82 and Leu89. For the human TLR4/MD-2 system, the size of the box was set to 33.00 Å in the x-axis, 40.50 Å in the y-axis and 35.25 Å in the z-axis, and the center of the box was located equidistant to the center of mass of residues Arg90 (MD-2), Lys122 (MD-2) and Arg264 (TLR4).

Molecular dynamics (MD) simulations. Selected docked complexes were submitted to MD simulations for 50 ns in the Amber14 suite following the same protocol described in section *3.1.1 Bradyrhizobium LPS.*

LogP calculations. LogP value of FP10, FP12, FP7, and FP116 were calculated within the Maestro package.[45]

3.2.2 Annex 1

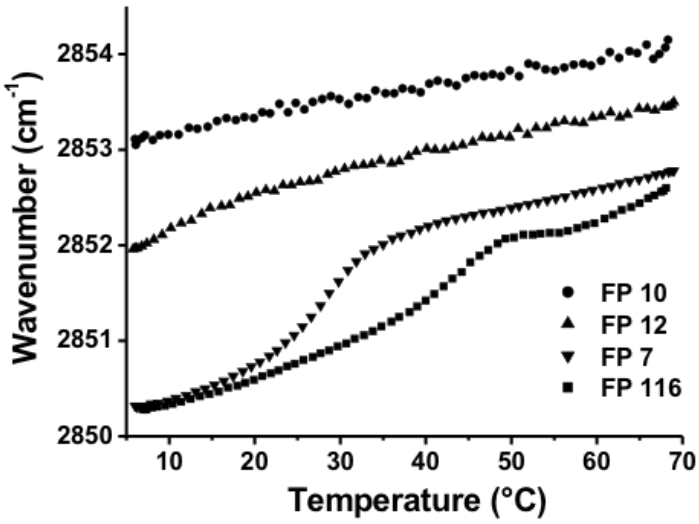


Figure A3.1. Acyl chain mobility of the aggregated FP compounds in dependence on temperature. The infrared absorption around wavenumbers 2850 - 2852 cm^{-1} corresponds to the symmetric stretching vibrations ν_s of the CH_2 groups of the acyl chains. The wavenumbers indicated were derived from the peak absorption of $\nu_s(CH_2)$ determined upon constant heating of the samples. Data are representative of two independent measurements.

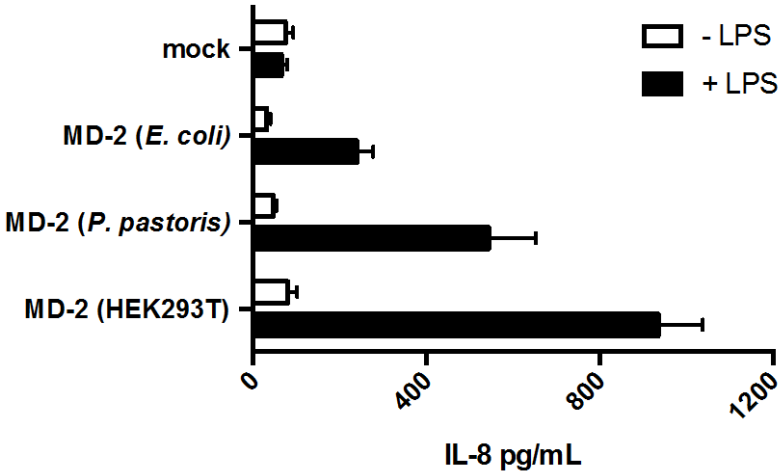


Figure A3.2. Activity of hMD-2 expressed in different hosts. The figure shows the maximum activation of TLR4 (quantified by IL-8 production) at the lowest concentration of hMD-2 under the different expressed conditions (bacteria 245 nM, yeast 15 nM, and mammalian 12 nM). Results are mean ± SEM from three parallels representative of at least three independent experiments.

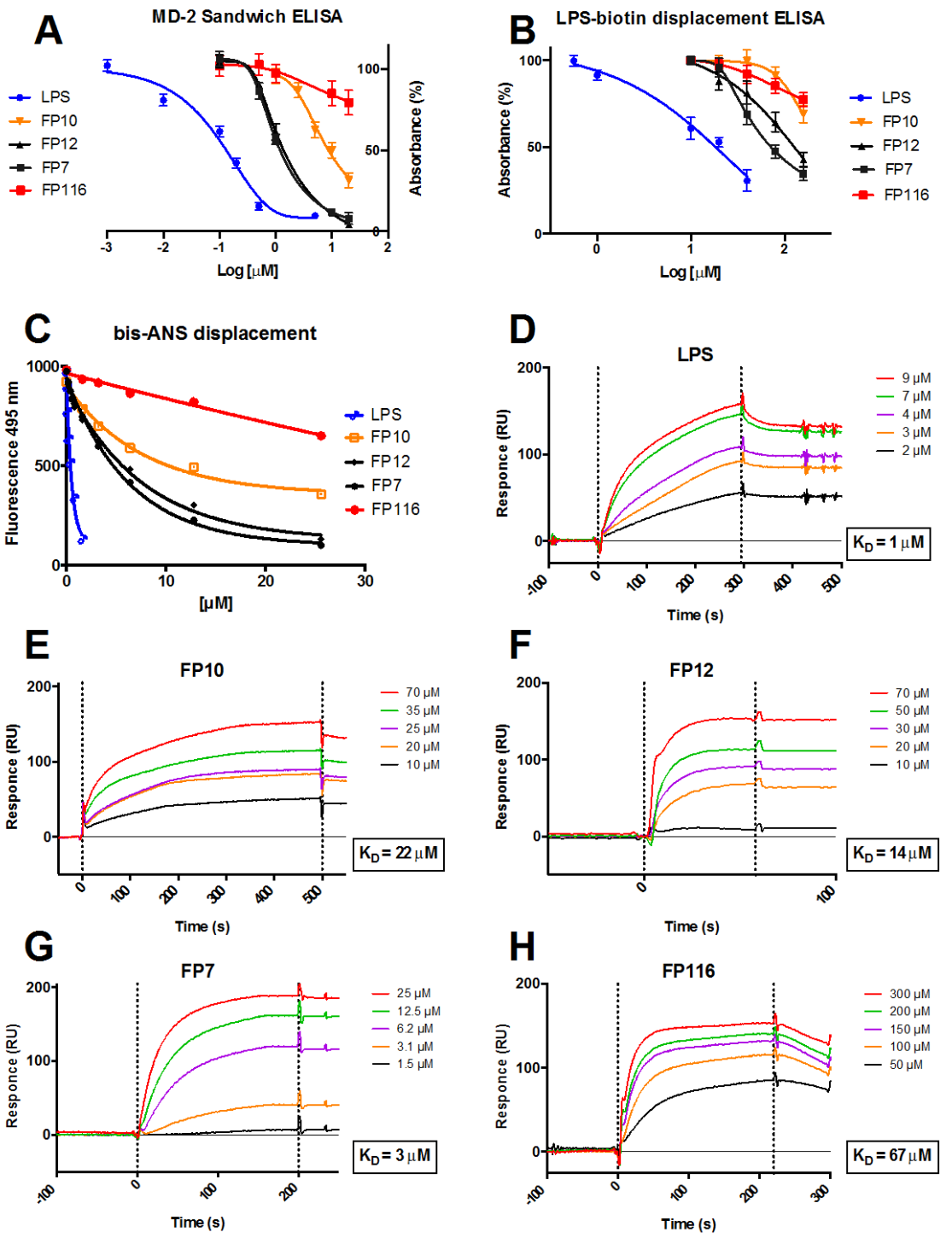
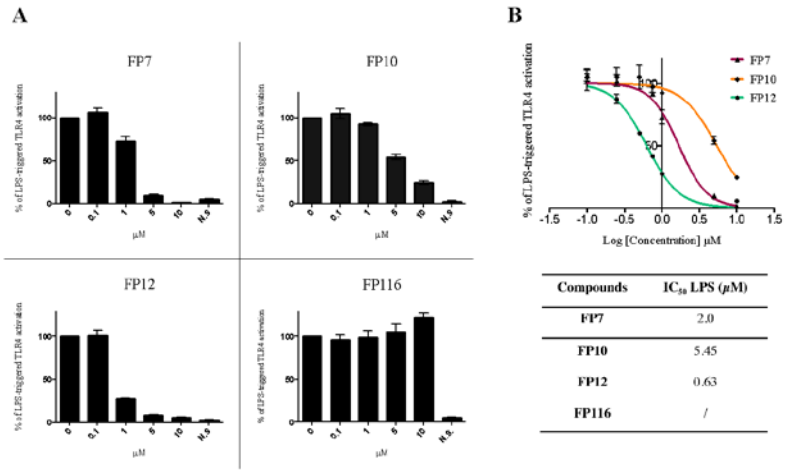


Figure A3.3. Cell-free binding studies on purified MD-2 receptor. (**A**) LPS, FP7, FP10 and FP12 prevent anti-human MD-2 monoclonal antibody binding in a dose-dependent manner; (**B**) LPS, FP7, FP10, FP12 and FP116 activity in competing with biotin-LPS for hMD-2 binding; (**C**) Fluorescence measurements show that LPS, FP7, FP10 and FP12 dose-dependently inhibit the binding of bis-ANS to MD-2; (**D-H**) SPR analysis show direct interaction between LPS, FP10, FP12, FP7, and FP116 and MD-2; K_D values are reported. Results are mean ± SEM from three parallels representative of at least three independent experiments.

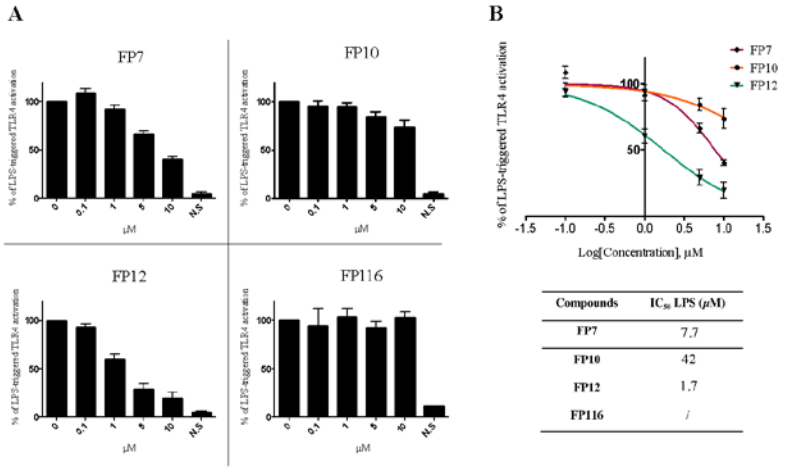
Modulation of LPS-stimulated TLR4 signaling in HEK-Blue cells



Compounds	IC_{50} LPS (µM)
FP7	2.0
FP10	5.45
FP12	0.63
FP116	/

Figure A3.4. Dose-dependent inhibition of LPS-triggered TLR4-dependent NF-kB activation in HEK-Blue hTLR4 cells by compounds FP7, FP10, FP12 and FP116. (**A**) HEK-Blue hTLR4 cells were pre-incubated with the indicated concentrations of compounds FP7, FP10, FP12 and FP116 and stimulated with LPS (100 ng/mL) after 30 minutes. Data were normalized to stimulation with LPS alone and expressed as the mean percentage ± SEM of at least three independent experiments. (**B**) Dose-response curves for compounds FP7, FP10 and FP12 in inducing the TLR4-dependent NF-κB reporter activity. Concentration-effect data were fitted to a sigmoidal 4-parameter logistic equation to determine IC_{50} values. Data points represent the mean of percentage ± SEM of at least 3 independent experiments.

Modulation of LPS-stimulated TLR4 signaling in murine macrophages



Compounds	IC_{50} LPS (µM)
FP7	7.7
FP10	42
FP12	1.7
FP116	/

Figure A3.5. Activity of compounds FP10, FP12 and FP116 on RAW-Blue cells**.** (**A**) RAW-Blue cells were pre-incubated with increasing concentrations of synthetic compounds and then stimulated with LPS (10 ng/mL, after 30 minutes). Data were normalized to response to LPS and expressed as the mean percentage ± SEM of at least three independent experiments. (**B**) Dose-dependent inhibition curves of compounds FP7, FP10 and FP12. IC_{50} values in the table on the bottom. Concentration-effect data were fitted to a sigmoidal 4-parameter logistic equation to determine IC_{50} values. Data points represent the mean of percentage ± SEM of at least 3 independent experiments.

Effect of FP variants on LPS-induced TLR4 signaling in THP-1 cells.

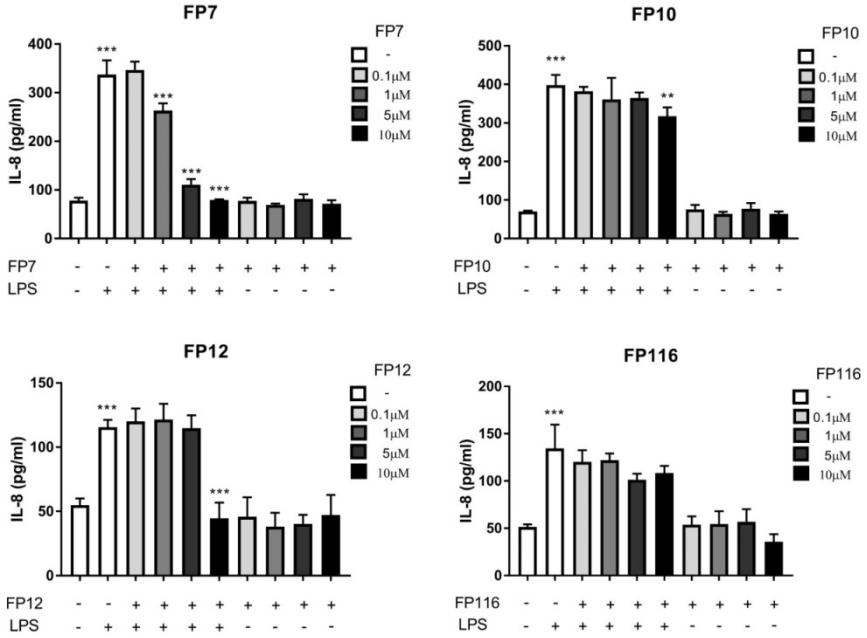


Figure A3.6. The Effect of FP variants on LPS/TLR4 induced production of IL-8 in THP-1 cells. THP-1 cells were pre-treated with FP variants (0 - 10 μM) for 1 h prior to LPS exposure. Cells were then left to incubate 16 h further in the presence or absence of LPS (100 ng/mL). IL-8 production was measured by ELISA. Results are displayed as mean concentration ± SD of three independent experiments. Significant results are indicated as *P>0.05 **P>0.01 ***P>0.001 for LPS vs Control and LPS vs FPs treated samples (Anova).

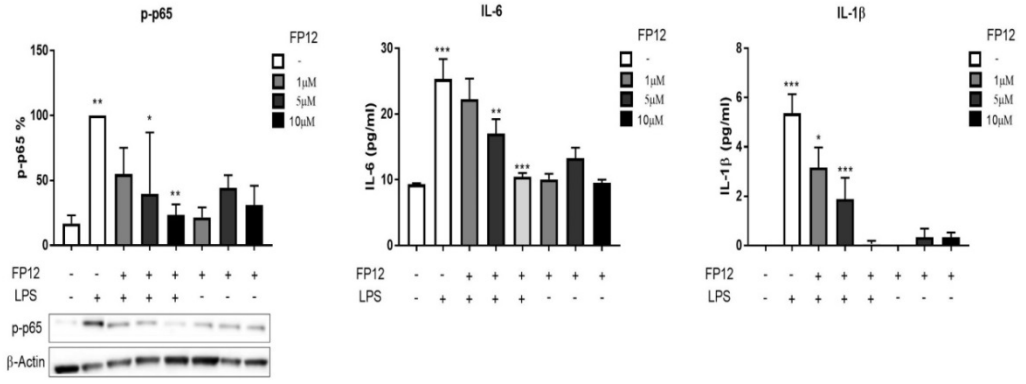


Figure A3.7. FP12 negatively regulates p65 NF-κB phosphorylation and production of IL-6 and IL-1β in THP-1 cells. THP-1 cells were pre-treated with compound FP12 (0 - 10 μM) for 1 h prior to LPS exposure. Cells were then left to incubate 1 and 16 h further in presence or absence of LPS (100 ng/mL). p65 NF-κB phosphorylation was determined in cell lysates using Western Blot analysis and cytokine production was measured by ELISA after 16 h of LPS exposure respectively. Results are displayed as mean concentration ± SD of three independent experiments. Significant results are indicated as *P>0.05 **P>0.01 ***P>0.001 for LPS vs Control and LPS vs FP12 treated samples (Anova).

SAXS studies

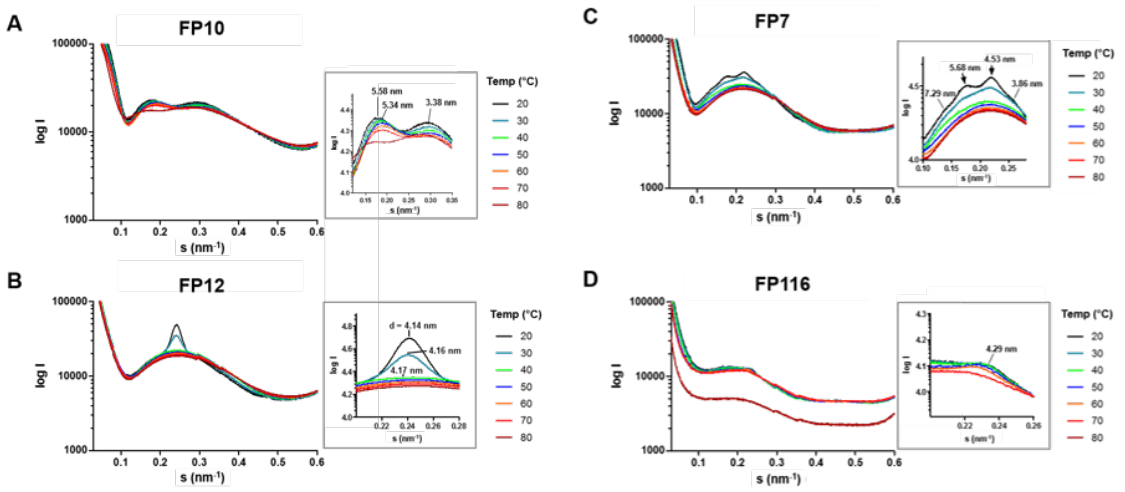


Figure A3.8. Small angle X-ray diffraction of aggregates in solution for FP10 (A), FP12 (B), FP7 (C), and FP116 (D). Scattering vectors are indicated for temperatures between 20 °C and 80 °C. Grey squares show enlargements of the relevant scattering vectors. The spacing of the diffraction maxima is indicted as d = 1/s (nm).

3.3 Non LPS-like modulators

3.3.1 Amphiphilic Guanidinocalixarenes Inhibit Lipopolysaccharide (LPS)- and Lectin-Stimulated Toll-like Receptor 4 (TLR4) Signaling

The content of this subsection was reported in a paper published in the *Journal of Medicinal Chemistry*.[23] Here the presentation of the results has been reorganized to emphasize the computational approach relevant to the thesis. This work was performed in collaboration with Prof. Peri (University Milano-Bicocca, Italy) within the MSCA-ITN TOLLerant project.

To block abnormal TLR4 signaling in bacterial sepsis, two different strategies have been developed. The first one is based on LPS neutralization by the formation of noncovalent adducts with cationic compounds: positively charged antimicrobial peptides (AMPs)[46] including polymixin B,[47] and synthetic dendrimeric polyamines[48-49] contain positively charged groups (most frequently amino and guanidinium groups) and form noncovalent complexes with negatively charged LPS, thus preventing LPS from interacting with the receptors.

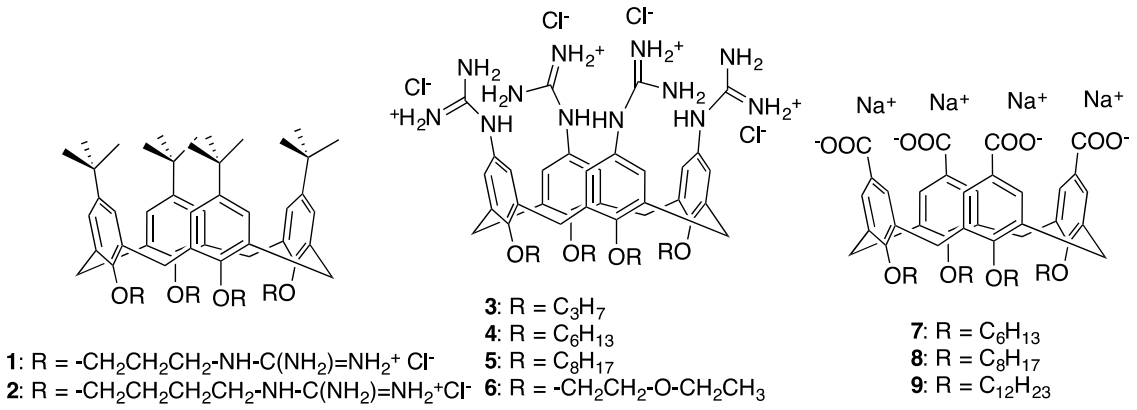
The second strategy is based on the use of molecules that compete with endotoxic LPS in binding to the same site on CD14 and MD-2, thereby inhibiting the induction of signal transduction by impairing LPS-initiated receptor dimerization. To date, several lipid A variants, which specifically block the LPS-binding site on human hMD-2, have been identified: natural compounds such as lipid IVa (a biosynthetic precursor of *E. coli* lipid A)[50] and a nonpathogenic lipid A from *R. sphaeroides*,[51] and synthetic molecules such as the tetraacylated disaccharide eritoran (E5564),[52-53] the aminoalkyl glucosaminide 4-phosphates (AGPs),[54-55] and some phosphorylated monosaccharide glycolipids.[37] These compounds inhibit TLR4 signaling by accommodating into the deep hydrophobic pocket of the co-receptor, MD-2, and blocking ligand-induced dimerization.[56] Eritoran[57] and other small molecules with TLR4 antagonist activity[58] also potently inhibit LPS binding to CD14. While the use of LPS neutralizing agents is limited to sepsis and septic shock, TLR4 antagonists that directly bind CD14 and MD-2 have potential also as therapeutics to treat neuroinflammations[59] and viral syndromes[60] caused by DAMP-TLR4 signaling. We recently observed that glycoamphiphiles with a sugar core (trehalose or glucose) functionalized with lipid chains and positively charged ammonium groups are able to inhibit LPS-stimulated TLR4 signal in vitro with IC_{50}

values ranging from about 5 to 0.2 μM and to reduce TLR4-dependent production of inflammatory cytokines in vivo.[61] The main structural feature of these molecules is their "facial" arrangement with positive charges and lipophilic chains disposed in spatially well-defined regions.

Therefore, we hypothesized that calixarene-based facial amphiphiles could also be suitable as scaffolds to obtain TLR4 ligands with antagonist activity. Actually, in a biological context, amphiphilic calixarenes showed remarkable properties significantly related to their amphiphilicity.[62] The calixarene scaffold represents a very versatile structure to build amphiphilic compounds due to the possibility of variably and selectively functionalizing both its upper (aromatic para positions) and lower (phenolic oxygens) rim. Moreover, the possibility of linking to the macrocyclic platform several binding moieties, resulting in preorganized arrays, gives rise to systems that, exploiting a multivalent effect, frequently show improved biological activity with respect to corresponding monovalent models.[62-63] From this point of view, also the tight compaction of hydrophobic chains located at one of the rims can result in the enhancement of some properties such as (self)assembling capabilities in an aqueous environment.[62-65]

We present here a study on the inhibition of TLR4/MD-2 signaling by a series of positively and negatively charged calixarene-based amphiphiles (compounds **1**-**6** and **7**-**9** in Figure 3.17, respectively) and the investigation of their mechanism of action. In the series we included calixarene 2 as reference compound whose activity in this biological context has been previously reported[66] and associated with its capacity to bind and neutralize LPS as topomimetic of LPS-binding peptides.

Since we hypothesized that calixarene derivatives could directly bind to human and murine MD-2 and CD14 in a similar fashion as LPS, we preliminarily performed docking calculations to support this mode of interaction. Moreover, we aimed here to verify if the TLR4 antagonist activity is a rather general property of positively charged amphiphilic calixarenes and if the antagonist effect also derives from the direct interaction of calixarenes with the receptors and not exclusively from LPS neutralizing action, as suggested for calixarene **2**.



1: R = -$CH_2CH_2CH_2$-NH-$C(NH_2)=NH_2^+$ Cl^-
2: R = -$CH_2CH_2CH_2CH_2$-NH-$C(NH_2)=NH_2^+Cl^-$

3: R = C_3H_7
4: R = C_6H_{13}
5: R = C_8H_{17}
6: R = -CH_2CH_2-O-CH_2CH_3

7: R = C_6H_{13}
8: R = C_8H_{17}
9: R = $C_{12}H_{23}$

Figure 3.17. Positively charged guanidinocalixarenes **1**-**6** and negatively charged carboxy calixarenes **7**-**9**.

Rational Design of Amphiphilic Calixarenes as CD14/ MD-2 Ligands.

We were inspired by the hypothesis that the calixarenes could be TLR4 modulators similar to lipid A variants and to trehalose or glucose-based glycoamphiphiles previously developed by one of the groups involved in the present study.[61] Positively charged guanidinocalix[4]arenes **1** and **3**-**6** and negatively charged carboxylate calixarenes **7**-**9** were designed in order to investigate the suitability of this macrocyclic scaffold to build CD14 and TLR4/MD-2 ligands (Figure 3.18). These calixarene derivatives have an amphiphilic character due to the presence of lipophilic tails on one rim and charged polar groups on the other. Only compound **6**, having ethoxyethyl chains at the lower rim, has a reduced amphiphilicity and was included in the library precisely to verify the possible relevance of this property in the biological activity.

Calixarenes **1** and **2**[66] present lipophilic upper rims bearing four *tert*-butyl groups and polar lower rims with positively charged guanidinium groups linked through, respectively, propyl and butyl chains. Calixarenes **3**-**6** present a reversed arrangement of lipophilic and charged groups: guanidinium groups are directly linked to the scaffold on the upper rim, and hydrocarbon chains of different length (C_3, C_6, and C_8 for compounds **3**, **4**, and **5**, respectively), or an ethoxy ethyl chain in the case of compound **6**, are linked at the lower rim. Finally, anionic calixarenes **7**-**9** were designed with the purpose of studying the influence of negatively charged groups. Thus, these anionic calixarenes present carboxylate groups at the upper rim, aiming to mimic the phosphate groups of LPS, and hydrocarbon chains of variable length (C_6, C_8, and C_{12}) at the lower rim.

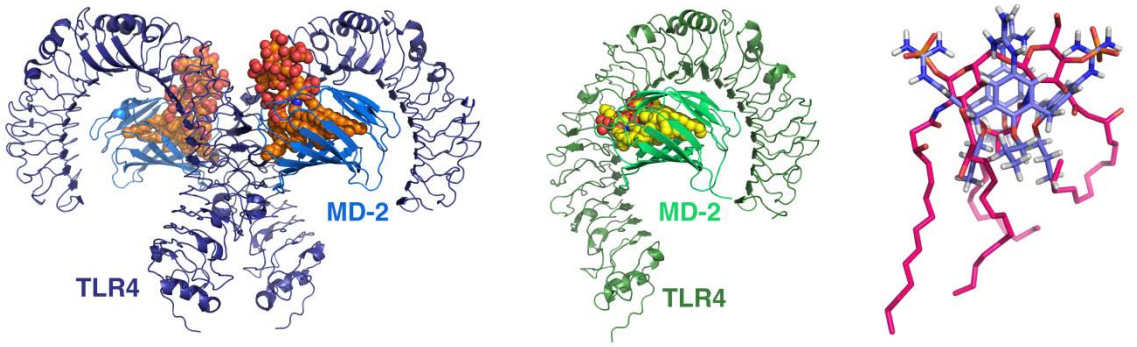


Figure 3.18. Left: 3D structure of human TLR4/MD-2/LPS dimer from PDB code 3FXI. Middle: 3D structure of TLR4/MD-2/lipid-IVa from PDB code 2E56. Right: Superimposition of lipid IVa (from PDB code 2E56, magenta) and calixarene **3** (purple).

Three-dimensional (3D) structures of compounds **1**-**9** were built and optimized by means of computational techniques (cf. *materials and methods* below). We superimposed the 3D structures of the calixarenes **2** and **3** with that of lipid IVa, a natural underacylated MD-2 ligand with activity as hTLR4 antagonist. From a comparison of lipid IVa (3D structure from the X-ray crystallography structure) with compound **3** (Figure 3.18-right), the oppositely charged groups (phosphate vs guanidinium) aligned perfectly, and also did the disaccharide over the aromatic calix backbones, and the acyl over the alkoxy chains. This preliminary result regarding the geometrical similarity prompted us to further study calixarenes **1**-**9** as putative TLR4/MD-2 and CD14 ligands.

First, compounds **2**, **3**, and **4**, as representative derivatives, were docked into the binding site of the human CD14 protein (PDB code 4GLP). For all these three compounds, docking calculations predicted favorable binding poses inside the human CD14 protein (Figure 3.19), where the guanidinium moieties are placed at the rim of CD14 and the hydrophobic chains are inserted into the hydrophobic pocket.

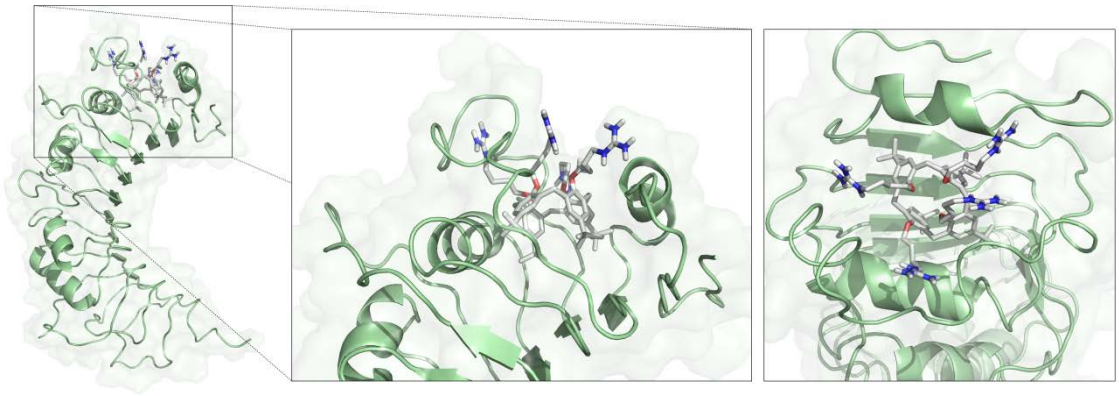


Figure 3.19. Docked pose for compound **3** inside CD14 (PDB code 4GLP). Left: full perspective. Middle: side view. Right: top view.

Docking calculations were also performed with compounds **1**-**9** into four different structures of the TLR4/MD-2 system: human and mouse, in agonist and antagonist conformations of MD-2 (cf. Figure 3.20, Figure 3.21, Figure 3.22, and Figure 3.23). Overall, all the ligands were predicted to bind inside the different TLR4/MD-2 structures, with the guanidinium/carboxylate moieties placed at the rim of MD-2, where polar interactions predominate, and the lipophilic groups (alkoxy or *tert*-butyl chains) inside the MD-2 pocket. These docked poses are in agreement with calculations reported by us of compounds binding both CD14 and MD-2 proteins. Although MD-2 is more specific in the ligand recognition, both MD-2 and CD14 binding pockets share some similarities regarding volume and accessible surface area.[33, 41]

Regarding reported compound **2**, in the docked poses in both agonist and antagonist conformations of human MD-2, the guanidinium groups establish H-bonds with the side chains of Glu92, Tyr102, and Ser118 and the backbone of Lys122 (Figure 3.20, Figure 3.22, and Figure 3.23 left), while one of the aromatic rings of the macrocycle is engaged in a π-π-stacking interaction with the side chain of Phe119.

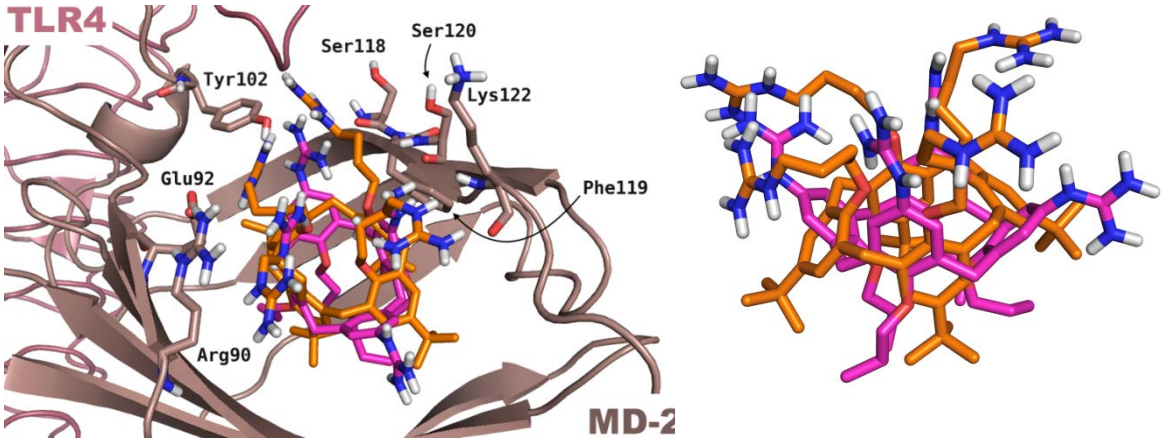


Figure 3.20. Superimposition of the best docked poses for compounds **2** (orange) and **3** (magenta) in TLR4/MD-2 heterodimer (PDB code 2Z65). A 90° rotated view is shown on the right (TLR4/MD-2 has been hidden for clarity).

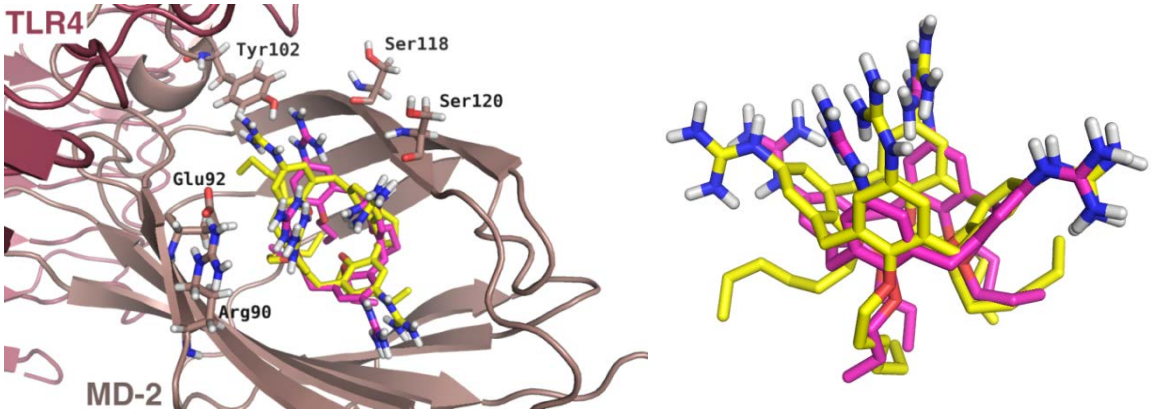


Figure 3.21. Superimposition of the best docked poses for compounds **3** (magenta) and **4** (yellow) in hTLR4/MD-2 heterodimer (PDB code 2Z65). A 90° rotated view is shown on the right (TLR4/MD-2 has been hidden for clarity).

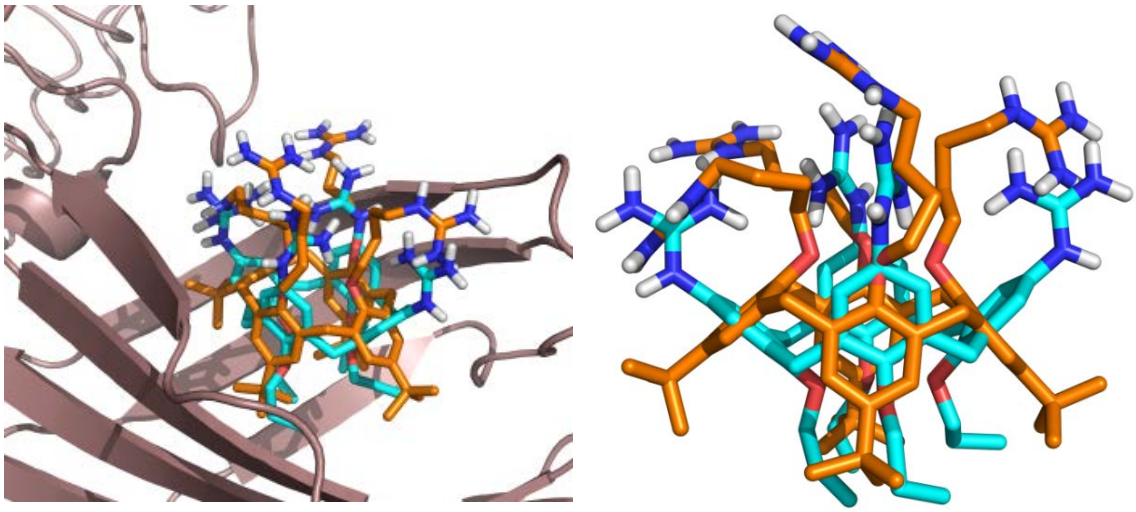


Figure 3.22. Left: superimposition of the best docked pose of compound **2** (orange) and compound **3** (cyan) into human TLR4/MD-2 system (PDB ID 3FXI). Right: details of the superimposed docked poses (the TLR4/MD-2 structure is not displayed).

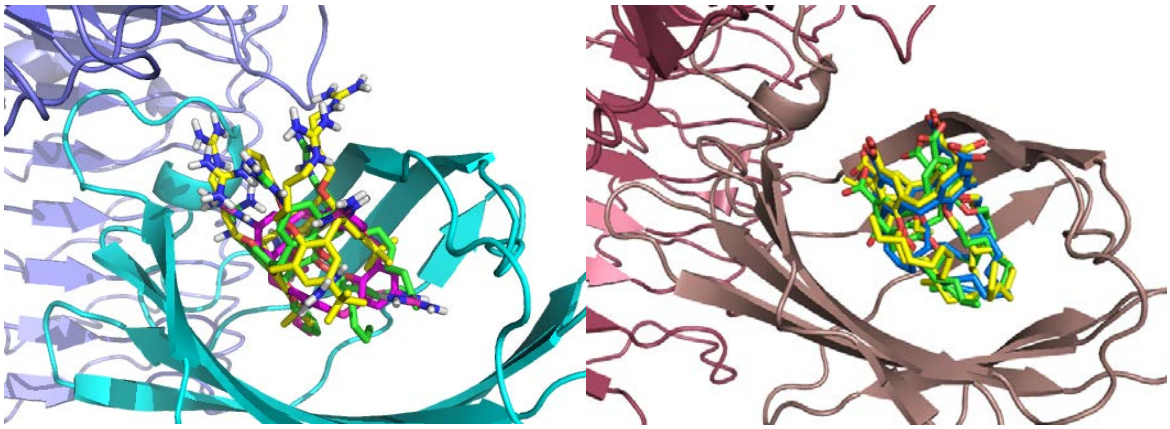


Figure 3.23. Superimposition of the best docked poses of compounds **2**, **3**, and **4**, in yellow, pink, and green, respectively, (on the left) and compounds **7**, **8**, and **9**, in green, blue, and yellow, respectively, (on the right) into mTLR4/MD-2 in the antagonist conformation (PDB ID 2Z65).

In detail, the guanidinium groups at the upper rim of compounds **3-5** establish H-bonds with the backbone of Ser120 and with the side chains of Glu92 and Tyr102 (Figure 3.21). The longer alkyl chains of compounds **4** and **5** occupy deeper regions of the MD-2 pocket. Interestingly, when comparing the best predicted docked poses for compounds **2** and **3**, it was observed that they are half-turn rotated one from another in regard to the calixarene moiety (Figure 3.20, and Figure 3.23-left). In both cases, the guanidinium moieties are accommodated at the entrance of the pocket while the hydrophobic groups (*tert*-butyl and propyl for compounds **2** and **3**, respectively) are buried inside the MD-2 hydrophobic pocket.

Regarding compounds **7-9**, they presented similar docked poses where the alkyl chains were also buried inside the hydrophobic MD-2 pocket and the carboxylate moieties were establishing polar interactions with the residues at the MD-2 rim. Compounds **8** and **9** presented docked poses protruding slightly more than compound **7**, probably due to the longer alkyl chains, although the difference was very subtle (Figure 3.23 right).

To ensure the stability of the docked poses of compound **3** with TLR4/MD-2 and to gain insights on the interactions that take place, we performed 90 ns molecular dynamic simulations of the hTLR4/MD-2/**3** complex starting from the docked geometries for both the antagonist and the agonist conformations of hTLR4/MD-2. In the simulation starting from the agonist conformation of MD-2 we could observe that compound **3** rotates of almost 90° around its plane of symmetry (a partial rotation happens at 5 ns of simulation and the full rotation at approximately 38 ns) to find a more stable bound

conformation that was maintained stable for the rest of the simulation (Figure 3.24 left). This rotation forced the MD-2 pocket to adopt an antagonist-like conformation (characterized by, inter alia, great motion of residue Phe126). In this new binding mode, two guanidinium groups of compound **3** continued to interact through hydrogen bonds with the side chains of Glu92 and Ser120, a third guanidinium group formed a new hydrogen bond with the CO group of Pro88, and the fourth guanidinium group was involved in polar interactions with the solvent. Moreover, later in the simulation (starting at 42 ns), the loop made by residues 80-90 undergoes a considerable deformation (Figure 3.24-right). In contrast, in the simulation of the TLR4/MD-2/**3** complex starting from the antagonist conformation, the geometries of both compound **3** and MD-2 were stable during the 90 ns run (Figure 3.24-right), not experiencing important conformational changes. These results clearly indicated that the complex of calixarene **3** with MD-2 in agonist conformation is less stable than the complex with the antagonist one, therefore providing explanations for the antagonist activity later observed (*3.4.2 Annex 2*). Taken together, our computational studies provided plausible binding poses for compounds **2-4** into CD14 and for compounds **1-9** into TLR4/MD-2, supporting a putative direct binding to these proteins.

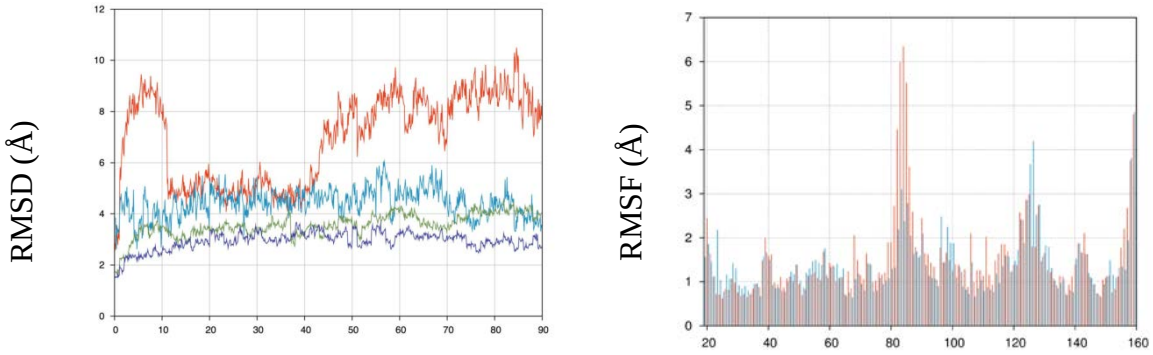


Figure 3.24. RMSD (left) and RMS fluctuation per residue (right) along the 90 ns MD simulation of the two TLR4/MD-2/**3** complexes starting from the agonist (green) and the antagonist (purple) conformation of MD-2. The RMSD (left) of compound **3** is also displayed: in complex with TLR4/MD-2 in the agonist (red) and the antagonist (cyan) conformation.

Experimental section.

Compounds **1-9** were synthesized through a procedure already reported[66] and other protocols thoroughly described in the paper of this study.[23] The results for all the

experiments performed on cells are reported in annex of the present chapter (*3.4.2 Annex 2*).

Discussion

Calixarenes with guanidinium groups on one rim and hydrocarbon chains on the other are facial amphiphiles in which the charged polar face and the hydrophobic apolar one are spatially organized. Because of this structural feature, they can be potential modulators of the TLR4 activation through direct binding to the receptor or one of the co-receptors involved in the signaling process. We designed the series of calixarenes **1**-**9** aiming to explore the plausible direct binding to CD14 and MD-2 co-receptors. Docking studies demonstrated that compounds **2**-**4** and **1**-**9** are in principle able to form complexes with CD14 and TLR4/MD-2 heterodimer (human and murine), respectively, independently from the relative disposition of the polar and apolar residues and from the nature of the charged groups. The lipophilic chains, linked at the upper (**1** and **2**) or at the lower (**3**-**5**) rim, were in all cases buried into the CD14 or MD-2 hydrophobic pocket, while the charged heads established contacts with polar residues located in proximity of the entry of the pockets. Therefore, our computational studies provide plausible binding poses into the TLR4 co-receptors for the investigated compounds, and this supports our hypothesis of a direct binding of calixarene derivatives to these proteins in competition with LPS. These findings thus open the possibility of exploring calixarenes as a platform for the design of TLR4/MD-2 modulators. Calculations of the stability of the complexes between the guanidinocalixarene **3** and TLR4/MD-2 suggested that this derivative, and for analogy at least all the other positively charged analogues, could act as antagonist.

The activity of positively charged calixarenes was tested on HEK cells expressing hTLR4/MD-2 and human and murine leukocytes. Cationic calixarenes **1**-**4** inhibited in a dosedependent way LPS-stimulated TLR4 activation in both human and murine cells. Cells were first stimulated by LPS and then treated with synthetic molecules. In agreement with the theoretical studies, compounds **1**-**4** showed an antagonist activity in the low micromolar range on human and murine TLR4. Paradoxically, negatively charged amphiphilic calixarenes, which should mimic better the anionic nature of lipid A, the natural MD-2 and CD14 ligand, were not active both in inhibiting and stimulating TLR4. Compound **6**, with more polar ether chains on the lower rim instead

of hydrocarbon chains turned out to be substantially inactive as TLR4 inhibitor in both cell types.

In addition, calculations of the log P predicted high values for compounds **7**, **8**, and **9**, with calculated log P values above 15, pointing to a high lipophilicity, while compound **6** was predicted to be extremely hydrophilic having calculated log P value below zero (Figure 3.25). These unfavorable log P values could be correlated with poor physical-chemical properties, thus explaining the lack of activity in the cells assays.

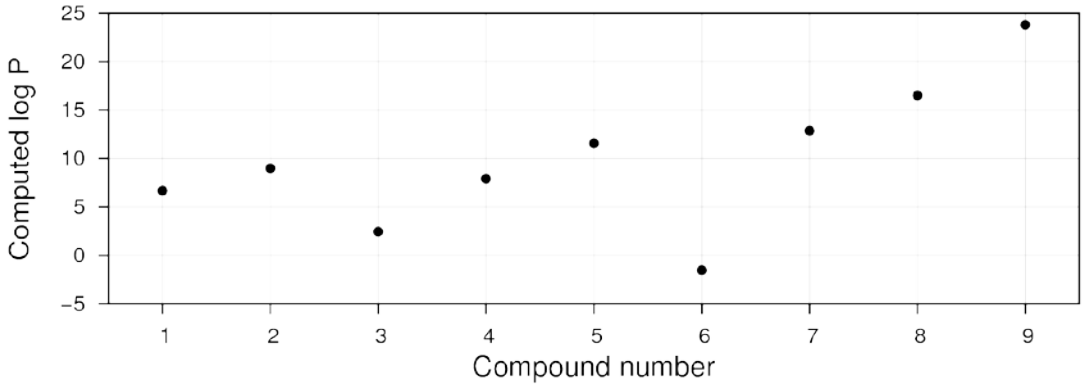


Figure 3.25. Calculated log P for compounds **1** to **9**.

The very close IC_{50} values (Table A3.1) found for guanidinocalixarenes **2** and **3** indicate that the relative disposition of polar and hydrophobic residues with respect to the macrocyclic cavity does not have a significant impact on the inhibition activity of these ligands. Furthermore, by comparing compounds **3**-**5**, it seems that an increasing lipophilicity is detrimental for the inhibition potency, even if the less amphiphilic derivative **6** is very poorly active with an IC_{50} two orders of magnitude higher than that of **3**. A subtle balance between lipophilic and hydrophilic portions in the ligand structure seems then to be the key to obtain activity.

It is worth noting that lead compounds **3** and **4** show antagonist activity on both human and murine TLR4. Several TLR4 modulators resembling lipid A have species-specific activity that is generally attributed, among other factors, to the dissimilarities in the shape of the hydrophobic binding pocket of human and mouse MD-2[67] and to the variations in the electrostatic potentials at the rim of the binding cavity of MD-2 and at the dimerization interface. The most significant example of this is the natural compound tetraacylated lipid IVa that acts as an antagonist on human but as an agonist on murine TLR4.[67] However, several synthetic phosphoglycolipids with a monosaccharide

scaffold also showed agonist activity on murine and antagonist activity on human TLR4.[34]

To understand whether the antagonist activity of calixarenes was due to their interaction with receptors or with LPSs we also undertook studies of the TLR4 activation with non-LPS ligands. We reasoned that if the contribution of calixarenes in inhibiting TLR4 activation is due to a neutralizing effect on endotoxin, the antagonist effect would be lost by stimulating cells with a TLR4 agonist structurally different from LPS. Besides the natural agonists LPS, lipid A, lipid A mimetics as monophosphoryl lipid A (MPL),[68] and aminoalkyl glucosaminide 4-phosphates,[54] TLR4 can also be activated by small molecules, such as the natural compound paclitaxel,[69] oxidized phospholipids, and synthetic pyrimidoindoles and neoseptins,[70-71] and by protein DAMPs such as high mobility group box 1 (HMGB1)[72] and lectins. Lectins constitute a very large class of carbohydrate-binding proteins, and plant lectins have immunostimulating activity that recently has been related to TLR agonism. In fact, the activity as potent TLR4 agonists of plant lectins KML-C (Korean mistletoe lectin)[73] and PHA (phytohemagglutinin from Phaseolus vulgaris)[74] has been described. Although the experimental data indicate a strong TLR4 agonist activity by lectins, the mechanism of action of these proteins is yet to be clarified.

Because lectins recognize and bind sugars, it is possible that lectins promote the formation of the $(TLR4/MD\text{-}2/LPS)_2$ heterodimer by binding to the sugars attached to the surface of glycosylated MD-2 and TLR4 proteins, thus bringing together two TLR4/MD-2 complexes.

Based on these data from the literature, we first validated plant PHA lectin as agonist in HEK-Blue cells. A dose-dependent activation of TLR4 signal was observed when cells were treated with PHA lectin in the presence of polymixin-B to neutralize the agonist effect of any possible LPS contamination. The addition of calixarenes **1**-**4** followed by lectins inhibited in a dose-dependent way the TLR4 signal, showing that cationic calixarenes antagonize TLR4 signal also in the case of non-LPS stimulation. This suggests a direct interaction of calixarenes with CD14 and MD-2 receptors, in agreement with predicted binding poses by docking calculations and MD simulations.

Calixarenes **3** and **4** showed a potent TLR4 antagonist activity in cells and inhibited the production of the inflammatory TNF-α in LPS-stimulated murine splenocytes and in

murine animal models. Although solubility and distribution properties of calixarenes **3** and **4** should be optimized for in vivo studies and preclinical development, the lack of toxicity (cf. *Supporting Info* of the paper[23]) and the potent TLR4 blocking activity point to these compounds as plausible drug hits targeting TLR4. The flexibility of calixarene scaffold will allow modulation of the hydrophilicity profile of cationic amphiphiles and optimization of their pharmacokinetics. The possibility of the calix cavity to complex metal ions or small organic fluorophores could also be exploited to generate labeled compounds for diagnostic and therapeutic applications.

Conclusion

We here report a series of amphiphilic guanidinocalixarenes whose structure were computationally optimized to dock into MD-2 and CD14 binding sites. Some of these calixarenes were active in inhibiting, in a dose-dependent way, the LPS-stimulated TLR4 activation and TLR4-dependent cytokine production in human and mouse cells. Moreover, guanidinocalixarenes also inhibited TLR4 signaling when TLR4 was activated by a non-LPS stimulus, the plant lectin PHA. While the activity of guanidinocalixarenes in inhibiting LPS toxic action has previously been related to their capacity to bind LPS, we suggest a direct antagonist effect of calixarenes on TLR4/MD-2 dimerization, supported by molecular modeling calculations and biological assays. These results point at the calixarene moiety as a potential scaffold for the development of new TLR4-directed therapeutics.

Materials and methods

Structure Construction. 3D structures of the ligands were built with PyMOL molecular graphics and modeling package[24] based on the coordinates of the calixarene scaffold retrieved from the PubChem database (CID: 562409). 3D coordinates for the agonist hTLR4/MD-2 complex, the antagonist mTLR4/MD-2 complex, the agonist mTLR4/MD-2 complex, and hCD14 were retrieved from the PDB database (www.rcsb.org) under codes 3FXI, 2Z64, 3VQ2, and 4GLP, respectively. The structures went through a restrained minimization procedure with Maestro using the OPLS3 force field. Gasteiger charges were computed within the AutoDock Tools program, and all nonpolar hydrogens were merged.

Structure Optimization. All compounds (from **1** to **9**) were optimized with *ab initio* calculations, using the density functional theory (DFT) with the hybrid functional

B3LYP with the Pople basis set 6-31+g(d,p) using Gaussian g09/e1.[43] Water solvation (with a dielectric constant of ε = 78.3553) was simulated with the Gaussian default SCRF method (i.e., using the polarizable continuum model (PCM) with the integral equation formalism variant (IEFPCM)).

Docking Procedure. Docking was performed independently with both AutoDock 4.2 and AutoDock Vina 1.1.2.[26, 44] In AutoDock 4.2, the Lamarckian evolutionary algorithm was chosen and all parameters were kept default except for the number of genetic algorithm runs which was set to 200 to enhance the sampling. AutoDockTools 1.5.6 was used to assign the Gasteiger-Marsili empirical atomic partial charges to the atoms of both the ligands and the receptors. The structure of the receptors was always kept rigid, whereas the structure of the ligand was set partially flexible by providing freedom to some appropriately selected dihedral angles. Concerning the boxes, spacing was set to 0.375 Å for AutoDock and is default to 1 Å for Vina. In the case of the human and mouse TLR4/MD-2 systems in their agonist and antagonist conformations, the size of the box was set to 33.00 Å in the x-axis, 40.50 Å in the yaxis, and 35.25 Å in the z-axis. For hCD14 the size of the box was set to 33.00 Å in the x-axis, 33.75 Å in the y-axis, and 33.75 Å in the z-axis. For the hTLR4/MD-2 complex the center of the box is located equidistant to the center of mass of residues Arg90 (MD-2), Lys122 (MD-2), and Arg264 (TLR4). For the mTLR4/MD-2 complex the center of the box is located equidistant to the center of mass of residues Arg90 (MD-2), Glu122 (MD-2), and Lys263 (TLR4). For hCD14 the center of the box is located equidistant to the center of mass of residues Phe69, Tyr82, and Leu89.

Parameters Derivation. Parameters for molecular dynamics simulations were set up with the standard Antechamber[28] procedure. Briefly, charges were calculated with Gaussian at the Hartree-Fock level (HF/6-31G* Pop=MK iop(6/33=2) iop(6/42=6)) from the solvated DFT B3LYP optimized structure, then derived and formatted for Ambertools15 and AMBER14[29] with Antechamber assigning the general AMBER force field (GAFF) atom types.[31] A new atom type for nitrogen was introduced (nj), within GAFF, to properly describe the guanidine moiety, mirroring the parameters of ff14SB[30] used to describe the guanidine fragment present in arginine. Parameters for this new atom are provided in annex (*3.4.2 Annex 2*).

Molecular Dynamics (MD) Simulations. All production runs were performed for 90 ns and followed the same protocol described in section *3.1.1 Bradyrhizobium LPS*.

Log P calculation. From the optimized 3D structure of compounds **1-9**, log P value was calculated with the Maestro package.[45]

Materials and methods regarding the synthesis and the cell tests is reported in the related paper and its *Supporting Info* document.[23]

3.4.2 Annex 2

Experimental results.

Inhibition of LPS-Stimulated TLR4 Signal in HEK-Blue Cells. Calixarenes **1-9** were first screened for their capacity to interfere with LPS-stimulated TLR4 activation and signaling on HEK-Blue cells. HEK-Blue cells are stably transfected with TLR4, MD-2, and CD14 genes. In addition, these cells stably express an optimized alkaline phosphatase gene engineered to be secreted (sAP), placed under the control of a promoter inducible by several transcription factors such as NF-κB and AP-1.[75] This reporter gene allows monitoring of the activation of TLR4 signal pathway by endotoxin. All calixarenes were inactive in stimulating TLR4 signal when provided alone, thus indicating the absence of agonist activity, in agreement with the molecular modeling studies. On the other hand, compounds **1-5** inhibited in a dose-dependent way the LPS-stimulated TLR4 signal (Figure A3.9), while calixarene **6** with oxygenated ethylene glycol chains instead of hydrocarbon chains showed weak antagonistic activity.

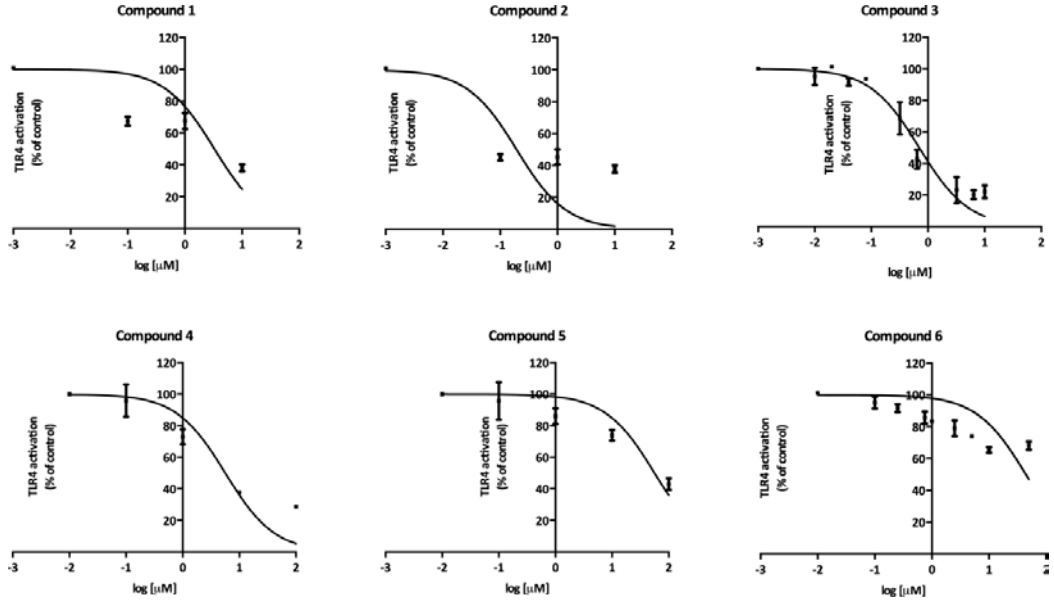


Figure A3.9. Dose dependent inhibition of LPS-stimulated HEK-Blue cells activation by calixarenes **1**-**5**. Human TLR4 HEK-Blue was treated with increasing concentrations of compounds and stimulated with LPS (100 ng/mL). The results represent normalized data with positive control (LPS alone) and expressed as the mean of percentage ± SD of at least three independent experiments.

Guanidinocalixarenes **1**-**5** inhibited TLR4 signal with potencies ranging from 0.2 to 63 μM (Table A3.1). Compounds **2**, **3**, and **4** were the most potent antagonists and inhibited LPS-stimulated TLR4 signal with IC50 of 0.2, 0.7, and 5.7 μM, respectively. In contrast, negatively charged amphiphilic calixarenes **7**-**9** with carboxylic groups on the upper rim showed no or very weak inhibition of LPS-TLR4 signal (data shown in the *Supporting Info* of the paper[23]).

Table A3.1. IC_{50} Values for the Inhibition of LPS-Stimulated TLR4 Signal in HEK Cells

Compound	IC_{50} LPS (μM)
1	10
2	0.2
3	0.7
4	5.7
5	63
6	45

Inhibition of PHA Lectin-Stimulated TLR4 Signal in HEK-Blue Cells. We were then interested in knowing if the inhibition of TLR4 signal is due to calixarene interaction with LPS or to a direct interaction with the TLR4 receptor system, evidenced as possible by calculations. To investigate this point, we stimulated HEK cells with the

plant lectin phytohemagglutinin (PHA from *Phaseolus vulgaris*) whose property to potently stimulate TLR4 signal acting as agonist has been recently described.[74] We first checked if PHA is able to activate TLR4 signal in HEK-Blue cells, and we found that the lectin was active in stimulating in a dose-dependent way TLR4-dependent SEAP production (cf. *Supporting Info* of the paper[23]). To exclude the TLR4 activity that could be derived from LPS contamination in the PHA, we performed the experiment in the presence of the LPS-neutralizing peptide polymixin-B. We also verified that control HEK-null cells, that is, HEK cells transfected with SEAP plasmid and lacking TLR4, MD-2, CD14 genes, were not activated by PHA lectin (cf. Supporting Info of the paper[23]). PHA lectin was then used instead of LPS as a TLR4 agonist to stimulate cells. The highly potent calixarene-based TLR4 antagonists, compounds **3** and **4**, were then investigated for their property to inhibit TLR4 activation by PHA lectin (Figure A3.10).

Guanidinocalixarenes **3** and **4** were indeed active in inhibiting PHA lectin-stimulated TLR4 signal in a concentration-dependent way, with potencies similar to those measured in the inhibition of LPS-stimulated TLR4 signal (Table A3.1). The fact that the antagonist activity was retained by calixarenes also when TLR4 was stimulated by a non-LPS agonist strongly suggests that the action of calixarenes is mainly based on direct interaction with CD14 and MD-2 receptors.

Inhibition of LPS-Stimulated TLR4 Signal in Human White Blood Cells. As HEK cells are a non-natural system to study TLR4 activation and to perform preliminary screening, the capacity of lead compounds **3** and **4** to inhibit LPS-stimulated TLR4 signaling was further investigated in human white blood cells hWBCs that naturally express CD14, MD-2, and TLR4 receptors. We evaluated the production of the main NF-κB-dependent proinflammatory cytokines tumor necrosis factor α (TNF-α), interleukin-6 (IL-6), and IL-8 by primary human peripheral blood mononuclear cells (hPBMCs) as readout for TLR4 pathway activation. hPBMCs isolated from the whole blood of healthy volunteers were treated with increasing concentrations (1-10 μM) of compounds **3** and **4** and stimulated after 30 min with LPS (100 ng/mL). Compound **3** reduced the production of all the proinflammatory cytokines monitored, while compound **4** showed a lower inhibitory activity, reducing only two of the three cytokines evaluated (Figure A3.11).

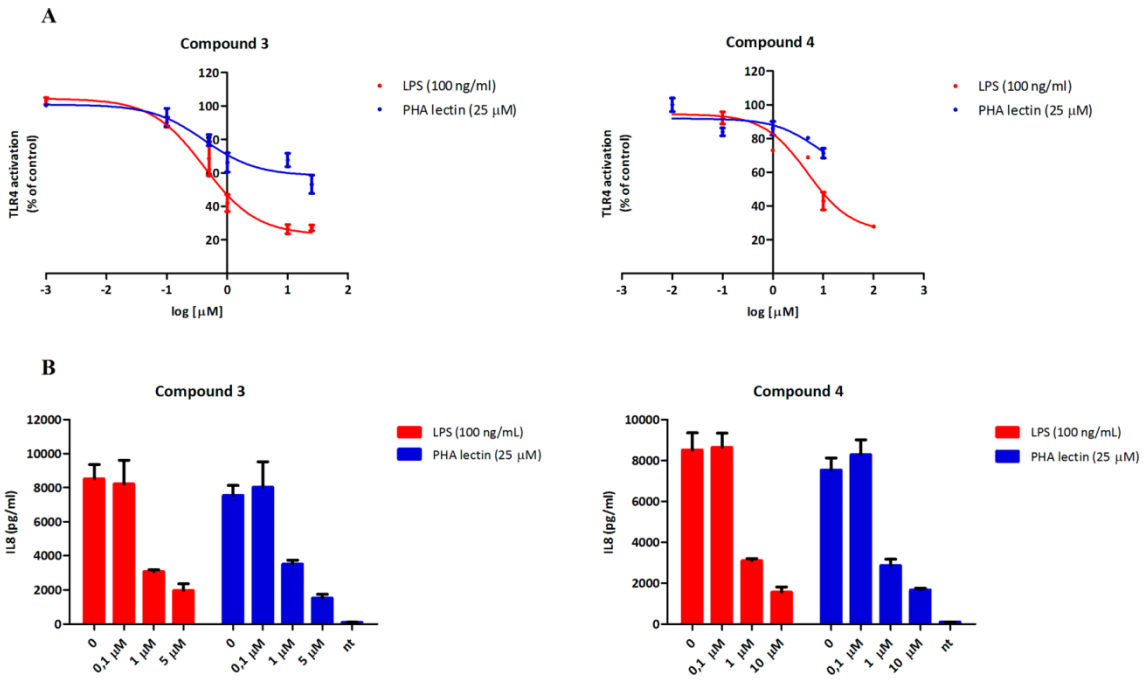


Figure A3.10. (**A**) Inhibition of TLR4 signaling in HEK-Blue cells stimulated with LPS (100 ng/mL) or PHA lectin (25 μM) and treated with calixarenes **3** and **4**. The results represent normalized data with positive control (LPS or PHA lectin alone). (**B**) Quantification of interleukin-8 (IL-8) in HEKBlue cells stimulated with LPS or PHA and treated with compounds **3** and **4** by performing ELISA assay. Data represent the mean of percentage ± SD of at least three independent experiments.

Inhibition of LPS-Stimulated TLR4 Signal in Murine White Blood Cells (mWBCs). It is known that human and murine MD-2 have dissimilarities in the LPS binding region, and some ligands have different activity on hMD-2 and mMD-2, in some cases switching from agonism to antagonism. We therefore aimed to compare the activity of calixarene on human and murine cells. The activity of compounds **3** and **4** was then evaluated in a murine macrophages cell line, RAW-Blue cells. As HEK-Blue cells, RAW-Blue cells are transfected to stably express the SEAP reporter gene in order to monitor the activation of TLR4 signal pathway. Compounds **3** and **4** inhibited in a dose-dependent way the LPS-stimulated TLR4 signal (Figure A3.12A), revealing that the two calixarenes were also effective on the murine TLR4 system.

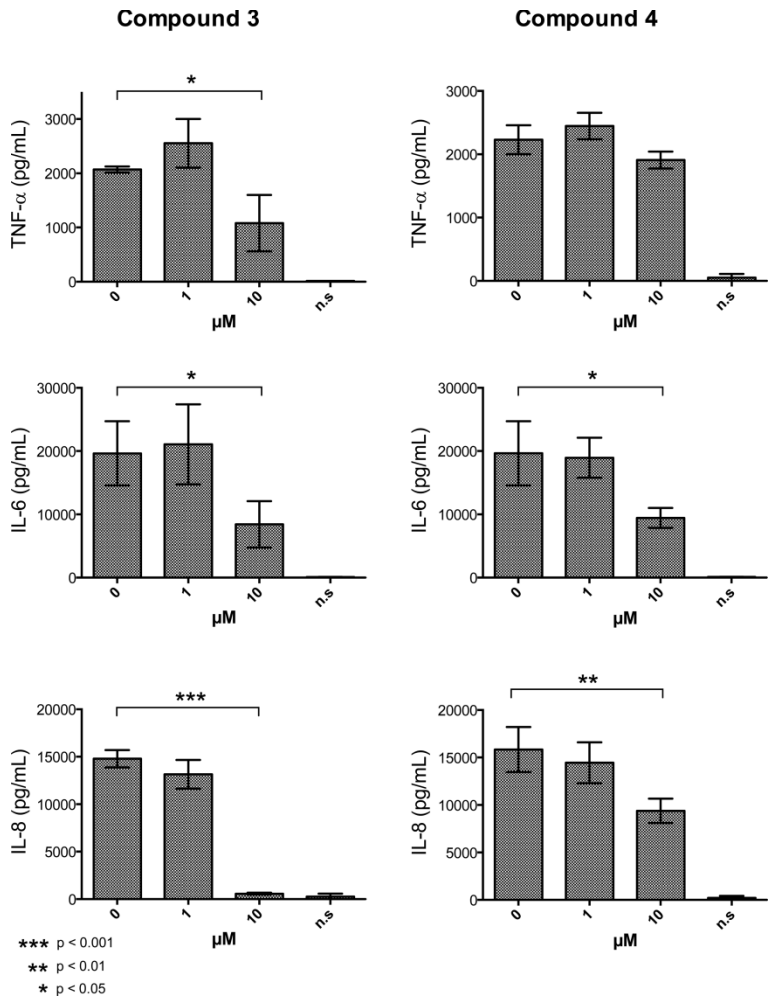


Figure A3.11. Inhibitory effect of compounds **3** and **4** on LPS-induced proinflammatory cytokines production by PBMCs. PBMCs isolated from whole blood were preincubated with synthetic compounds for 30 min and then stimulated with LPS (100 ng/mL). TNFα, IL-6, and IL-8 production was quantified after one night's incubation. Data represent the mean ± SEM of at least three independent experiments.

The abilities of compounds **3** and **4** were further investigated in murine splenocytes. TNF-α relative expression was determined from TLR4-MyD88 pathway activation. Splenocytes from balb/c mice were treated with two concentrations (1 and 10 μM) of compounds **3** and **4** in RPMI and stimulated after 30 min with LPS (100 ng/mL). The LPS-induced TNF-α expression after a 5 h incubation was measured by qPCR. The lower concentration of compounds **3** and **4** (1 μM) was weakly active in reducing LPS-induced TNF-α expression, whereas the higher concentration (10 μM) of both compounds completely inhibited the expression of TNF-α (Figure A3.12B).

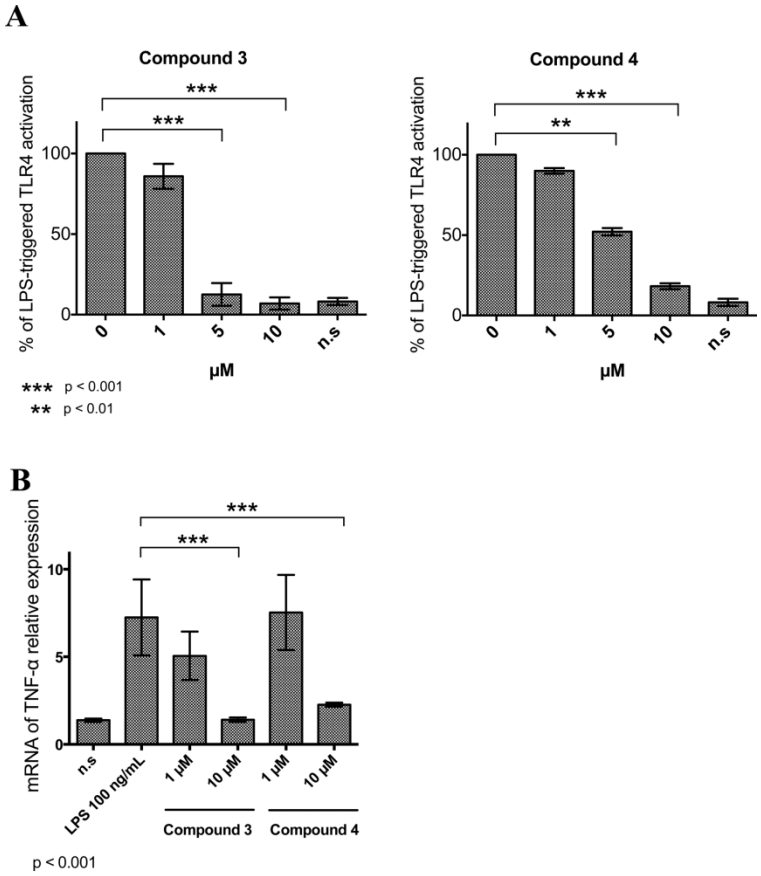


Figure A3.12. Effects of compounds **3** and **4** on RAW-Blue cells and on murine splenocytes. (**A**) RAW-Blue cells stably transfected with NF-κB-dependent SEAP reporter plasmid were treated with increasing concentrations of compounds **3** and **4** and stimulated with LPS (100 ng/mL) after 30 min. Data represent the mean of percentage of at least three independent experiments. (**B**) Murine splenocytes isolated from murine spleen were preincubated with two concentrations (1 and 10 μM) of compounds **3** and **4** for 30 min and then stimulated with LPS (100 ng/mL). Readout was the TNF-α expression after 5 h of incubation. Normalized data are representative of three independent experiments.

Parameters for nitrogen atom type "nj".

LogFile:
addAtomTypes {
{ "nj" "N" "sp2" }
}

FRCMOD:
MASS
nj 14.01 0.530 sp2 N in amino groups (from ff14SB N2)
BOND
ca-nj 481.0 1.340 JCC,7,(1986),230; ARG,CYT,GUA (from parm10 CA-N2)
hn-nj 434.0 1.010 JCC,7,(1986),230; ADE,CYT,GUA,ARG (from parm10 H -N2)
ANGLE
ca-nj-hn 50.0 120.00 (from parm10 CA-N2-H)
nj-ca-nj 70.0 120.00 AA arg (from parm10 N2-CA-N2)

ca-nj-ca 50.0 123.20 AA arg (from parm10 CA-N2-CT)
hn-nj-hn 35.0 120.00 (from parm10 H -N2-H)
ca-ca-nj 70.0 120.00 (from parm10 CA-C -OH)
DIHE
hn-nj-ca-nj 1 0.000 0.0 -4. (H -N2-CA-N2 from ff14SB)
hn-nj-ca-nj 1 2.400 180.0 2. ---
ca-nj-ca-nj 1 0.000 0.0 -4. Arg Lys copied (C8-N2-CA-N2 from ff14SB)
ca-nj-ca-nj 1 2.400 180.0 2. ---
ca-ca-nj-ca 1 0.065 0.0 -4. (CA-C -OH-HO from ff14SB)
ca-ca-nj-ca 1 0.883 180.0 2. (CA-C -OH-HO from ff14SB)
ca-ca-ca-nj 1 1.1 180.0 2. (CA-CA-C -OH from ff14SB)
ca-ca-nj-hn 1 0.065 0.0 -4. (CA-C -OH-HO from ff14SB)
ca-ca-nj-hn 1 0.883 180.0 2. (CA-C -OH-HO from ff14SB)

Bibliography

1. Oldroyd, G. E.; Murray, J. D.; Poole, P. S.; Downie, J. A. The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* **2011,** *45*, 119-144.
2. Masson-Boivin, C.; Giraud, E.; Perret, X.; Batut, J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol.* **2009,** *17* (10), 458-466.
3. Kuykendall, L.; Saxena, B.; Devine, T.; Udell, S. Genetic diversity in Bradyrhizobium japonicum Jordan 1982 and a proposal for Bradyrhizobium elkanii sp. nov. *Can. J. Microbiol.* **1992,** *38* (6), 501-505.
4. Zhang, Y. F.; Wang, E. T.; Tian, C. F.; Wang, F. Q.; Han, L. L.; Chen, W. F.; Chen, W. X. Bradyrhizobium elkanii, Bradyrhizobium yuanmingense and Bradyrhizobium japonicum are the main rhizobia associated with Vigna unguiculata and Vigna radiata in the subtropical region of China. *FEMS Microbiol. Lett.* **2008,** *285* (2), 146-154.
5. Lerouge, I.; Vanderleyden, J. O-antigen structural variation: mechanisms and possible roles in animal/plant–microbe interactions. *FEMS Microbiol. Rev.* **2002,** *26* (1), 17-47.
6. Carlson, R. W.; Forsberg, L. S.; Kannenberg, E. L. Lipopolysaccharides in Rhizobium-legume symbioses. In *Endotoxins: structure, function and recognition*, Springer: 2010; pp 339-386.
7. Busset, N.; De Felice, A.; Chaintreuil, C.; Gully, D.; Fardoux, J.; Romdhane, S.; Molinaro, A.; Silipo, A.; Giraud, E. The LPS O-Antigen in photosynthetic Bradyrhizobium Strains is dispensable for the establishment of a successful symbiosis with Aeschynomene legumes. *PLoS One.* **2016,** *11* (2), e0148884.
8. Silipo, A.; Vitiello, G.; Gully, D.; Sturiale, L.; Chaintreuil, C.; Fardoux, J.; Gargani, D.; Lee, H.-I.; Kulkarni, G.; Busset, N. Covalently linked hopanoid-lipid A improves outer-membrane resistance of a Bradyrhizobium symbiont of legumes. *Nat. Commun.* **2014,** *5*, 5106.
9. Komaniecka, I.; Choma, A.; Mazur, A.; Duda, K. A.; Lindner, B.; Schwudke, D.; Holst, O. Occurrence of an unusual hopanoid-containing lipid A among lipopolysaccharides from Bradyrhizobium species. *J. Biol. Chem.* **2014,** *289* (51), 35644-35655.
10. Di Lorenzo, F.; Palmigiano, A.; Duda, K. A.; Pallach, M.; Busset, N.; Sturiale, L.; Giraud, E.; Garozzo, D.; Molinaro, A.; Silipo, A. Structure of the Lipopolysaccharide from the Bradyrhizobium sp. ORS285 rfaL Mutant Strain. *ChemistryOpen.* **2017,** *6* (4), 541-553.
11. Kannenberg, E. L.; Carlson, R. W. Lipid A and O-chain modifications cause Rhizobium lipopolysaccharides to become hydrophobic during bacteroid development. *Mol. Microbiol.* **2001,** *39* (2), 379-392.
12. Ferguson, G. P.; Datta, A.; Baumgartner, J.; Roop, R. M.; Carlson, R. W.; Walker, G. C. Similarity to peroxisomal-membrane protein family reveals that Sinorhizobium and Brucella BacA affect lipid-A fatty acids. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (14), 5012-5017.
13. Brown, D. B.; Huang, Y.-C.; Kannenberg, E. L.; Sherrier, D. J.; Carlson, R. W. An acpXL mutant in Rhizobium leguminosarum bv. phaseolus lacks 27-hydroxyoctacosanoic acid in its lipid A and is developmentally delayed during symbiotic infection of the determinate nodulating host plant Phaseolus vulgaris. *J. Bacteriol.* **2011**, JB. 00392-11.
14. Raetz, C. R.; Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **2002,** *71* (1), 635-700.

15. Netea, M. G.; van Deuren, M.; Kullberg, B. J.; Cavaillon, J.-M.; Van der Meer, J. W. Does the shape of lipid A determine the interaction of LPS with Toll-like receptors? *Trends Immunol.* **2002,** *23* (3), 135-139.
16. Di Lorenzo, F.; Kubik, Ł.; Oblak, A.; Lorè, N. I.; Cigana, C.; Lanzetta, R.; Parrilli, M.; Hamad, M. A.; De Soyza, A.; Silipo, A. Activation of human Toll-like receptor 4 (TLR4)· myeloid differentiation factor 2 (MD-2) by hypoacylated lipopolysaccharide from a clinical isolate of burkholderia cenocepacia. *J. Biol. Chem.* **2015,** *290* (35), 21305-21319.
17. Noel, K.; Duelli, D. Rhizobium lipopolysaccharide and its role in symbiosis. *Prokaryotic nitrogen fixation: a model system for the analysis of a biological process.* **2000**, 415-431.
18. Vandenplas, M. L.; Carlson, R. W.; Jeyaretnam, B. S.; McNeill, B.; Barton, M. H.; Norton, N.; Murray, T. F.; Moore, J. N. Rhizobium sin-1 lipopolysaccharide (LPS) prevents enteric LPS-induced cytokine production. *J. Biol. Chem.* **2002,** *277* (44), 41811-41816.
19. Urbanik-Sypniewska, T.; Choma, A.; Kutkowska, J.; Kaminska, T.; Kandefer-Szerszen, M.; Russa, R.; Dolecka, J. Cytokine inducing activities of rhizobial and mesorhizobial lipopolysaccharides of different lethal toxicity. *Immunobiology.* **2000,** *202* (4), 408-420.
20. Komaniecka, I.; Zdzisinska, B.; Kandefer-Szerszen, M.; Choma, A. Low endotoxic activity of lipopolysaccharides isolated from Bradyrhizobium, Mesorhizobium, and Azospirillum strains. *Microbiol. Immunol.* **2010,** *54* (12), 717-725.
21. Tsukushi, Y.; Kido, N.; Saeki, K.; Sugiyama, T.; Koide, N.; Mori, I.; Yoshida, T.; Yokochi, T. Characteristic biological activities of lipopolysaccharides from Sinorhizobium and Mesorhizobium. *J. Endotoxin Res.* **2004,** *10* (1), 25-31.
22. Lorenzo, F. D.; Pallach, M.; Billod, J.-M.; Francisco, S.; Fresno, M.; Holgado, A.; Beyaert, R.; Martín-Santamaría, S.; Bernardini, M.-L.; Silipo, A. Bradyrhizobium lipopolysaccharide lipid A: antagonistic properties and molecular basis of its binding to the MD-2/TLR4 complex. *Manuscript in preparation.*
23. Sestito, S. E.; Facchini, F. A.; Morbioli, I.; Billod, J.-M.; Martin-Santamaria, S.; Casnati, A.; Sansone, F.; Peri, F. Amphiphilic guanidinocalixarenes inhibit lipopolysaccharide (LPS)-and lectin-stimulated toll-like receptor 4 (TLR4) signaling. *J. Med. Chem.* **2017,** *60* (12), 4882-4892.
24. L. DeLano, W. PyMOL(TM) Molecular Graphics System, Version 1.6.0.0. Schrödinger, LLC.
25. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory Comput.* **2015,** *12* (1), 281-296.
26. Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
27. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-655.
28. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006,** *25* (2), 247-260.
29. Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Gohlke, H.; Goetz, A.; Gusarov, S.; Homeyer, N.; Janowski, P.;

Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. AMBER 14, University of California, San Francisco. **2014**.
30. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
31. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-1174.
32. Facchini, F. A.; Zaffaroni, L.; Minotti, A.; Rapisarda, S.; Calabrese, V.; Forcella, M.; Fusi, P.; Airoldi, C.; Ciaramelli, C.; Billod, J.-M.; Schromm, A.; Braun, H.; Palmer, C.; Beyaert, R.; Jerala, R.; Pirianov, G.; Martin-Santamaria, S.; Peri, F. Structure-activity relationship (SAR) in monosaccharide-based Toll-like receptor 4 (TLR4) antagonist. *J. Med. Chem.* **2018,** *in press*.
33. Cighetti, R.; Ciaramelli, C.; Sestito, S. E.; Zanoni, I.; Kubik, Ł.; Ardá-Freire, A.; Calabrese, V.; Granucci, F.; Jerala, R.; Martín-Santamaría, S. Modulation of CD14 and TLR4· MD-2 Activities by a Synthetic Lipid A Mimetic. *Chembiochem.* **2014,** *15* (2), 250-258.
34. Matsuura, M.; Kiso, M.; Hasegawa, A. Activity of monosaccharide lipid A analogues in human monocytic cells as agonists or antagonists of bacterial lipopolysaccharide. *Infect. Immun.* **1999,** *67* (12), 6286-6292.
35. Tamai, R.; Asai, Y.; Hashimoto, M.; Fukase, K.; Kusumoto, S.; Ishida, H.; Kiso, M.; Ogawa, T. Cell activation by monosaccharide lipid A analogues utilizing Toll-like receptor 4. *Immunology.* **2003,** *110* (1), 66-72.
36. Perrin-Cocon, L.; Aublin-Gex, A.; Sestito, S. E.; Shirey, K. A.; Patel, M. C.; André, P.; Blanco, J. C.; Vogel, S. N.; Peri, F.; Lotteau, V. TLR4 antagonist FP7 inhibits LPS-induced cytokine production and glycolytic reprogramming in dendritic cells, and protects mice from lethal influenza infection. *Sci. Rep.* **2017,** *7*, 40791.
37. Funatogawa, K.; Matsuura, M.; Nakano, M.; Kiso, M.; Hasegawa, A. Relationship of structure and biological activity of monosaccharide lipid A analogues to induction of nitric oxide production by murine macrophage RAW264. 7 cells. *Infect. Immun.* **1998,** *66* (12), 5792-5798.
38. Mueller, M.; Lindner, B.; Kusumoto, S.; Fukase, K.; Schromm, A. B.; Seydel, U. Aggregates are the biologically active units of endotoxin. *J. Biol. Chem.* **2004,** *279* (25), 26307-26313.
39. Gutsmann, T.; Schromm, A. B.; Brandenburg, K. The physicochemistry of endotoxins in relation to bioactivity. *Int. J. Med. Microbiol.* **2007,** *297* (5), 341-352.
40. Gioannini, T. L.; Teghanemt, A.; Zhang, D.; Coussens, N. P.; Dockstader, W.; Ramaswamy, S.; Weiss, J. P. Isolation of an endotoxin–MD-2 complex that produces Toll-like receptor 4-dependent cell activation at picomolar concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (12), 4186-4191.
41. Ciaramelli, C.; Calabrese, V.; Sestito, S. E.; Pérez-Regidor, L.; Klett, J.; Oblak, A.; Jerala, R.; Piazza, M.; Martín-Santamaría, S.; Peri, F. Glycolipid-based TLR4 Modulators and Fluorescent Probes: Rational Design, Synthesis, and Biological Properties. *Chem. Biol. Drug Des.* **2016,** *88* (2), 217-229.
42. http://glycam.org (accessed February 2018).
43. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. Gaussian 09. Gaussian, Inc., Wallingford CT: 2009.

44. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
45. https://schrodinger.com/maestro (accessed February 2018).
46. Andreu, D.; Rivas, L. Animal antimicrobial peptides: an overview. *Peptide Science.* **1998,** *47* (6), 415-433.
47. Rifkind, D. Studies on the interaction between endotoxin and polymyxin B. *J. Infect. Dis.* **1967**, 433-438.
48. David, S. A. Towards a rational development of anti-endotoxin agents: novel approaches to sequestration of bacterial endotoxins with small molecules. *J. Mol. Recognit.* **2001,** *14* (6), 370-387.
49. Burns, M. R.; Jenkins, S. A.; Kimbrell, M. R.; Balakrishna, R.; Nguyen, T. B.; Abbo, B. G.; David, S. A. Polycationic sulfonamides for the sequestration of endotoxin. *J. Med. Chem.* **2007,** *50* (4), 877-888.
50. Ohto, U.; Fukase, K.; Miyake, K.; Satow, Y. Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science.* **2007,** *316* (5831), 1632-1634.
51. Takayama, K.; Qureshi, N.; Beutler, B.; Kirkland, T. Diphosphoryl lipid A from Rhodopseudomonas sphaeroides ATCC 17023 blocks induction of cachectin in macrophages by lipopolysaccharide. *Infect. Immun.* **1989,** *57* (4), 1336-1338.
52. Mullarkey, M.; Rose, J. R.; Bristol, J.; Kawata, T.; Kimura, A.; Kobayashi, S.; Przetak, M.; Chow, J.; Gusovsky, F.; Christ, W. J. Inhibition of endotoxin response by e5564, a novel Toll-like receptor 4-directed endotoxin antagonist. *J. Pharmacol. Exp. Ther.* **2003,** *304* (3), 1093-1102.
53. Rossignol, D. P.; Lynn, M. Antagonism of in vivo and ex vivo response to endotoxin by E5564, a synthetic lipid A analogue. *J. Endotoxin Res.* **2002,** *8* (6), 483-488.
54. Johnson, D. A.; Sowell, C. G.; Johnson, C. L.; Livesay, M. T.; Keegan, D. S.; Rhodes, M. J.; Ulrich, J. T.; Ward, J. R.; Cantrell, J. L.; Brookshire, V. G. Synthesis and biological evaluation of a new class of vaccine adjuvants: aminoalkyl glucosaminide 4-phosphates (AGPs). *Bioorg. Med. Chem. Lett.* **1999,** *9* (15), 2273-2278.
55. Lewicky, J. D.; Ulanova, M.; Jiang, Z.-H. Improving the immunostimulatory potency of diethanolamine-containing lipid A mimics. *Biorg. Med. Chem.* **2013,** *21* (8), 2199-2209.
56. Kim, H. M.; Park, B. S.; Kim, J.-I.; Kim, S. E.; Lee, J.; Oh, S. C.; Enkhbayar, P.; Matsushima, N.; Lee, H.; Yoo, O. J. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell.* **2007,** *130* (5), 906-917.
57. Shirey, K. A.; Lai, W.; Scott, A. J.; Lipsky, M.; Mistry, P.; Pletneva, L. M.; Karp, C. L.; McAlees, J.; Gioannini, T. L.; Weiss, J. The TLR4 antagonist Eritoran protects mice from lethal influenza infection. *Nature.* **2013,** *497* (7450), 498.
58. Piazza, M.; Yu, L.; Teghanemt, A.; Gioannini, T.; Weiss, J.; Peri, F. Evidence of a specific interaction between new synthetic antisepsis agents and CD14. *Biochemistry.* **2009,** *48* (51), 12337-12344.
59. De Paola, M.; Sestito, S. E.; Mariani, A.; Memo, C.; Fanelli, R.; Freschi, M.; Bendotti, C.; Calabrese, V.; Peri, F. Synthetic and natural small molecule TLR4 antagonists inhibit motoneuron death in cultures from ALS mouse model. *Pharmacol. Res.* **2016,** *103*, 180-187.
60. Shirey, K. A.; Lai, W.; Patel, M. C.; Pletneva, L. M.; Pang, C.; Kurt-Jones, E.; Lipsky, M.; Roger, T.; Calandra, T.; Tracey, K. Novel strategies for targeting innate immune responses to influenza. *Mucosal Immunol.* **2016,** *9* (5), 1173.

61. Rodriguez Lavado, J.; Sestito, S. E.; Cighetti, R.; Aguilar Moncayo, E. M.; Oblak, A.; Lainšček, D. k.; Jimenez Blanco, J. L.; Garcia Fernandez, J. M.; Ortiz Mellet, C.; Jerala, R. Trehalose-and glucose-derived glycoamphiphiles: small-molecule and nanoparticle Toll-like receptor 4 (TLR4) modulators. *J. Med. Chem.* **2014,** *57* (21), 9105-9123.
62. Giuliani, M.; Morbioli, I.; Sansone, F.; Casnati, A. Moulding calixarenes for biomacromolecule targeting. *Chem. Commun.* **2015,** *51* (75), 14140-14159.
63. Salvio, R.; Volpi, S.; Cacciapaglia, R.; Casnati, A.; Mandolini, L.; Sansone, F. Ribonuclease activity of an artificial catalyst that combines a ligated cuII ion and a guanidinium group at the upper rim of a cone-calix[4]arene platform. *J. Org. Chem.* **2015,** *80* (11), 5887-5893.
64. Avvakumova, S.; Fezzardi, P.; Pandolfi, L.; Colombo, M.; Sansone, F.; Casnati, A.; Prosperi, D. Gold nanoparticles decorated by clustered multivalent cone-glycocalixarenes actively improve the targeting efficiency toward cancer cells. *Chem. Commun.* **2014,** *50* (75), 11029-11032.
65. Mochizuki, S.; Nishina, K.; Fujii, S.; Sakurai, K. The transfection efficiency of calix[4]arene-based lipids: the role of the alkyl chain length. *J. Org. Chem.* **2015,** *3* (2), 317-322.
66. Chen, X.; Dings, R. P.; Nesmelova, I.; Debbert, S.; Haseman, J. R.; Maxwell, J.; Hoye, T. R.; Mayo, K. H. Topomimetics of amphipathic β-sheet and helix-forming bactericidal peptides neutralize lipopolysaccharide endotoxins. *J. Med. Chem.* **2006,** *49* (26), 7754-7765.
67. 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.* **2012,** *109* (19), 7421-7426.
68. Ulrich, J. T.; Myers, K. R. Monophosphoryl lipid A as an adjuvant. In *Vaccine design*, Springer: 1995; pp 495-524.
69. Resman, N.; Gradišar, H.; Vašl, J.; Keber, M. M.; Pristovšek, P.; Jerala, R. Taxanes inhibit human TLR4 signaling by binding to MD-2. *FEBS Lett.* **2008,** *582* (28), 3929-3934.
70. Morin, M. D.; Wang, Y.; Jones, B. T.; Su, L.; Surakattula, M. M.; Berger, M.; Huang, H.; Beutler, E. K.; Zhang, H.; Beutler, B. Discovery and Structure–Activity Relationships of the Neoseptins: A New Class of Toll-like Receptor-4 (TLR4) Agonists. *J. Med. Chem.* **2016,** *59* (10), 4812-4830.
71. Wang, Y.; Su, L.; Morin, M. D.; Jones, B. T.; Whitby, L. R.; Surakattula, M. M.; Huang, H.; Shi, H.; Choi, J. H.; Wang, K.-w. TLR4/MD-2 activation by a synthetic agonist with no similarity to LPS. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (7), E884-E893.
72. Goligorsky, M. S. TLR4 and HMGB1: partners in crime? *Kidney Int.* **2011,** *80* (5), 450-452.
73. Park, H.-J.; Hong, J.-h.; Kwon, H.-J.; Kim, Y.; Lee, K.-H.; Kim, J.-B.; Song, S. K. TLR4-mediated activation of mouse macrophages by Korean mistletoe lectin-C (KML-C). *Biochem. Biophys. Res. Commun.* **2010,** *396* (3), 721-725.
74. Unitt, J.; Hornigold, D. Plant lectins are novel Toll-like receptor agonists. *Biochem. Pharmacol.* **2011,** *81* (11), 1324-1328.
75. Sansone, F.; Dudič, M.; Donofrio, G.; Rivetti, C.; Baldini, L.; Casnati, A.; Cellai, S.; Ungaro, R. DNA condensation and cell transfection properties of guanidinium calixarenes: Dependence on macrocycle lipophilicity, size, and conformation. *J. Am. Chem. Soc.* **2006,** *128* (45), 14528-14536.

CHAPTER 4

TLR4 activation

4.1 Introduction

Cell membranes, also known as plasma membrane or cytoplasmic membrane, consist of a lipid bilayer that separates the cytosol from the extracellular fluid.[1] Substances can cross the membrane by passive diffusion, active transport or through transport proteins (forming protein channels) and information, useful for the survival of the cell, is transmitted both ways through embedded proteins. Toll-like receptor 4 (TLR4), as a pattern recognition receptor (PRR), perceives the presence of both damage-associated and pathogen-associated molecular patterns (DAMPs and PAMPS respectively), e.g. bacterial lipopolysaccharides (LPSs), on the outside of the membrane and transmit this signal inside the cell initiating the activation of defense mechanisms.[2-5]

The nature of the lipids composing biological membranes is important for many physiological processes. A number of diseases such as cancers, diabetes, Alzheimer's disease, HIV entry, and atherosclerosis, have been associated to changes in expression levels of individual lipid species.[6-7] A typical plasma membrane is formed by hundreds of amphipathic lipids.[8] Phospholipids together with glycolipids are the most abundant and the large majority of non-lipid membrane components are sterols.[9] The fatty chains (FA) in phospholipids and glycolipids may be saturated or unsaturated and usually contain an even number of carbon atoms, typically between 16 and 20, with 16- and 18-carbon FAs the most common ones. The polar head groups are exposed to water and the nonpolar lipid tail groups from the upper bilayer interact with the ones from the lower bilayer, forming a hydrophobic block.[10-12] The presence of unsaturation in the FA chains is correlated with liquid-disordered (Ld) phases whereas the absence of unsaturation in the FA chains of the lipids and the presence of cholesterol are associated with liquid-ordered (Lo) phases. Regarding the composition, mammalian membrane contains phosphatidylcholine (PC), sphingomyelin (SM), and gangliosides (GM) in the outer leaflet and phosphatidylethanolamine (PE), phosphatidylserine (PS), and other charged lipids in the inner leaflet; in addition eukaryotic plasma membranes contain between 20 to 50% sterols.[10, 13]

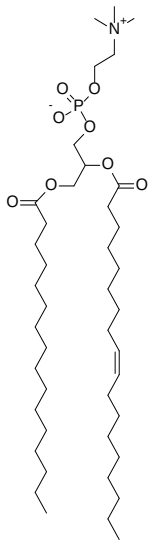
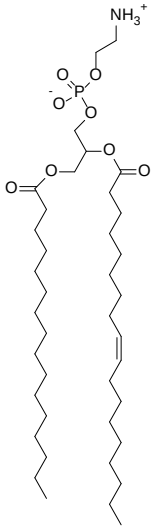
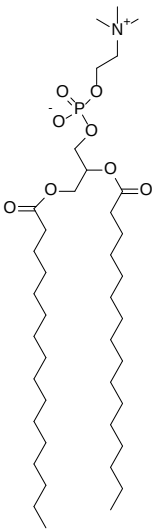
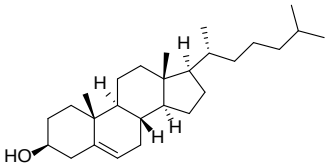
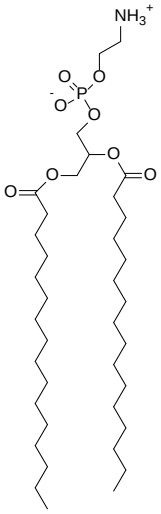
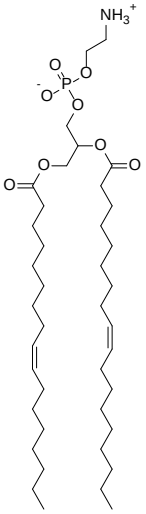
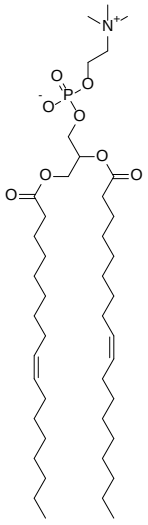
The study of the dynamics of the TLR4/MD-2 complex requires the consideration of different membrane environments, since these are directly involved in the dimerization processes thus governing the activation process.[14-15] We aimed at deepening the understanding of this dimerization process at atomic level using various computational

techniques. We studied independently each TLR4 subdomains, namely the ectodomain (ED), the transmembrane domain (TD), and the intracellular domain (ID), aiming at proposing full TLR4/MD-2 models accounting for its mechanism of activation. We here report MD simulations of the ED in its dimeric and monomeric forms, in complex with MD-2 engaged by *E. coli* LPS. The TD was simulated in Lo and Ld membrane phases to account for the recruitment of TLR4 in lipid-rafts over activation.[14-15] Building our way up to a more complex model, we performed MD simulations of the TD attached to the ID, attempting to explain the reported importance of the long linker joining the two domains.[16-18] A model for the ID/ID complex has also been addressed. Our final goal is, from the information gathered in the previous modelled, to propose a full TLR4/MD-2 dimer model that explain most of the molecular information known to date regarding TLR4 activation.

4.2 Computational considerations about membrane models.

To understand the dynamics of the bilayers themselves before including the TLR4 we constructed a number of symmetric models: POPC, POPE:POPC [1:1], CHL:POPC [1:1], DPPC:POPC [1:1], DPPC:POPE [1:1]. The names and structures corresponding to these abbreviations are reported in Table 4.1. Each of these membrane systems were simulated for 50 ns under anisotropic pressure coupling conditions at a temperature of 303 K. The area per lipid over the simulation time is reported in Figure 4.1 left panel, in order of decreasing compactness our models can be ranked as follows: DPPC:POPE > POPE:POPC > DPPC:POPC > POPC >> CHL:POPC. The electron-density was calculated over the last nanosecond of simulation and is shown in Figure 4.1 right panel. Based on this analysis we can rank the membrane models based on their thickness in ascending order: CHL:POPC < POPC < DPPC:POPC < POPE:POPC < DPPC:POPE.

Table 4.1. Different type of lipids and sterols used in this study.

POPC 1palmitoyl-2-oleoyl-sn-glycero-3-phosphocoline PA-PC-OL	**POPE** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine PA-PE-OL	**DPPC** 1,2-dipalmitoyl-sn-glycero-3-phosphocholine PA-PC-PA	**CHL** Cholesterol
			
DPPE 1,2-dipalmitoyl-sn-glycero-3-phosphoethanolamine PA-PE-PA	**DOPE** 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine OL-PE-OL	**DOPC** 1,2-dipalmitoyl-sn-glycero-3-phosphocholine OL-PC-OL	
			

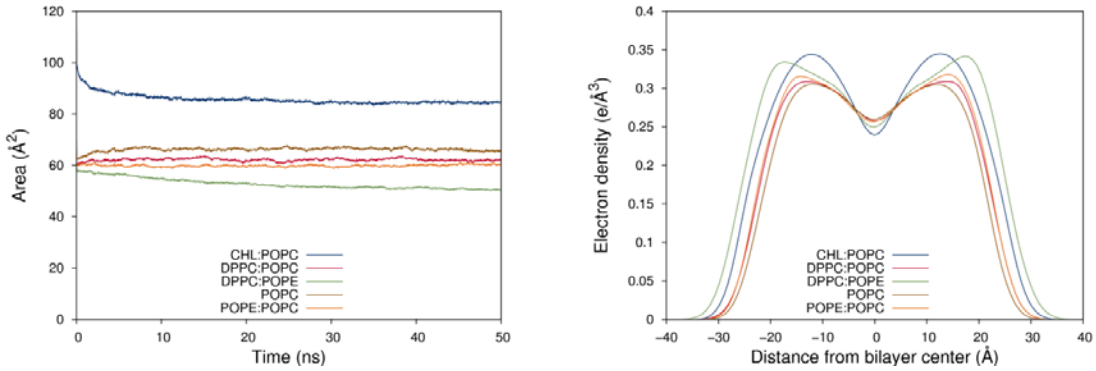


Figure 4.1. On the left panel: area per lipids. On the right: total electron density profiles of the five membrane types (water excluded).

The importance of membrane domains, known as rafts, for proteins signaling and trafficking has been extensively reported.[19-21] The recruitment of TLR4 into lipid rafts domain has also been reported.[22-24] To better represent the activation of TLR4, and based on our previous membranes models, we introduce two new models, namely a liquid-ordered (Lo) membrane model, representing a membrane raft, and a liquid-disordered (Ld) models. Each layer of the Ld model is composed of 64 units of DOPC and each layer of the Lo model of a mixture of 38 DPPC and 26 CHL, approximating a 60:40 ratio. Both membrane models were simulated for 50 ns. The electron density plot from the simulations gives a rough membrane thickness estimates of 45 Å for the Lo model and of 24 Å for the Ld model, similar to the POPC model previously reported (Figure 4.2 right). The area per lipids is around 68 Å^2 for the Lo model, also similar than the one of the POPC model, and around 70 Å^2 for the Ld model (Figure 4.2 left).

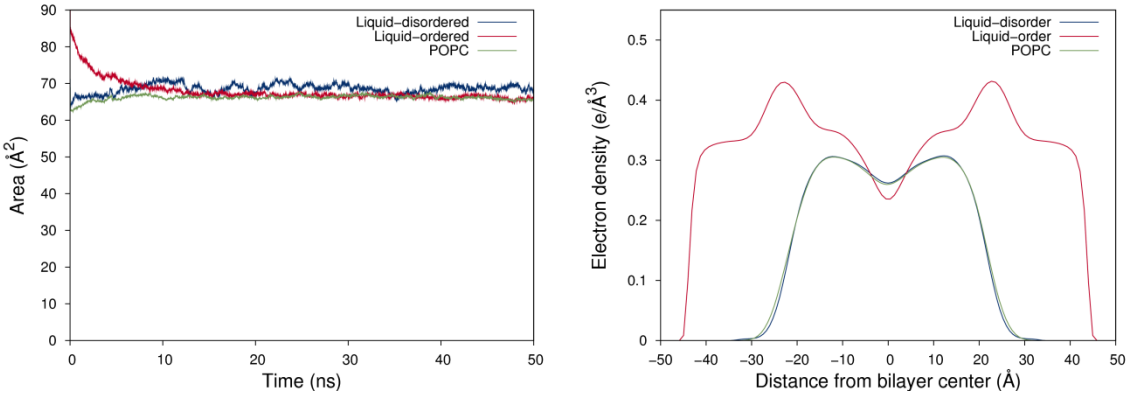


Figure 4.2. On the left: area (in Å^2) occupied by the membrane models Ld, Lo and POPC over time (ns). On the right: total electron density profiles of the five membrane types (water excluded).

In addition to membrane models introduced previously, we considered a more physiologically relevant mammalian plasma membrane, by choosing different

concentration of lipids and sterols in the upper and the lower bilayer, thus building an asymmetric model. The outer leaflet of this membrane is composed by 35% of CHL and 65% of lipids, among which: 30% of DPPC and DPPE, 10% of POPE and POPC, and 10% of DOPE and DOPC. The inner leaflet is composed by 30% of cholesterol and 70% of lipids, among which: 25% of DPPC and DPPE, 10% of POPE and POPC, and 15% of DOPE and DOPC (Figure 4.3). The lipid fractions of the outer and the inner leaflets are thereof respectively composed by 60% and 50% of lipids without unsaturation, 20% and 20% of lipids with only one unsaturation, and by 20% and 15% of lipids with more than one unsaturation, in agreement with data described in Marrink and coworker's paper.[8]

OUTER	**for 263 molecules**	**INNER**	**for 250 moleucles**
65% of lipids:	171 molecules	**70% of lipids:**	175 molecules
0 unsaturation		**0 unsaturation**	
DPPC (30%)	51 molecules	DPPC (25%)	44 molecules
DPPE (30%)	52 molecules	DPPE (25%)	44 molecules
1 unsaturation		**1 unsaturation**	
POPE (10%)	17 molecules	POPE (10%)	17 molecules
POPC (10%)	17 molecules	POPC (10%)	18 molecules
>1 unsaturation		**>1 unsaturation**	
DOPE (10%)	17 molecules	DOPE (15%)	26 molecules
DOPC (10%)	17 molecules	DOPC (15%)	26 molecules
35% CHL	92 molecules	**30% CHL**	75 molecules

Figure 4.3. Composition of the asymmetric membrane model.

In our simulations conditions and during the simulated time we do not expect transverse diffusion of lipids thus it is important to make sure than both layers, when reaching a stable simulation regime, are occupying the same area. To achieve that we simulated both layers composition as independent bilayers changing the quantity of lipids, accordingly to the fraction defined above, until reaching acceptable similarities in simulation boxes sizes after 100 ns of simulations (Figure 4.4, left panel). The final model, which the exact composition is given in Figure 4.3, was simulated for 100 ns. Decomposed electron density profile of the model was plotted based on the trajectory of the last ns of simulation. The difference in unsaturation can be noted in Figure 4.4, right panel, in which the alkene group in the outer leaflet (in the negative coordinates range) has a higher electron density peak that the one in the inner leaflet (in the positive coordinates range).

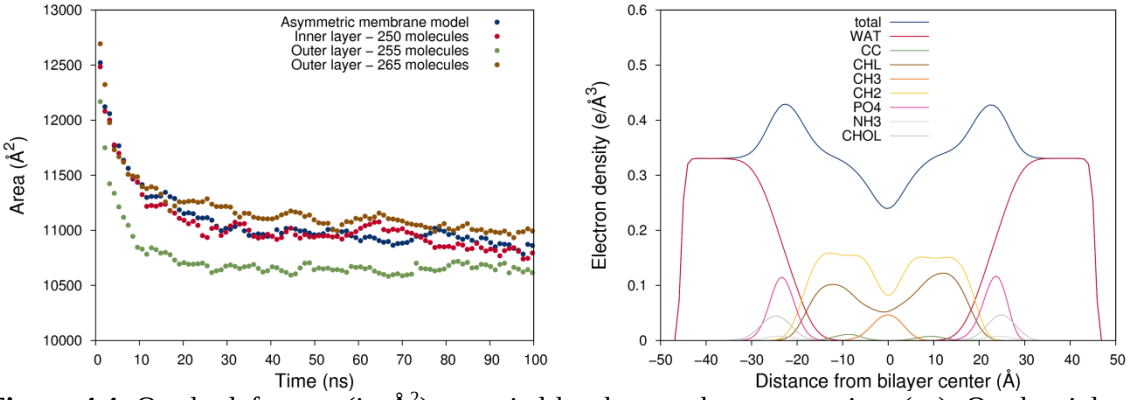


Figure 4.4. On the left: area (in Å^2) occupied by the membrane over time (ns). On the right: decomposed electron density profile for the asymmetric model.

4.3 Computational studies on the TLR4/MD-2 receptor complex

4.3.1 TLR4 ectodomain

The structure of the agonist form of TLR4 in complex with its accessory protein MD-2 engaged by the most potent agonist known to date, *E. coli* LPS, was resolved by X-ray crystallography (PDB ID 3FXI).[25] We previously reported MD simulations of the TLR4/MD-2 complex engaged by ligands of different kind comprising naturally occurring modulators (e.g. LPS from *Burkholderia cenocepacia*),[26] LPS-like molecules (e.g. fluorescent probes),[27] and non-LPS like small molecules (e.g. amphiphilic guanidinocalixarenes). [28] We also previously reported the impact of a single point mutation on the receptor activity.[29] Here we report MD simulations of both the inactivated monomeric form and the activated dimeric form of the ED of TLR4, both in complex with MD-2 engaged by *E. coli* LPS. In the simulation of the dimer, both TLR4s undergo little deviation from their original locations, in relation to MD-2, at the beginning of the simulation and stabilize for the rest of the simulation (Figure 4.5 left panel, blue and red lines). In the simulation of the monomer TLR4 keeps deviating from its original location, in relation to MD-2, along the entire simulation, indicating that the monomeric TLR4/MD-2/*E. coli* LPS complex is not as stable as the dimeric (TLR4/MD-2/*E. coli* LPS)$_2$ complex (Figure 4.5 left panel, right line). This suggests that the presence of a second TLR4/MD-2/*E. coli* LPS unit stabilizes the first, and vice versa, which may be due to the polar interactions taking place at the dimer interface. The ligand, *E. coli* LPS, displays similar deviation pattern in both the simulation of the monomer and the dimer, as indicated by the RMSD plot (Figure 4.5 right panel). However a subtle loss of symmetry between the two LPSs in the dimer simulation can be observed (comparing blue and red lines of Figure 4.5 right panel) due the high

flexibility of the OS-core of LPS, which adopts different conformations in both monomers. The lipid A moiety, buried in the MD-2 pocket, behave similarly along the simulation in both monomers, only a subtle difference in the disposition of the R3' lipid chain can be observed: in one monomer the chain stays within the MD-2 channel, as it is in the crystal structure, and in the other monomer the chain moves towards the MD-2 pocket. Nevertheless, critical interactions of this chains with residues like Phe440* and Phe126 are maintained in both cases. MD-2 shows similar amplitude of conformational deviation in both simulations (Figure 4.6 left panel). Phe126, known as the switch ON/OFF of the receptor complex,[30] remains in an agonist-like conformation in both simulations characterized by an arbitrarily selected angle (Figure 3.14B in *Chapter 3*) plotted in Figure 4.6 (right panel).

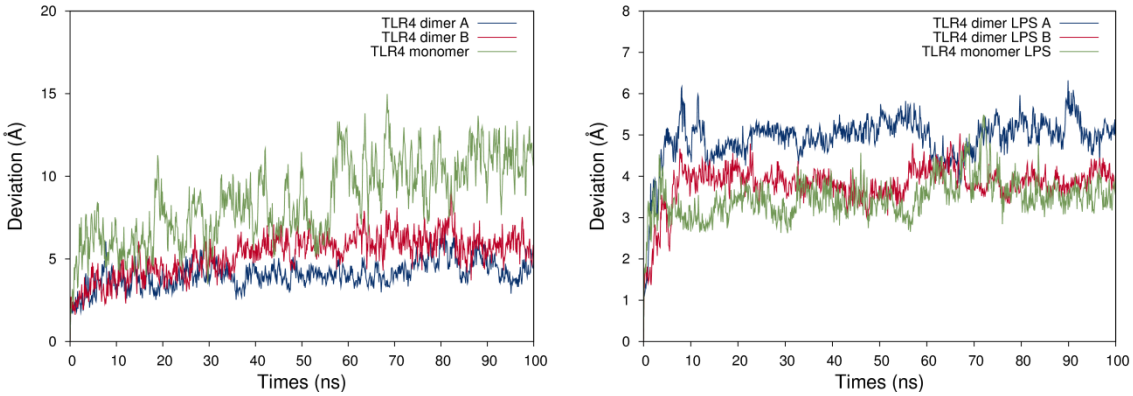


Figure 4.5. In both plots the minimum fit was performed on the αC of MD-2 in complex with a given TLR4, giving information of how the different components of the simulations deviate from their initial position in relation to MD-2. On the left: RMSD of both TLR4 in the simulation of the dimeric model (in blue and red) and TLR4 in the simulation of the monomer (in green). On the right: RMSD of each of the LPS.

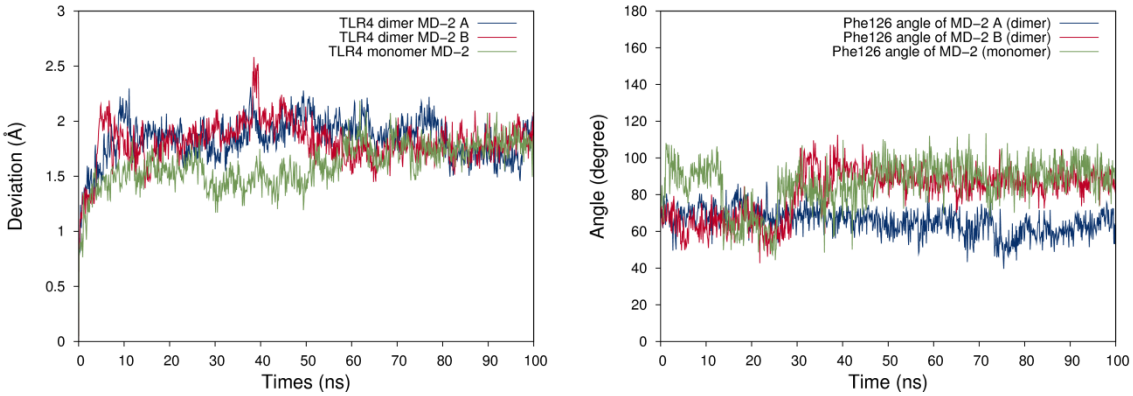


Figure 4.6. On the left: RMSD of both MD-2 in the simulation of the dimeric model (in blue and red) and MD-2 in the simulation of the monomer (in green). On the right: angle between two vectors defined in Figure 3.14B in *Chapter 3*, providing conformational information about the critical residue Phe126 along the simulation.

4.3.2 TLR4 transmembrane domain

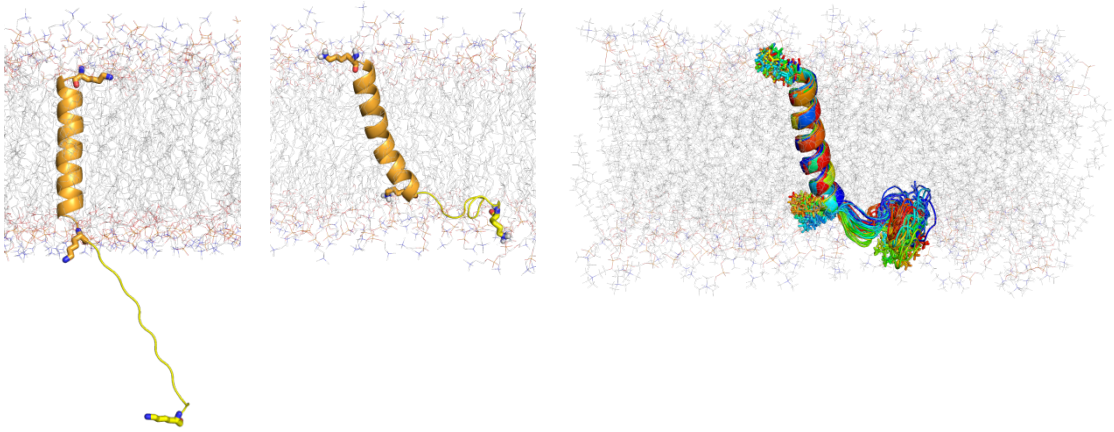
The transmembrane domain of TLR4 (TD) is predicted to span from Lys631 to Lys653 and to consist of an α-helix of lipophilic residues, with few polar residues. In addition, the amino acid sequence directly following this domain, called hydrophobic region (further abbreviated HR), lower delimitated by Lys666, has been largely argue to also actively interact with the membrane, either extending the TD α-helix or interacting with the head group of the lipids.[17-18, 31] Very recently, Mineev *et al.* reported a NMR study of both the TD and the HR of TLR4.[16] The interpretation of the NMR data points to a helical conformation of the HR in DMPG/DHPC bicelles (PDB ID 5NAM). They further performed protein-protein docking experiments and selected a dimeric model based on NMR data. However, it is important to keep in mind that membrane protein secondary structure and membrane protein-protein interactions are highly dependent of the medium in which they are studied. We decided to use computational tools to study the TLR4 TD-TD protein interactions and to propose a model for the TD-TD dimerization.

4.3.3 TLR4 transmembrane domain and hydrophobic region

We performed a 250ns MD simulation of the TD and the HR structured as a long α-helix, as reported in the NMR study,[16] in a POPC membrane. In this simulation the entire peptide enters the membrane and adopts a very tilted disposition (around 45^{o}). The polar side chain of Lys653, found between the TD and the HR, interacts with the head group of the lipids inducing a soft kink in the helix. Lys666 is found outside the membrane exposed to the lipids head group and the solvent (Figure 4.7).

In addition, we performed the same simulation but starting with an extended unstructured HR (Figure 4.8A). Along the simulation the HR residues explore the surrounding of the membrane and largely interact with the lipid head groups without penetrating the membrane. The TD adopts a slightly titled orientation in the POPC membrane (Figure 4.8B). Once the HR starts to interact extensively with the membrane it stabilizes for the rest of the MD simulation (Figure 4.8C), not experiencing major changes during the final 50ns.

Figure 4.7. MD simulations of TLR4 TD and HR (uninterrupted α-helix) domain. At t=0ns of the MD simulation (left), the protein is perpendicular to the mambrane plan. At t=250ns (right), the protein adopt a titled position in relation with the membrane plan.



(**A**) t=0ns (**B**) t=200ns (**C**) t=150-200ns

Figure 4.8. MD simulations of TLR4 TD (α-helix) and HR (extended coil) domain. The TD was built as an α-helix and HR, built in an extended conformation, was attached to it (**A**). At the end of the simulation the HR is folded against the membrane (**B**). (**C**) Superimposition of one frame per nanosecond from the last 50 ns of MD simulation. Lys631, Lys653 and Lys666 are represented as sticks; the TD and the HR are represented as orange and yellow cartoon, respectively. The membrane is in thin lines and the solvent was hidden.

4.3.4 TLR4 TD-TD dimerization

We used the DAFT approach[32] within the Martini Coarse Grained (CG) force field to explore TD-TD interaction in POPC membranes. Among the 550 simulations that were run, 487 were successfully performed for 1024ns. We ranked them based on Lennard-Jones energy of interaction between the two TD (as calculated by the gromacs *gmx energy* command). We selected the top 5% of the successful simulations for closer analysis. 13 out of 24 feature very similar protein-protein interactions (Figure 4.9). In these models we observed hydrophobic interactions between the two equivalent Val636 and hydrogen bonds between the two equivalents Ser640. The model with the highest score, which belongs to largest cluster, was back-mapped to all atom following the standard backmapping protocol described by Wassenaar *et al.*[33] and submitted to a 10 ns

all atom MD simulation with AMBER in a DOPC membrane. No major interaction changes were observed.

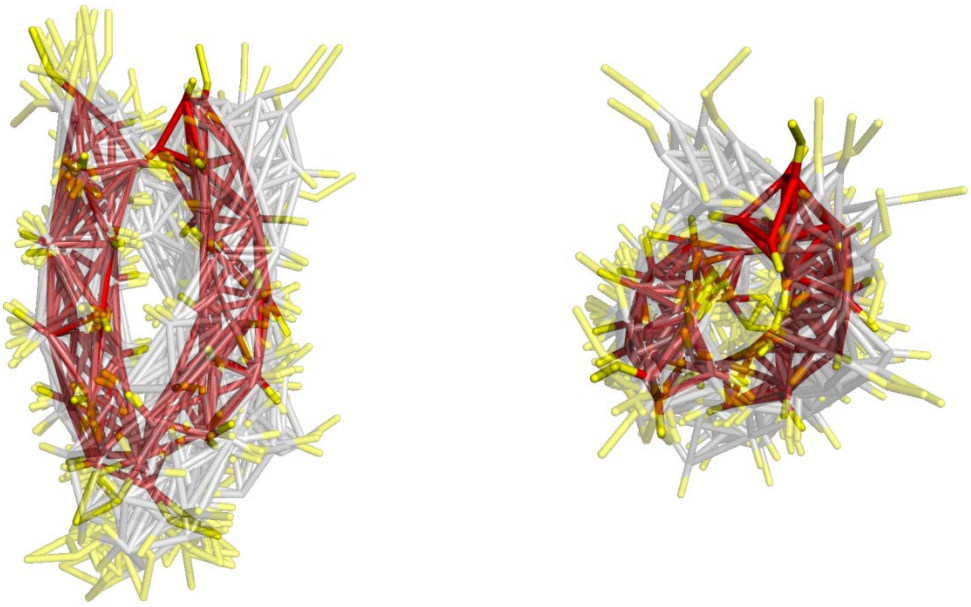


Figure 4.9. Superimposition, based on one monomer, of the 13 selected dimirezed TD-TD poses. Backbone of the pose with the best L-J interactions score is shown in solid red sticks, backbone and side chains of the other poses are shown respectively, in grey and yellow, semi-transparent sticks.

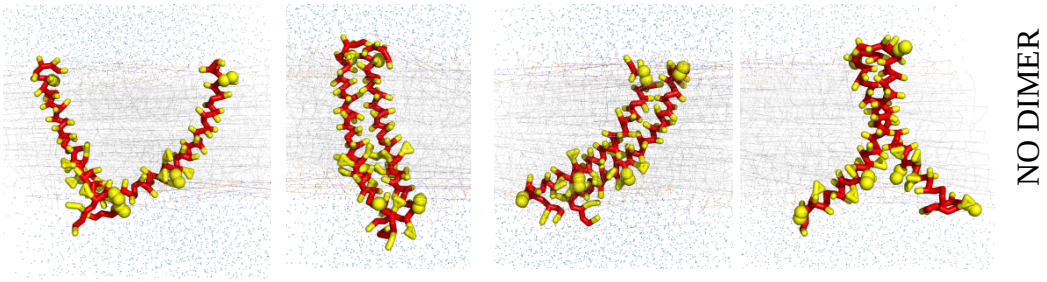
4.3.5 TLR4 TD HR-TD HR dimerization

We performed additional sets of DAFT experiments (4 sets of 120 simulations) to investigate the dimerization behavior of the TD and the HR together, exploring unstructured and α-helical conformations of the HR in both Ld and Lo membrane types.

As it was shown in the MD simulations of the monomer of the TD and the HR structured as a continuous α-helix (Figure 4.7) in order for the dimer of the TD and the HR (from Lys631 to Lys666), structured as an uninterrupted α-helix, to be fully inserted in the membrane, it has to adopt a very tilted angle in comparison to the membrane plan.

In the Ld membrane model, in order to be fully inserted into the membrane, the helix must be tilted, either as a tilted dimer, representing 31% of the poses, leading to a nonsymmetrical dimer which might be compatible with dimerized ED-ED and ID-ID, either as two separated TD domains, representing 32% of the poses, incompatible with a dimerized ED (Figure 4.10). The symmetrical TD-TD dimer is observed in a low percentage (15%) and requires the HR region to be outside the membrane, where it is likely to adopt a non-helical conformation according to our MD simulations (Figure 4.8). This behavior points toward an inactive architecture of the

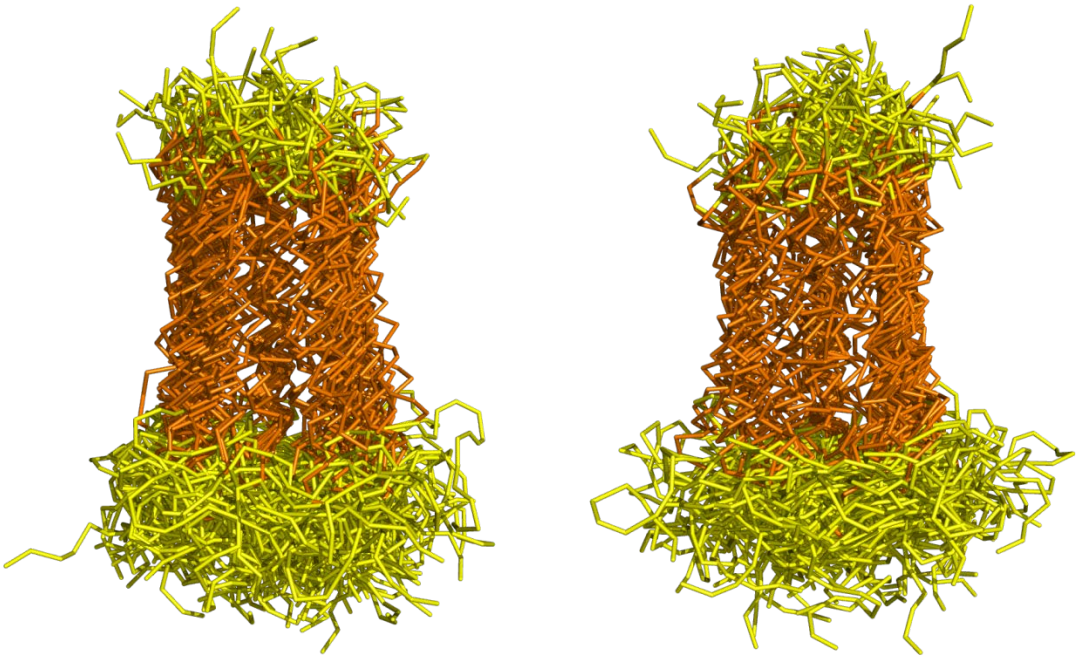
TLR4/MD-2/TLR4*/MD-2* complex in the Ld membrane. In the case of the Lo membrane model, the symmetrical TD-TD dimer is observed in a high percentage of the MD simulation (46%) with the HR region outside the membrane. The tilted TD-TD dimer is also observed (33%). The HR region not being organized as a α-helix outside the membrane would provide flexibility to accommodate the IDs. The presence of an unstructured HR would thus allow the dimerization of the IDs and would point to an active architecture of the TLR4/MD-2/TLR4*/MD-2* complex in the Ld membrane.



	a	b	c	d	
Ld	32%	15%	31%	10%	12%
Lo	12%	46%	33%	3%	0%

Figure 4.10. TLR4 $(TD\text{-}HR)_2$ protein-protein binding mode at the end of the simulations in which TD and HR were parametrized as α-helix.

CG simulations of plausible TD-TD dimers were carried out. The two sets of simulations (in Lo and Ld phases) in which only the HR was imposed to be structure as a α-helix do not point to a preferantial binding mode as it can be seen in Figure 4.11. A deeper analysis of these simulations is in progress.



Liquid-ordered membrane Liquid-disordered membrane

Figure 4.11. Superimposition, based on one monomer, of all the poses from the CG simulations that presented a dimer in the final step of simulation. Backbone of residues between Lys631 and Lys653, which are the residues spanning through the membrane, are represented in orange sticks. Backbone of other residues are in yellow sticks.

4.3.6 TLR4 intracellular domain

The homology modelling algorithm retained 6 templates to base the building process on: the TIR domain of human TLR1, TLR2 (P681h mutant), TLR2 (C713s mutant), TLR6 and TLR10, which it then retrieved from the PDB under the accession codes 1FYV, 1FYX, 1O77, 4OM7 and 2J67, respectively. The program produced 17 models based on these templates, which were then used to build a final hybrid model that was considered best based on its Z-score. The model is a dimer as most of the templates also feature a dimer. We compared our model with all the TLR4 ID models reported to date, one from Gond *et al.*,[34] three from Guven-Maiorov *et al.*[35] and another from Miguel *et al.*,[36] and found that our model present great similarities with the latter one (Figure 1.10 in the *Introduction* chapter).

A monomer was extracted from the HM and submitted to 100 ns of MD simulations. The TIR domain proved to be stable along the simulation (Figure 4.13, left panel, red line). The TIR domain is highly conserved among TIR-domain-containing proteins and is thus likely to have been accurately modelled. The entire ID domain is less stable

(Figure 4.13, left panel, blue line), due to the high motion of the C- and N- terminal linkers that are highly flexible as shown in Figure 4.13, right panel.

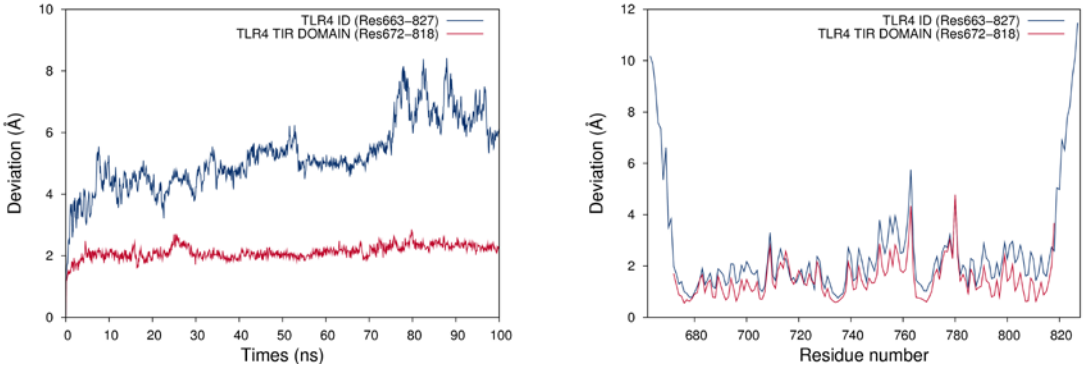


Figure 4.12. MD simulation of monomeric TLR4 ID in water. Left: RMSD. RMS fluctuation per residues

The TLR4 ID dimer model was submitted to 200 ns of MD simulations in water. We characterized the motion of both TIR domains independently and noted higher motion of one of the dimer (Figure 4.14). The higher motions correspond to the CD loop, shown in Figure 1.10 in the *Introduction* chapter (cf. RMSF plot in Figure 4.14, right panel).

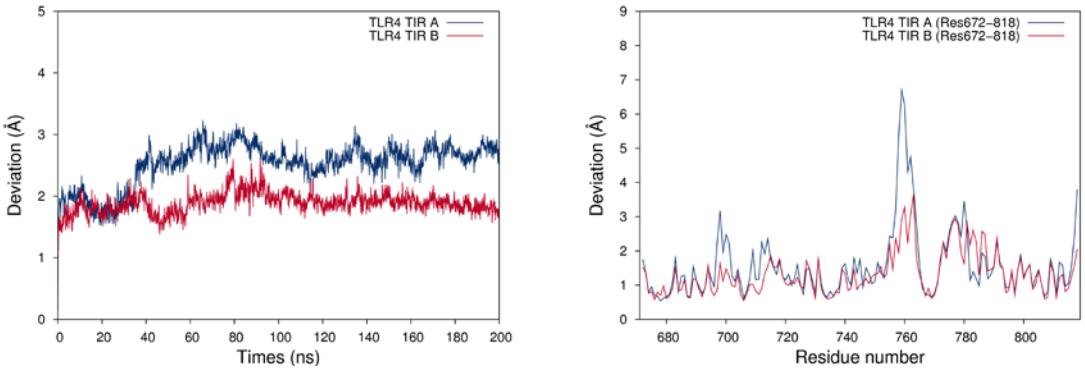


Figure 4.13. MD simulation of dimeric TLR4 ID in water. Left: RMSD. RMS fluctuation per residues

Regarding protein-protein interactions, in the model (starting point of the MD simulation) both the BB loop and the Cys747 are interacting with their counterparts (Figure 4.15). During the simulation, the BB loops interactions are discontinued and one of the monomer rotates 90º in relation to the other, leaving both BB loops exposed to the solvent in a symmetric manner.

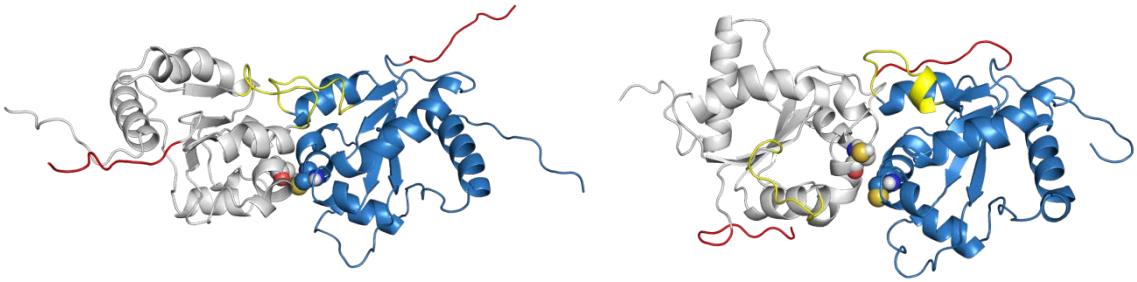


Figure 4.14. TLR4 ID from 0ns (left) to 200ns (right) of MD simulation in water. The dimer is in cartoon representation, Cys747 from both monomer are represented as spheres, chain A is in blue, chain B in grey, the BB loops in yellow, and the N-terminal linkers in red.

In our model we assessed the binding sites of the downstream adaptors based on the findings of Guven-Maiorov *et al.* (2015).[35] The protein-protein interaction model after 200 ns of MD simulation is somewhat similar to a model reported by the same authors (dubbed FF in the original paper) but the overall mode of interaction greatly differs as one monomer is rotated 180 degrees compared to our model. However the interactions with downstream adaptors, as described in the paper, can still take place. The Mal binding site is widely exposed and accessible as the TRAM binding site as presented in the paper.[37] Concerning protein-protein interactions and complexation with downstream adaptors, both models seem to be possible.

4.4 Conclusion

Computational studies of the different independent domains composing the TLR4 were undertaken aiming at uncovering details of the precise mechanism of activation of the receptor. Understanding, at the atomic scale, the dimerization of both the transmembrane domain and the intracellular domain of TLR4 permitted to favor certain binding modes and specific secondary structures increasing the knowledge available regarding the activation. This work is still in progress.

4.5 Materials and methods

Setting up the AA membranes. The lipid bilayers were built using the online CHARMM-GUI membrane builder.[38] Composition of the different system is given in SI. Systems were converted to Lipid14 compatible PDB format using the charmmlipid2amber.sh script.[39]

Full TLR4/MD-2 model construction. The 3D structure of TLR4 and MD-2 were retrieved from the protein data bank under the accession code 3FXI. The TD was built as an α-helix based on its sequence (Uniprot ID O00206). The 3D structure of the ID

was predicted and built through homology modelling as explained in details in the homology modelling section.

Homology modelling. The 3D structure of TLR4 ID was predicted and built within the homology modelling feature of the YASARA program.[40] TLR4 AA sequence spanning from residue 653 to residue 839 was giving as input to the program. Modeling speed was set to slow, which yield best results, and other parameters were kept default.

Molecular dynamics simulations. AA MD simulations were performed with either Amber14 or Amber16. The force fields ff14SB,[41] Lipid14,[42] and a combination of GAFF[43] and GLYCAM06[44] were used to described the proteins, the membrane and *E. coli* LPS, respectively. AA simulations of systems containing membranes went through the same simulation protocol. Steepest descent gradient algorithm is iterated for 5000 steps followed by 5000 iterations of conjugate gradient algorithm under no constraint. The system is then heated from 0 to 100K for 2500 steps in the NVT ensemble while the proteins and the lipids are held by a 10 $kcal.mol^{-1}A^{-1}$ harmonic potential. In the subsequent step the system is heated from 100K to 303K for 50000 steps. In membrane system the dimension of the box can change considerably during the first nanoseconds of simulation, thus to allow the program to recalculate them frequently the first 10 steps of the production run are performed for a maximum of 500ps. In all the steps the temperature is controlled by a Langevin thermostat. The warming up phase and the production run are performed under an anisotropic NPT ensemble to account for different physical properties along the dimensions tangential to the membrane than the one normal to it. The analysis was performed using the cpptraj module of AmberTools15.[45]

Bibliography

1. Malanovic, N.; Lohner, K. Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. *Biochim. Biophys. Acta.* **2016,** *1858* (5), 936-46.
2. Akira, S.; Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **2004,** *4* (7), 499.
3. Beutler, á. TLR4 as the mammalian endotoxin sensor. In *Toll-like receptor family members and their ligands*, Springer: 2002; pp 109-120.
4. Klett, J.; Reeves, J.; Oberhauser, N.; Perez-Regidor, L.; Martin-Santamaria, S. Modulation of toll-like receptor 4. Insights from x-ray crystallography and molecular modeling. *Curr. Top. Med. Chem.* **2014,** *14* (23), 2672-2683.
5. Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B.-S.; Lee, H.; Lee, J.-O. The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature.* **2009,** *458* (7242), 1191.
6. van Meer, G. Cellular lipidomics. *EMBO J.* **2005,** *24* (18), 3159-65.
7. Holthuis, J. C.; Menon, A. K. Lipid landscapes and pipelines in membrane homeostasis. *Nature.* **2014,** *510* (7503), 48-57.
8. Ingolfsson, H. I.; Melo, M. N.; van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; de Vries, A. H.; Tieleman, D. P.; Marrink, S. J. Lipid organization of the plasma membrane. *J. Am. Chem. Soc.* **2014,** *136* (41), 14554-9.
9. Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature.* **2009,** *459* (7245), 379-385.
10. van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **2008,** *9* (2), 112-24.
11. Sampaio, J. L.; Gerl, M. J.; Klose, C.; Ejsing, C. S.; Beug, H.; Simons, K.; Shevchenko, A. Membrane lipidome of an epithelial cell line. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (5), 1903-7.
12. Klose, C.; Surma, M. A.; Simons, K. Organellar lipidomics--background and perspectives. *Curr. Opin. Cell. Biol.* **2013,** *25* (4), 406-13.
13. Mouritsen, O. G.; Zuckermann, M. J. What's so special about cholesterol? *Lipids.* **2004,** *39* (11), 1101-13.
14. Cuschieri, J.; Bulger, E.; Billgrin, J.; Garcia, I.; Maier, R. V. Acid sphingomyelinase is required for lipid Raft TLR4 complex formation. *Surg. Infect. (Larchmt.).* **2007,** *8* (1), 91-106.
15. Triantafilou, M.; Miyake, K.; Golenbock, D. T.; Triantafilou, K. Mediators of innate immune recognition of bacteria concentrate in lipid rafts and facilitate lipopolysaccharide-induced cell activation. *J. Cell Sci.* **2002,** *115* (12), 2603-2611.
16. Mineev, K. S.; Goncharuk, S. A.; Goncharuk, M. V.; Volynsky, P. E.; Novikova, E. V.; Aresinev, A. S. Spatial structure of TLR4 transmembrane domain in bicelles provides the insight into the receptor activation mechanism. *Sci. Rep.* **2017,** *7* (1), 6864.
17. Treeby, M.; Vašl, J.; Ota, P.; Friedrich, J.; Jerala, R. Different functional role of domain boundaries of Toll-like receptor 4. *Biochem. Biophys. Res. Commun.* **2009,** *381* (1), 65-69.
18. Panter, G.; Jerala, R. The ectodomain of the Toll-like receptor 4 prevents constitutive receptor activation. *J. Biol. Chem.* **2011,** *286* (26), 23334-44.
19. Simons, K.; Toomre, D. Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* **2000,** *1* (1), 31-9.
20. Simons, K.; Ikonen, E. Functional rafts in cell membranes. *Nature.* **1997,** *387* (6633), 569-72.

21. Janes, P. W.; Ley, S. C.; Magee, A. I. Aggregation of lipid rafts accompanies signaling via the T cell antigen receptor. *J. Cell Biol.* **1999,** *147* (2), 447-61.
22. Wong, S. W.; Kwon, M. J.; Choi, A. M.; Kim, H. P.; Nakahira, K.; Hwang, D. H. Fatty acids modulate Toll-like receptor 4 activation through regulation of receptor dimerization and recruitment into lipid rafts in a reactive oxygen species-dependent manner. *J. Biol. Chem.* **2009,** *284* (40), 27384-92.
23. Szabo, G.; Dolganiuc, A.; Dai, Q.; Pruett, S. B. TLR4, ethanol, and lipid rafts: a new mechanism of ethanol action with implications for other receptor-mediated effects. *J. Immunol.* **2007,** *178* (3), 1243-9.
24. Plociennikowska, A.; Hromada-Judycka, A.; Borzecka, K.; Kwiatkowska, K. Co-operation of TLR4 and raft proteins in LPS-induced pro-inflammatory signaling. *Cell Mol. Life Sci.* **2015,** *72* (3), 557-81.
25. Park, B. S.; Song, D. H.; Kim, H. M.; Choi, B. S.; Lee, H.; Lee, J. O. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature.* **2009,** *458*, 1191–1195.
26. Di Lorenzo, F.; Kubik, Ł.; Oblak, A.; Lorè, N. I.; Cigana, C.; Lanzetta, R.; Parrilli, M.; Hamad, M. A.; De Soyza, A.; Silipo, A. Activation of human Toll-like receptor 4 (TLR4)· myeloid differentiation factor 2 (MD-2) by hypoacylated lipopolysaccharide from a clinical isolate of burkholderia cenocepacia. *J. Biol. Chem.* **2015,** *290* (35), 21305-21319.
27. Ciaramelli, C.; Calabrese, V.; Sestito, S. E.; Pérez-Regidor, L.; Klett, J.; Oblak, A.; Jerala, R.; Piazza, M.; Martín-Santamaría, S.; Peri, F. Glycolipid-based TLR4 Modulators and Fluorescent Probes: Rational Design, Synthesis, and Biological Properties. *Chem. Biol. Drug Des.* **2016,** *88* (2), 217-229.
28. Sestito, S. E.; Facchini, F. A.; Morbioli, I.; Billod, J.-M.; Martin-Santamaria, S.; Casnati, A.; Sansone, F.; Peri, F. Amphiphilic guanidinocalixarenes inhibit lipopolysaccharide (LPS)-and lectin-stimulated toll-like receptor 4 (TLR4) signaling. *J. Med. Chem.* **2017,** *60* (12), 4882-4892.
29. Vašl, J.; Oblak, A.; Peternelj, T. T.; Klett, J.; Martín-Santamaría, S.; Gioannini, T. L.; Weiss, J. P.; Jerala, R. Molecular basis of the functional differences between soluble human versus murine MD-2: Role of Val135 in transfer of lipopolysaccharide from CD14 to MD-2. *J. Immunol.* **2016,** *196* (5), 2309-2318.
30. Scior, T.; Alexander, C.; Zaehringer, U. Reviewing and identifying amino acids of human, murine, canine and equine TLR4/MD-2 receptor complexes conferring endotoxic innate immunity activation by LPS/lipid A, or antagonistic effects by Eritoran, in contrast to species-dependent modulation by lipid IVa. *Comput. Struct. Biotechnol. J.* **2013,** *5* (6), e201302012.
31. Ruysschaert, J. M.; Lonez, C. Role of lipid microdomains in TLR-mediated signalling. *Biochim. Biophys. Acta.* **2015,** *1848* (9), 1860-7.
32. Wassenaar, T. A.; Pluhackova, K.; Moussatova, A.; Sengupta, D.; Marrink, S. J.; Tieleman, D. P.; Böckmann, R. A. High-Throughput Simulations of Dimer and Trimer Assembly of Membrane Proteins. The DAFT Approach. *J. Chem. Theory Comput.* **2015,** *11* (5), 2278-2291.
33. Wassenaar, T. A.; Pluhackova, K.; Bockmann, R. A.; Marrink, S. J.; Tieleman, D. P. Going Backward: A Flexible Geometric Approach to Reverse Transformation from Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014,** *10* (2), 676-90.
34. Gong, J.; Wei, T.; Stark, R. W.; Jamitzky, F.; Heckl, W. M.; Anders, H. J.; Lech, M.; Rossle, S. C. Inhibition of Toll-like receptors TLR4 and 7 signaling pathways by SIGIRR: a computational approach. *J. Struct. Biol.* **2010,** *169* (3), 323-30.

35. Guven-Maiorov, E.; Keskin, O.; Gursoy, A.; VanWaes, C.; Chen, Z.; Tsai, C. J.; Nussinov, R. The Architecture of the TIR Domain Signalosome in the Toll-like Receptor-4 Signaling Pathway. *Sci. Rep.* **2015,** *5*, 13128.
36. Nunez Miguel, R.; Wong, J.; Westoll, J. F.; Brooks, H. J.; O'Neill, L. A.; Gay, N. J.; Bryant, C. E.; Monie, T. P. A dimer of the Toll-like receptor 4 cytoplasmic domain provides a specific scaffold for the recruitment of signalling adaptor proteins. *PLoS One.* **2007,** *2* (8), e788.
37. Kawai, T.; Akira, S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat. Immunol.* **2010,** *11* (5), 373.
38. Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **2009,** *97* (1), 50-8.
39. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*, University of California, San Francisco., 2012.
40. Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: force-field parameterization in crystal space. *Proteins.* **2004,** *57* (4), 678-83.
41. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-713.
42. Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014,** *10* (2), 865-879.
43. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25* (9), 1157-74.
44. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008,** *29* (4), 622-55.
45. Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013,** *9* (7), 3084-3095.

Conclusions

Regarding naturally occurring modulators, we studied the LPS from *Bradyrhizobium* species. Rhizobia are Gram-negative bacteria able to establish symbiotic relationship with legumes and to reduce atmospheric nitrogen into ammonium, thus providing nitrogen nutrition for the host plants. Bacteria belonging to the *Bradyrhizobium* genus promote nitrogen-fixing nodules development on roots and stems of both wild-growing and cultivated *Aeschynomene* legumes. It was previously demonstrated that the lipopolysaccharide (LPS) macromolecule in Rhizobia plays a key role throughout the symbiotic process and that its structural features are altered in response to plant signals. Different lipid A structures from *Bradyrhizobium* were recently elucidated. They are highly heterogeneous regarding the number, length and nature of their acyl chains. Some contained very long-chains fatty acids and, more surprisingly, a covalently linked hopanoid molecule. That novelty prompted us to evaluate the activity these *Bradyrhizobium* lipid As may have on the innate immune system. Experimental studies, including cell assays on both murine and human bone marrow-derived macrophages and HEK 293-TLR4/MD-2/CD14 cells, revealed an extremely low capability to elicit an immune response. More intriguingly, a potent antagonistic activity towards the toxic *E. coli* LPS was observed. Our computational studies allowed the proposal of plausible binding modes of two of these *Bradyrhizobium* lipid As to the TLR4/MD-2 system. These binding modes account for the potent activity antagonizing the binding of *E. coli* LPS to the MD-2/TLR4 complex thus inhibiting its toxic effects. It is likely that the TLR4 signaling modulation occurs by direct interaction with the TLR4/MD-2 complex, both in its hopanoid-containing and hopanoid-free forms. Our studies do not point toward a primary role of the hopanoid moiety in the biological activity regarding TLR4 signaling.

With respect to LPS-like synthesis modulators, we studied a group of glucosamine derivative. FP7, a glucosamine derivative with two phosphate groups and two myristic (C_{14}) FA chains, is active in inhibiting in a dose-dependent way human and murine TLR4 activation by LPS. NMR experiments suggest that FP7 interact with MD-2, probably inserting its FA chains into hydrophobic binding cavity. We designed new TLR4 modulators, based on FP7, and performed structure-activity relationship (SAR) studies to understand how their FA chains length determine their potency as TLR4

modulators. These FP7 variants differ only in FA chains lengths (10, 12, 14 and 16 carbon atoms). In this study we took into account both the interaction with MD-2 and the aggregation properties of the molecules. We reported structural and functional biological data demonstrating the ability of novel FP variants to negatively regulate TLR4 signaling in different cell model systems. Our computational studies were relevant in the context of the SAR study and to propose the rationale for the mechanism of binding. Our models suggest that there is an optimum length for the FA chains for an appropriate TLR4 antagonist activity related to the binding mode and to the physical-chemical properties of the FP variants.

On the subject on non LPS-like modulators we studied amphiphilic guanidinocalixarenes. To block abnormal TLR4 signaling in bacterial sepsis, two different strategies have been developed. The first one is based on LPS neutralization by the formation of noncovalent adducts with cationic compounds thus preventing LPS from interacting with the receptors. The second strategy is based on the use of molecules that compete with endotoxic LPS in binding to the same site on CD14 and MD-2, thereby inhibiting the induction of signal transduction by impairing LPS-initiated receptor dimerization. Among the amphiphilic guanidinocalixarenes studied, we included one whose activity in this biological context had previously been reported as reference compound. Its biological activity was associated with its capacity to bind and neutralize LPS as topomimetic of LPS-binding peptides. Our computational studies challenged this view. We hypothesized that calixarene-based facial amphiphiles could also be suitable as scaffolds to obtain TLR4 ligands with antagonist activity. In a biological context, amphiphilic calixarenes showed remarkable properties significantly related to their amphiphilicity. Since we hypothesized that calixarene derivatives could directly bind to human and murine MD-2 and CD14 in a similar fashion than LPS, we preliminarily performed docking calculations to support this mode of interaction. In addition, we studied whereas the TLR4 antagonist activity is a rather general property of positively charged amphiphilic calixarenes and if this antagonist effect also derives from the direct interaction of calixarenes with the receptors and not exclusively from LPS neutralizing action, as it was suggested. Experimental evidences showed that some of these calixarenes were active in inhibiting, in a dose-dependent way, the LPS-stimulated TLR4 activation and TLR4-dependent cytokine production in human and mouse cells. Moreover, guanidinocalixarenes also inhibited TLR4 signaling when TLR4

was activated by a non-LPS stimulus, the plant lectin PHA. These results point at the calixarene moiety as a potential scaffold for the development of new TLR4-directed therapeutics.

As for the activation of TLR4, computational studies of the different independent domains composing the TLR4 were undertaken aiming at uncovering details of the precise mechanism of activation of the receptor. Understanding, at the atomic scale, the dimerization of both the transmembrane domain and the intracellular domain of TLR4 permitted to favor certain binding modes and specific secondary structures increasing the knowledge available regarding the activation.