UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

TESIS DOCTORAL

Molecular pattern recognition receptors: approaches from computational chemistry for drug design and innate immunity modulation

Los receptores para el reconocimiento de patrones moleculares: aportaciones de la química computacional para el diseño de fármacos y la modulación de la inmunidad innata

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juan Guzmán Caldentey

Directora

Sonsoles Martín Santamaría

Madrid

FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA ORGÁNICA Y FARMACÉUTICA

MOLECULAR PATTERN RECOGNITION RECEPTORS: APPROACHES FROM COMPUTATIONAL CHEMISTRY FOR DRUG DESIGN AND INNATE IMMUNITY MODULATION

PhD Dissertation presented by

JUAN GUZMAN CALDENTEY

Madrid, 2019

This Thesis have been performed in Centro de Investigaciones Biológicas (CSIC)

Supervisor: Sonsoles Martín Santamaría

FACULTAD DE FARMACIA
DEPARTAMENTO DE QUÍMICA ORGÁNICA Y FARMACÉUTICA

LOS RECEPTORES PARA EL RECONOCIMIENTO DE PATRONES MOLECULARES: APORTACIONES DE LA QUÍMICA COMPUTACIONAL PARA EL DISEÑO DE FÁRMACOS Y LA MODULACIÓN DE LA INMUNIDAD INNATA

Memoria para optar el grado de Doctor presentada por

JUAN GUZMÁN CALDENTEY

Madrid, 2019

Esta Tesis ha sido realizada en el Centro de Investigaciones Biológicas (CSIC)

Directora: Sonsoles Martín Santamaría

UNIVERSIDAD COMPLUTENSE MADRID

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña. Juan Guzmán-Caldentey,
estudiante en el Programa de Doctorado Química médica,
de la Facultad de Farmacia de la Universidad Complutense de Madrid, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Los receptores para el reconocimiento de patrones moleculares: Aportaciones de la química computacional para el diseño de fármacos y la modulación de la inmunidad innata

Molecular pattern recognition receptors: Approaches from computational chemistry for drug design and innate immune modulation

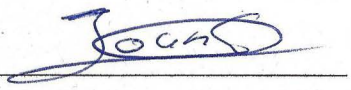
y dirigida por: Prof. Sonsoles Martín-Santamaría

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita.

Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada de conformidad con el ordenamiento jurídico vigente.

En Madrid, a 10 de junio de 2019

Fdo.: 

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Acknowledgements

The writing of this thesis marks the end of a journey that started 5 years ago. I experienced countless changes and transformations since then but, finally, here we are.

First of all, I would like to thank my supervisor, Prof. Sonsoles Martín-Santamaría, the opportunity to work in her lab all these years and to supervise my job with great expertise. She has been more than a boss in various occasions. Thanks a lot!

I would to thank as well Prof. Fukase for welcoming me in his lab in Osaka. I spent three months in Japan and it was an unforgettable experience. I would also like to extend this acknowledgment to all collaborators that have contributed to the work presented in this thesis: Atsushi Shimoyama, Ichinoo Takushi, Carmen Punzón, Prof. Alba Silippo, Prof. Tony Molinaro, Prof. Jesús Jimenez-Barbero, Prof. Anna Bernardi, Prof. Paco Corzana, Prof. José Carlos Menéndez, Prof. Manuel Fresno and Prof. Stefan Oscarson. I would also like to thank the Spanish MINECO for the financial support.

When I first entered the laboratory I met three people that are not ready here, but they have been more than co-workers: Javi, Lucia and Alessandra. Without them, all this time would have been very different, and probably worst. I will always remember that team like if it was the first day. Thanks for the advices, the smiles, the critics, the beer, the trips, the cochinillo segoviano, and more, and more, and more.

I also have to mention all the people that have been in the lab more or less time: Jean-Marc, Laura, Malik, Lide, Elena, Sandra, Quique, Miguel, Carlos, Alejandra, Itziar, … I have learnt precious lessons from all of them. Probably I am missing someone, I apologize in advance.

Outside the lab, but inside the CIB building, I would like to thank Grazia. Thanks for listening to me in the darkest ours. The contribution of ours cigarette breaks to the completion of this Thesis needs to be mentioned.

I would like to mention the Michael Just Furniture band, which was created at the CIB. With them, I accomplished one of my dreams: to record a CD. Carlos, Cris and Iván,

thanks! It is the best band I have never played with. Nevertheless, between them, there is one man that contributed to the completion of this Thesis in a special manner. With him, I had one of the first opportunities to call someone “friend”. Iván, thank you. Our coffee breaks sharing good and bad feelings have (almost literally) saved my life.

Music has played an important role in my life during these years (and longer). I would like to mention here the people from O’Mandala band, I have never faced a group of people where I laugh more. Thanks Gemma, Luis, Alfon, Noe, Bea, Estela, Vicky, Sebas. Moreover, in that band I met another man that I can truly call “friend”. Maikel, I have no idea what I could have done without all the funny experiences we have been through. Certainly, your contribution to this work has to be highlighted as well. Thank you!

Of course, my parents need to be mentioned in these words. During these years, and thanks the growth and maturation I have experienced, I realized how lucky I am to have parents like you. Thanks for supporting me in every decision I make, whatever it is. I need to mention also the financial support from them. It contributed to make my life much easier. Thanks!

Finally, last but not least, I would like to dedicate the last lines to a woman that has specially contributed to let me know what it means to love someone. Martina, you opened to me a full world of experiences and I have learnt (and I am learning) countless lessons from the way you live your life. You are one of the reasons why I keep laughing but you taught me that there are countless reasons out there to do it as well. Thanks!

Table of contents

Table of Abbreviations

AA	All atom
AMPs	Antimicrobial Peptides
APBS	Adaptive Poisson-Boltzmann Solver
CD14	Cluster of Differentiation 14
CG	Coarse grain
CL	Cardiolipin
DAMPs	Damage-Associated Molecular Patterns
DC	Dendritic Cell
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-Grabbing Non-integrin
ED	Ectodomain
EM	Electron Microscopy
FA	Fatty Acid
GlcN	Glucosamine
HEK293	Human embryonic kidney cells 293
HM	Homology Modeling
HMGB1	High Mobility Group Box 1
hPBMCs	human Peripheral Blood Mononuclear Cells
ID	Intracellular Domain
IL	Interleukin
INF	Interferon
IRF-3	Interferon Regulatory Factor 3
LBP	Lipopolysaccharide-Binding Protein
Ld	Liquid-disordered
Lo	Liquid-ordered
LOS	Lipooligosaccharide
LPS	Lipopolysaccharide
LPS_{BV}	*Bacteroides vulgatus* LPS
Mal	MyD88-adapter-like
mCD14	mouse CD14
MDMP	MD-2 Mimetic Peptide
MD simulations	Molecular Dynamics simulations

MD-2	Myeloid Differentiation factor 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
PBE	Predicted Binding Energy
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

In this Thesis we have aimed the study of the molecular recognition processes of receptors involved in the innate immunity. More concretely, we have focused in two different types of lectins, Galectins and DC-SIGN, and in Toll-like receptor 4. We have made use of computational techniques, including docking and virtual screening, molecular dynamics simulations, conformational analysis and quantum mechanics calculations. The work has been organized into several chapters that are summarized as follows:

Chapter 1 corresponds to the current knowledge and perspectives about receptors related to immunity, in particular: galectins, DC-SIGN, and Toll-like receptor 4, corresponding to the molecular recognition events and modulation by small molecules.

Chapter 2 describes the state-of-the-art methods in molecular modeling and computational chemistry applied to the study of molecular recognition processes and drug design.

Chapter 3 reports the fragment-based design of modulators of Galectins 1, 3 and 7, by studying the pocket adjacent to the carbohydrate recognition domain (CRD). While there are no differences in the CRDs among the different galectins, major structural differences can be observed in the adjacent pocket. A fragment-based virtual screening (VS) approach was applied in order to identify fragments able to bind to these pockets in a selective way among the different galectins. With the best fragments, more than 500 lactose-based ligands were designed and docked into all galectins. In collaboration with other groups, selected compounds have been synthesized and preliminary NMR and ITC results are very promising, showing lactose derivatives with high selectivity and affinity towards galectins 1, 3

and 7. These selective ligands may contribute to the understanding of the highly relevant biological functions of galectins and their role in several pathological processes, such as cancer and immune diseases.

Following on chapter 4, a computational study of the DC-SIGN has been carried out aiming the design of glycomimetics with high affinity and selectivity. We have demonstrated by means of MD simulation that the mannose binding site of DC-SIGN and the surrounding regions have a structural plasticity that is crucial to understand the molecular recognition process. We have elucidated that Val351 can adopt two orientations, being one of them clearly favored. We have performed fragment-based VS and identified a pocket where positively charged fragments are hosted with high predicted affinity. This "ammonium binding" pocket, which is adjacent to the mannose binding site, has served as an anchorage site to design new ligands based on a pseudo-dimannoside core. The peculiarity, in this case, is that the modifications have been carried out at the carbon 2, not at the carbon 6, already described in the literature. Newly designed pseudo-dimannosides have been predicted to bind with improved affinity to DC-SIGN.

In chapter 5, we report the studies performed in several TLR4 modulators. First of all, we have demonstrated that RS01 and RS09 peptides, TLR4 agonists, have different binding modes to the TLR4/MD-2 system. RS09 is buried inside MD-2 pocket, while RS01 is located at the rim of the pocket. After MD simulations, we observed that only in the case of RS01, the agonist conformation of Phe126 loop is maintained. This is agreement with the experimental results indicating that RS01 is a better agonist than RS09. The TLR4 antagonist peptide MDMP has also been studied. This peptide is a mimetic of the TLR4 binding region of MD-2, with a disulfide bond in its structure. After a structure optimization, we performed docking and MD simulations of this peptide bound to TLR4. We clearly observed

that MDMP can bind TLR4 in a similar region that MD-2 in a very similar way to MD-2. Thus, we have proposed a binding mode for several peptides to TLR4 accounting for their observed biological activity towards TLR4.

We also have studied the binding mode of cardiolipin (CL), a lipid present in the mitochondria membrane with observed synergistic effect together with LPS, to TLR4. Our computational studies indicate that CL can bind to MD-2, having the FA chains inserted in the hydrophobic pocket. Cardiolipin counters its lower number of FA chains (4) compared to *E. coli* LPS with the length of them (18 against 14), so is able to mimic the interactions that occur at the hydrophobic pocket of MD-2. Then, we assessed the mechanism for the observed synergistic effect by constructing a mixed TLR4/MD-2 dimer, with one monomer occupied by CL and the other by *E. coli* LPS. We observed that this complex is stable and that the agonist conformation of MD-2 in maintained, so we proposed that CL and *E. coli* LPS can form a stable complex that main explain the observed synergistic effect.

Finally, a mixture of LPS from *Bacteroides vulgatus* was studied. This mixture contains a tetracylated LPS and a pentacylated LPS, and our collaborators observed experimentally that the mixture has a weak agonist activity towards TLR4. After docking and MD simulations of both species in TLR4/MD-2 system we observed that only the pentacylated fully accomplish the features of a TLR4 agonist. The tetracylated, on the contrary, was not able to maintain the agonist conformation of MD-2. So, given that the pentacylated has an agonist behavior and the tetracylated has an antagonist behavior, the mixture may act as a weak TLR4 agonist.

On chapter 6, a VS protocol has been applied to several public, commercial and *in-house* libraries in the TLR4/MD-2 system. The VS was performed following both structure-based and ligand-based strategies. Several new chemical scaffolds that could bind to MD-2 pocket have been identified. After the virtual screening protocol and having identified the possible TLR4 binders, we tested the activity of around 50 compounds on HEK-BLUE TLR4 cells and we observed antagonist activity on 7 of them. Assays on 744 macrophage cell line confirmed the antagonist activity of 2 of these compounds, while no cytotoxic effect was observed. Thus, we have identified novel drug-like compounds with TLR4/MD-2 antagonist activity.

Finally, on chapter 7, computational studies of the different independent domains composing the TLR4 were undertaken aiming to uncover 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 available knowledge regarding the activation. We have performed a series of coarse grain simulations of the transmembrane domain in different membrane models. We observed that the dimerization of the transmembrane domain depends on the composition of the membrane. We also have proposed two possible models for the dimerization of the intracellular domain that are in agreement with mutagenesis studies. Finally, we also have provided a model for the binding of MAL adaptor protein to each intracellular dimer.

Resumen

En esta tesis hemos estudiado los procesos reconocimiento molecular de receptores involucrados en la inmunidad innata. Más concretamente, nos hemos centrado en dos tipos diferentes de lectinas, Galectinas y DC-SIGN, y en el receptor Toll-like 4 (TLR4). Hemos utilizado técnicas computacionales, incluyendo *docking* y cribado virtual, simulaciones de dinámica molecular, análisis conformacional y cálculos de mecánica cuántica. El trabajo se ha organizado en diferentes capítulos que se resumen como sigue:

El capítulo 1 corresponde al estado del arte y las perspectivas relacionadas con los estudios de reconocimiento molecular proteína-carbohidrato y diseño de nuevos moduladores con actividad biológica en receptores de la inmunidad, en particular galectinas, DC-SIGN y el receptor Toll-like 4.

El capítulo 2 describe el estado actual de los métodos en modelado molecular y química computacional aplicados al estudio de los procesos de reconocimiento molecular y diseño de fármacos.

El capítulo 3 describe el diseño basado en fragmentos de moduladores de las galectinas 1, 3 y 7, estudiando el bolsillo adyacente al dominio de reconocimiento de carbohidratos. Aunque no hay diferencias en este dominio entre las diferentes galectinas, sí se pueden observar diferencias estructurales en este bolsillo vecino. Aplicamos un protocolo de cribado virtual basado en fragmentos con la finalidad de identificar fragmentos capaces de unirse a este bolsillo de una forma selectiva a las diferentes galectinas. Se seleccionaron los mejores fragmentos para cada bolsillo y se diseñaron más de 500 ligandos basados en lactosa. En colaboración con otros grupos de investigación, se sintetizaron algunos de los ligandos diseñados y

fueron, a continuación, estudiados por técnicas biofísicas (RMN y calorimetría). Los resultados indican que los compuestos presentan una afinidad superior a la de la lactosa (compuesto de referencia) y selectividad por las galectinas. Estos ligandos pueden contribuir al conocimiento de las funciones biológicas de las galectinas y de su papel en distintos procesos patológicos, como el cáncer y las enfermedades inmunológicas.

En el capítulo 4, se ha realizado un estudio computacional de la proteína DC-SIGN para diseñar glicomiméticos con afinidad y selectividad mejoradas. Utilizando métodos de simulación de dinámica molecular, se demostró que el sitio de unión a manosa y los sitios adyacentes a este sitio de unión presentan una plasticidad que es crucial para entender el proceso de reconocimiento molecular. Se ha observado que la Valina 351 puede adoptar dos conformaciones y que una de ellas está claramente favorecida. Se realizó un cribado virtual basado en fragmentos, a través del cual hemos identificado un bolsillo donde los fragmentos cargados positivamente se unen con gran afinidad. Este "bolsillo de unión de amonio", vecino al bolsillo de unión a manosa, ha servido de base para el diseño de nuevos ligandos derivados de un pseudo-dimanósido. En este caso, la peculiaridad reside en que las modificaciones se han realizado en el carbono 2 del manósido, y no en el carbono 6, el cual ya estaba descrito. Estos nuevos ligandos se unen con una gran afinidad a la proteína DC-SIGN.

En el capítulo 5, describimos los estudios realizados con distintos moduladores del TLR4. Primero, hemos demostrado que los péptidos agonistas de TLR4, RS01 y RS09, tienes distintos modos de unión al complejo TLR4/MD-2. RS09 está insertado dentro del bolsillo de MD-2, mientras que RS01 se aloja el borde de este mismo bolsillo. Tras simulaciones de dinámica molecular, observamos que solo en el caso del péptido RS01, la conformación agonista de MD-2 se mantiene. Esta observación

es concordante con resultados experimentales que indican que el péptido RS01 es más potente como agonista del TLR4 que el péptido RS09. También se ha estudiado el péptido antagonista MDMP. Este péptido es un mimético de la región de unión a TLR4 de MD-2. Después de optimizar su estructura, realizamos estudios de *docking* y de dinámica molecular de este péptido unido a TLR4. Hemos observado claramente que MDMP se une a TLR4 de una manera muy similar a MD-2. Así pues, en este capítulo se ha propuesto el modo de unión de diversos péptidos a TLR4 que pueden justificar su actividad experimental.

También se ha estudiado el modo de unión a TLR4 de la cardiolipina (CL), un lípido endógeno presente en la membrana mitocondrial, para determinar el mecanismo del efecto sinérgico observado con LPS. Nuestros estudios computacionales indican que la CL puede unirse a MD-2, con todas sus cadenas alifáticas dentro del bolsillo hidrofóbico de esta proteína. El menor número de cadenas alifáticas (4), comparado con el LPS de *E. coli* (6), se contrarresta con la mayor longitud de éstas (18 frente a 14), y es capaz de mimetizar su modo de unión con los residuos del bolsillo hidrofóbico de MD-2. Además, hemos construido un dímero mixto de TLR4/MD-2, con un monómero ocupado por la CL y el otro por el LPS de *E. coli*. Hemos observado que este complejo es estable y que la conformación agonista de MD-2 se mantiene. Así pues, hemos propuesto que la CL y el LPS de *E. coli* pueden formar un complejo estable que podría explicar el efecto sinérgico observado.

Por último, también hemos estudiado el LPS de *Bacteroides vulgatus*. Este LPS es una mezcla de moléculas de LPS tetra-acilado y penta-acilado, y presenta una actividad como agonista débil del receptor TLR4. Después de estudios de *docking* y simulación de dinámica molecular de estas dos especies de LPS en complejo con TLR4/MD-2, observamos que únicamente el LPS penta-acilado cumple totalmente

las características de un agonista de TLR4. Por el contrario, el tetra-acilado no es capaz de mantener la conformación agonista del MD-2. Así pues, dado que el LPS penta-acilado se comporta como un agonista y el tetra-acilado como un antagonista, la mezcla puede presentar actividad como un agonista débil.

En el capítulo 6, se ha aplicado un protocolo de cribado virtual utilizando quimiotecas (comerciales, públicas y de colaboradores) de compuestos tipo-fármaco y fármacos genéricos en el receptor TLR4. El cribado virtual se realizó siguiendo estrategias basadas en la estructura del ligando y la estructura del receptor. A través de este método, se han identificado nuevos compuestos que podrían unirse al bolsillo de MD-2. Después de este cribado virtual, la actividad de los 50 compuestos seleccionados fue ensayada en la línea celular Hek Blue TLR4 y se observó actividad antagonista en 7 de ellos. Posteriores ensayos en la línea de macrófagos J774 confirmaron la actividad de dos de ellos, mientras que no se observaba citotoxicidad. Así pues, se han identificado nuevos compuestos tipo-fármaco (evitando la estructura de glicolípido) con actividad antagonista sobre el complejo TLR4/MD-2.

Finalmente, en el capítulo 7, se han realizado estudios computacionales de los diferentes dominios del receptor TLR4, con el objetivo de comprender más profundamente los detalles precisos de la activación de este receptor y de su proceso de dimerización a nivel atómico. Hemos realizado simulaciones "coarse grain" del dominio transmembrana en diferentes modelos de membrana. Hemos observado que la dimerización de este dominio depende de la composición de la membrana. También se habn propuesto dos posibles modelos de dimerización del dominio intracelular concordantes con estudios experimentales descritos de mutagénesis. Finalmente, hemos propuesto un modelo de unión de la proteína MAL para cada modelo de dímero intracelular.

CHAPTER 1:

Introduction

INTRODUCTION

The immune system of all living organisms emerged as a protection against environmental microorganisms that can be hazardous for them. In the case of vertebrate immune system, it comprises innate and acquired immunity. Both types of immunity are essential for the survival of the organism, but innate immunity is the first barrier of defense and, thus, less specific. The lower specificity of this system indicates that the same component of the innate immune system can recognize and neutralize different pathogens.[1] One of these components of the innate immune system comprises the so called Pathogen Recognition Receptors (PRRs). These receptors recognize distinct structural motifs that are present in a set of microorganisms, giving to opportunity to the organism to give a quick and effective response. The structural motifs that are associated with hazardous microorganisms are called Pathogen Associated Molecular Patterns (PAMPs). Therefore, PAMPs recognition by PRRs is one of the main events of the innate immunity.

1.1 TLRs

Toll-like receptors (TLRs) are PRRs that have a fundamental role in innate immunity. They are responsible of the initiation and propagation of the inflammation and they have a role in inflammatory and immune diseases, cancer and other pathogenic processes.[2] In 2011, Beutler and Hoffmann were awarded with the Nobel Prize in Medicine[3] for their studies on TLR and the activation of the innate immune system, highlighting the increasing importance of these receptors in drug design. TLRs antagonist may be beneficial to control an inappropriate TLR stimulation and, thus, helpful to treat inflammation and autoimmune diseases. On the other hand, TLRs agonists can help in the treatment of cancer, viral infections

and as adjuvants for vaccines.[4] The first identified TLR was TLR4, the mammalian endotoxin sensor.[5] This TLR recognizes lipopolysaccharide (LPS) from Gram-negative bacteria. The structure of TLR4 comprises a Toll-like extracellular domain, and helical transmembrane domain and a TIR (Toll/Interleukin 1 receptor) intracellular domain.

In 1989, Janeway proposed that cells use pattern recognition to identify pathogens.[6] Receptors recognize and subsequently bind to structural shapes or patterns called pathogen-associated molecular patterns (PAMPs) which are present in entire groups of pathogens, but not in the host. According to Janeway's theory, receptors cannot precisely recognize a particular microbe, but they can identify it as a foreign entity. Ten years after Janeway´s proposal, the first human pattern-recognition receptors (PRRs) were identified. Using the amino acid sequence of the Toll gene from the fruit fly,[7] related sequences were searched in the Human Genome Project database, finally leading to the identification of TLRs.

Since some TLR binders are originated from the host, these new ligands are hypothesized to act as damage signals (damage-associated molecular patterns, or DAMPs) to alert the body about cell or tissue injury. For example, this is evident in cases of necrosis, ischemic injury, etc.[8] Blocking various TLRs (such as TLR2 and TLR4) with antagonists may be useful in these circumstances to prevent an overactive immune response. There is also evidence that TLRs contribute to the development of atherosclerosis and Alzheimer's disease through sensing damage signals in the form of oxidized lipoproteins.[9]

Given their therapeutic potential, there is considerable interest in pharmaceuticals able to modulate TLR activation. TLR antagonists hold great clinical promise for the treatment of numerous inflammatory diseases and they are under investigation for the treatment of viral infections, through redirecting allergic T cell responses

and as anticancer therapeutics. Some TLR agonists have also proven to be safe and efficacious as vaccine adjuvants in humans and they are currently used in Europe.[10]

The human TLR family comprises 10 to 12 type I transmembrane glycoproteins with a single transmembrane domain, a conserved cytoplasmic Toll-like/interleukin-1 receptor signaling domain, and an extracellular antigen recognition domain comprising of 19–25 tandem leucine-rich repeat (LRR) modules.[11] The LRR modules have approximately 20-30 amino acid residues with conserved ''LxxLxLxxN'' motifs.[12]

TLRs function mainly as heterodimers (Figure 1). There are a wide variety of ligands with distinct PAMPs and that is the reason why ten human TLRs are able to recognize more than ten different PAMPs. In fact, the list of known TLR binders keeps growing. Heterodimer formation also increases binder diversity: for example, TLR1/TLR2 heterodimer recognizes triacylated lipopeptides, while TLR2/TLR6 recognizes diacylated lipopeptides.[13] The association with proteins outside of the TLR family also increases the complexity of the molecular recognition process; for example, TLR4 recognizes LPS in association with the accessory proteins MD-2 (myeloid differentiation factor 2) and CD14 (cluster of differentiation 14). Upon binding of the ligands to the extracellular domains of TLRs, rearrangement of the receptor complex is promoted, thus triggering the recruitment of specific adaptor proteins in the intracellular TIR domains.[14] In particular, MyD88 is a universal adapter protein used by almost all TLRs (except TLR3) to activate the transcription factor NF-κβ. Mal (also known as TIRAP) is another adaptor protein necessary to recruit MyD88 into TLR2 and TLR4. TLR expression is particularly significant in different types of white blood cells: mast cells, macrophages, and dendritic cells. The innate immune response is initiated by

mast cells and macrophages, whereas the adaptive immune response is primarily initiated by dendritic cells.[1]

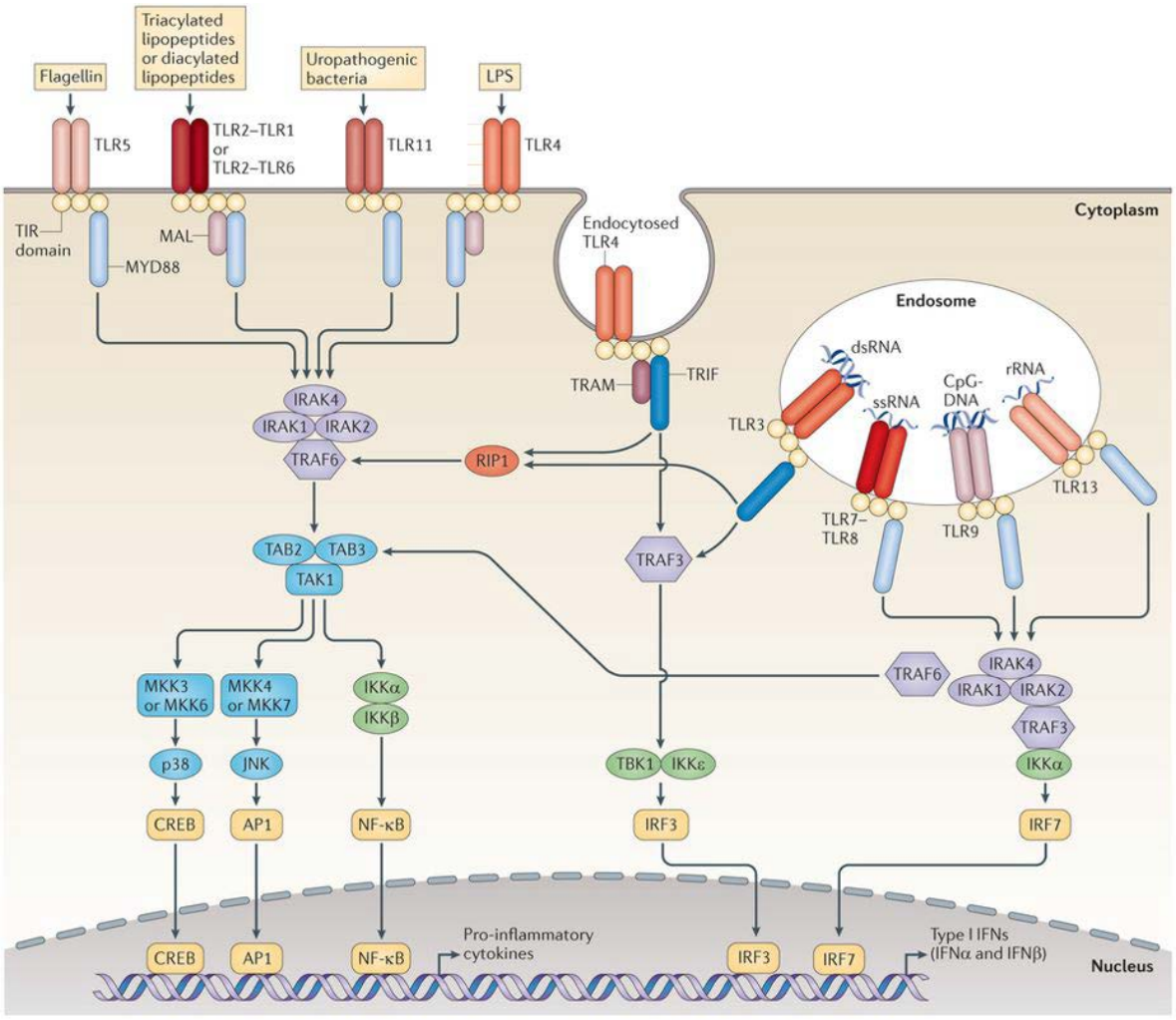


Figure 1: The Toll-like receptor family. Adapted from[15].

TLRs 1, 2, 4, 5, and 6 are located primarily in the plasma membrane, where they recognize components of microbial cell walls and membranes that are present exclusively in pathogens. The best characterized ligands are bacterial. Examples include lipopolysaccharide (LPS) and lipoteichoic acid from the cell wall, lipoproteins from the cell membrane, and flagellin, a structural component of bacterial flagella. TLRs 3, 7, 8, and 9 are situated in the membranes of endosomes and lysosomes. These TLRs bind to microbial nucleic acids, such as DNA from

most organisms, and double and single stranded RNA from RNA viruses. Since these TLRs cannot distinguish self-nucleic acids (those of the host cell), recognition of foreign nucleic acids (those from the pathogen) largely depends on the localization in the cell.[16]

1.2 TLR4

TLR4 is an interesting case of study not only because is the only TLR that requires an accessory protein, called MD-2 (Myeloid differentiation factor 2), to perform its function but because it can activate the immune response through two different signaling pathways and, more importantly, subtle changes in the structure of PAMP lipopolysaccharides can lead to an agonist or an antagonist response of TLR4.[17]

The activation mechanism of TLR4 by LPS goes as follows: after an infection by Gram-negative bacteria, the LPS present in the cell wall binds to lipopolysaccharide binding protein (LBP). LBP brings the LPS to CD14, the following component of the LPS-sensing machinery. CD14 is present extracellularly in the plasma membrane and transfers LPS to MD-2, the accessory protein of TLR4. TLR4 and MD-2 form a highly stable heterodimer and, after the binding of LPS to MD-2, the formation of TLR4/MD-2/TLR4*/MD-2* dimer is induced. Although both LBP and CD14 are necessary to the TLR4 activation by LPS, huge amounts of LPS can also activate it. Finally, the TLR4/MD-2/TLR4*/MD-2* dimer formation brings the intracellular domains close to each other, inducing its dimerization (Figure 2). Once the intracellular domains are in the form of a dimer, the recruitment of the signaling adaptors takes place and induces the activation of the immune response. The activation of the TLR4 complex triggers the release of pro-inflammatory cytokines. After that, the maturation of dendritic cell

is induced, as well as the expression of type-1 interferon and IFN-regulated genes. TLR4 can activate the immune response through two different pathways: the MyD88 dependent pathways and the TRIF-dependent pathways. This different activation leads to the recruitment of different downstream adaptors. In the MyD88 dependent pathway, the MyD88-adapter-like (MAL) protein and Myeloid differentiation factor 88 (MyD88) are recruited. This pathway regulates the activation of NF-ƙB nuclear factor. On the other hand, TIR-domain-containing adapter-inducing interferon-β (TRIF) and TRIF-related adapter molecule (TRAM) are recruited in the TRIF-dependent pathway and TNF-α is activated. Furthermore, this pathway requires the endocytosis of the activated TLR4 complex completely.[18]

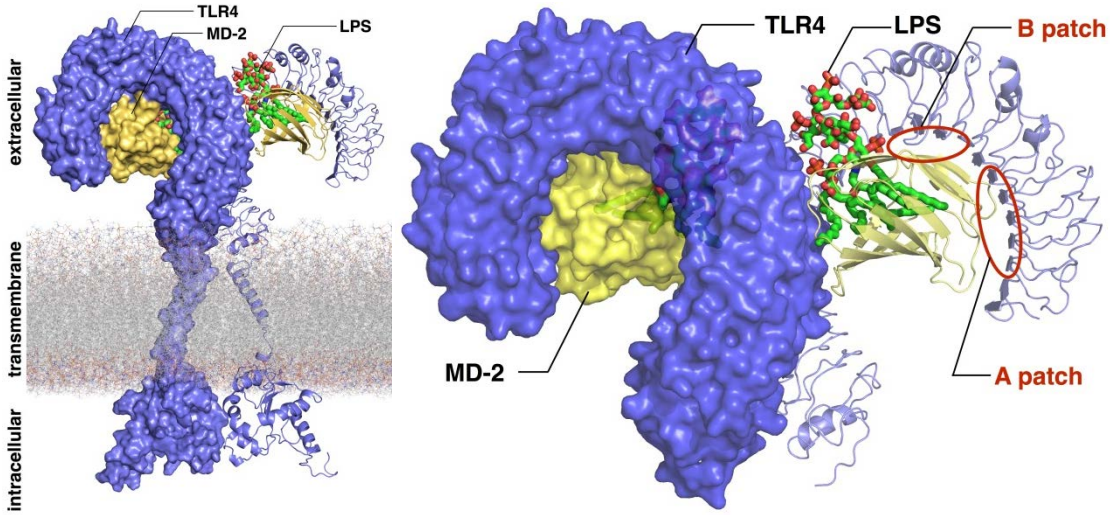


Figure 2: Left) Representation of the full TLR4/MD2 heterodimer. Right) Detail of the TLR4 ectodomain.

To describe the structure of the TLR4 complex, which comprises TLR4 protein and its accessory protein MD-2, we have to describe the structure of its component separately.

TLR4 belongs to the superfamily of the LRR proteins, which they have a typical horse-shoe-like conformation. This conformation is characterized form presenting

parallel β-strands in its concave surface and unstructured loop in its convex surface (Figure 2). The whole sequence can be divided in three regions: N-terminal domain, containing modules 1 to 6; central domain, with modules from 7 to 12; and C-terminal domain, containing modules 13 to 22. The central and N-terminal domains provide charge complementarity to bind MD-2.[19]

The structure of the MD-2 is characterized by two antiparallel β-sheets, containing three and six β-strands, respectively (Figure 3). These two β-sheets adopt a β-cup-like fold, forming a large internal hydrophobic pocket of approximately 1000 $Å^2$. This pocket will be able to host the large lipophilic fatty acid (FA) chains from the LPS. This pocket is completely built by hydrophobic residues on the interior, and positively charged residues surrounding the entrance, at the rim of the pocket, allowing the binding of the phosphate groups and the sugars present in the LPS structure. The formation of the disulfide bond between Cys25 and Cys51, and the hydrogen bond between Tyr34 and Tyr36, are crucial for the stability of the MD-2 structure. Here, we should mention the loop between residues Ile124 and Phe126. This loop is the only part of the protein that changes its conformation, depending of the nature of the ligand that is bound to MD-2 (i.e. if it is agonist or antagonist). When an agonist ligand is bound, Phe126 is buried in the MD-2 pocket, while Ile124 is more exposed to the solvent. On the contrary, when an antagonist ligand is inside MD-2, Phe126 changes its conformation and exposes its side chain to the solvent, while Ile124 occupies the initial position of Phe126. It is said that this twist in the loop acts as a switch ON/OFF (Figure 3).[20]

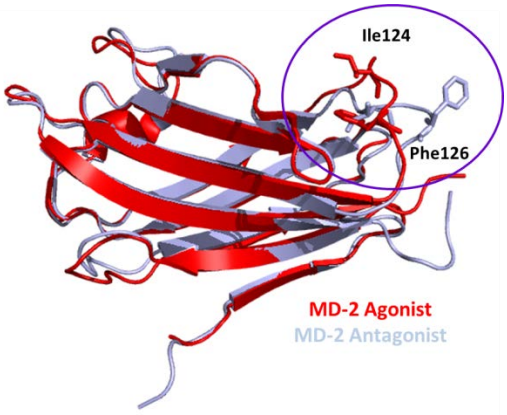


Figure 3: MD-2 agonist structure (red, PDB ID 3FXI) and MD-2 antagonist structure (grey, PDB ID 2E59). Ile124 and Phe126 are highlighted in sticks.

The TLR4/MD-2 complex interacts through a narrow interface, the so-called primary contact interface. The formation of the TLR4/MD-2 complex occurs before the binding of LPS, so the LPS binds to the preformed TLR4/MD-2 dimer (Figure4). The narrow TLR4/MD-2 binding interface is divided into A and B patches, which contain residues of the N-terminal and central domains of TLR4, respectively.[19]

The binding of LPS to TLR4/MD-2 is determined mainly by two parts of the LPS structure: the lipid A and the sugar core (Figure 5). Lipid A is formed by 6 saturated fatty acid chains linked to two N-acetylglucosamines with 1 phosphate group in its carbon 4.[21] The fatty acid chains of lipid A are buried inside the hydrophobic pocket of MD-2, interacting with the hydrophobic residues of this region while the sugars interact with the residues located at the rim of MD-2 through hydrogen bonds mainly. The phosphate groups stablish polar contacts with both MD-2 and TLR4. On the contrary, the sugar core of LPS stablishes a network of hydrogen bonds with the residues of TLR4 located and the central domain of the protein.[22]

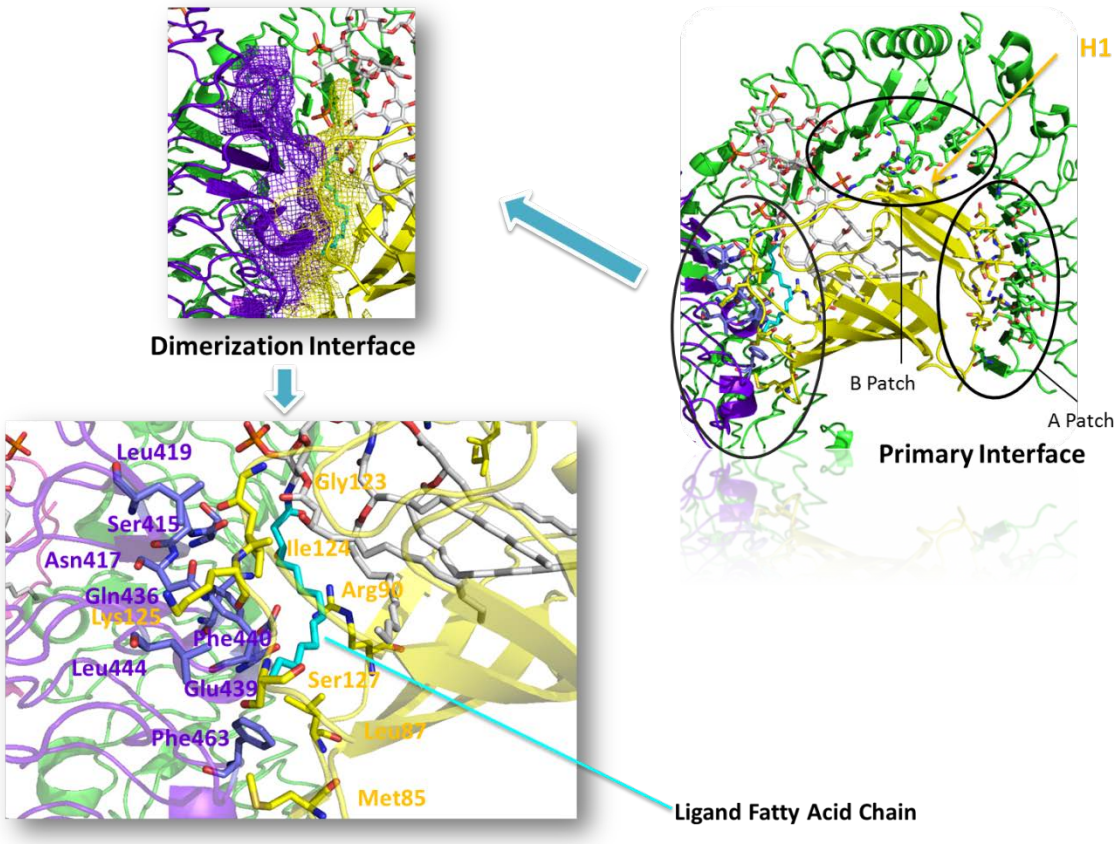


Figure 4: Representation of the dimerization interface of TLR4.

Once the LPS is bound to TLR4/MD-2, the formation of the multimeric complex composed by two copies of TLR4/MD-2/LPS occurs. TLR4 interacts with its partner through the C-terminal domains. This is the primary dimerization interface of the complex. Polar interactions can be observed mainly in this interface, for example between His458 of both TLRs as well as the interaction between Asn433 and Asn433 of the partner TLR4. There is also another dimerization interface that helps to the maintenance of the dimer formed by the MD-2 and the partner TLR4*. In this interface, which includes residues of the C-terminal domain of the TLR4 and residues located at the Phe126 loop, there can be seen, apart from polar contacts, an hydrophobic interaction between Leu87 of MD-2 and Phe463 of TLR4*. An important detail to point out in this interface is the fact that the 6th FA chain of LPS

protrudes from the hydrophobic pocket of MD-2 and interacts with Phe440 and Phe463 of the partner TLR4. Finally the sugars at the sugar core of LPS can reach the partner TLR4, stablishing hydrogen bonds with it and, thus, contributing to the stability of the complex.[19]

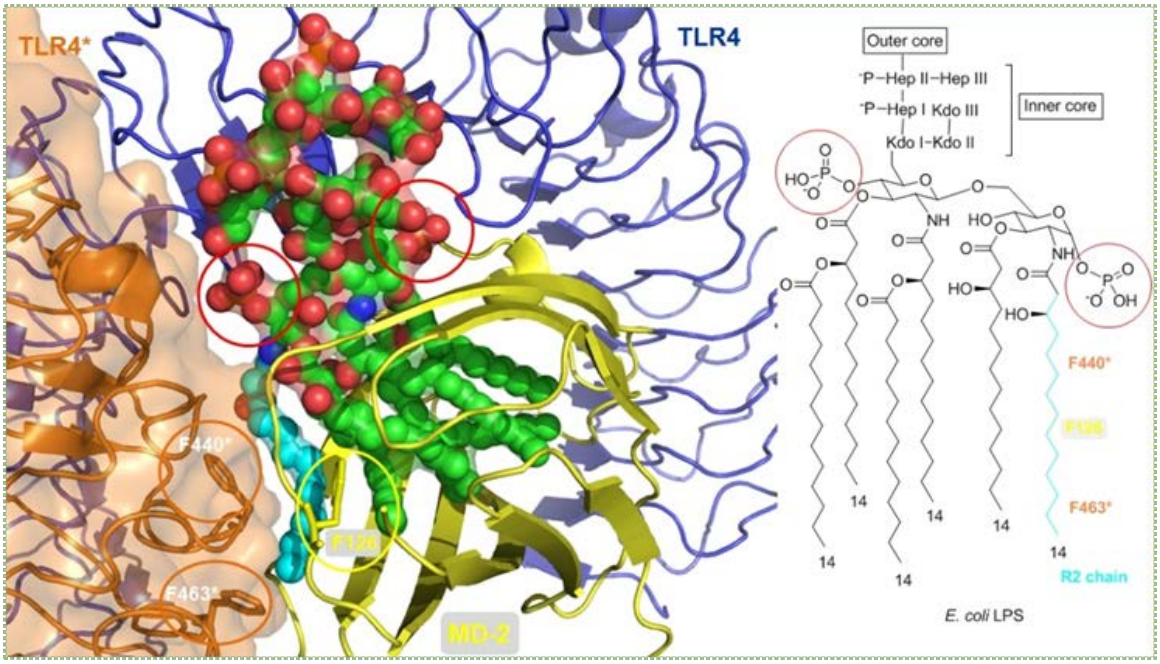


Figure 5: Structure of E. coli LPS and its binding mode on TLR4/MD-2. FA chain that interacts with the partner TLR4 in the activated form of this receptor is marked in blue.

The agonistic activity of LPS has been attributed to the number, length and chemical structure of its FA chains, as well as to its phosphorylation degree. However, recent findings have questioned this paradigm because LPSs bearing penta-acylated lipid A and positively-charged residues decorating their lipid A molecules with agonistic activity have been reported.[23] These data suggest that subtle changes in lipid A structure may have a deep effect on its activity. For example, tetracylated lipid IVa, a lipid A precursor, has an antagonist behavior in human and an agonist behavior in mouse (Figure 6). X-ray crystallographic structures of *h*MD-2/lipid IVa (PDB ID 2E59) and *m*MD-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.[20, 24] Thus, lipid IVa presents two different molecular patterns of interaction for human and mouse. 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.

MD-2 can only be found in the antagonist conformation, either bound to antagonists (MD-2/lipid IVa, PDB ID 2E59, and TLR4/MD-2/Eritoran, PDB ID 2Z65), or without any ligand (TLR4/MD-2, PDB ID 2Z64). This observation may suggest that the conformational change of the Phe126 loop which leads to the dimerization, and to the activation of the immune response, is promoted upon agonist binding. The MD-2 conformational change could be explained by the induced fit paradigm rather than by the conformational selection from the MD-2 conformational landscape. This has also been suggested by reported NMR studies on hexaacylated endotoxin bound to wild-type and F126A mutant MD-2, which indicate that re-orientation of the aromatic side chain of Phe126 is induced by binding of hexaacylated endotoxin, preceding interaction with TLR4.[25] Multimeric complexes are available with natural agonist binder LPS (PDB ID 3FXI, human TLR4/MD-2), and with lipid IVa (PDB ID 3VQ1, mouse TLR4/MD-2).

In this way, lipid IVa acts as a dual binder of TLR4 depending on the species, able to activate the TLR4 immune response in mice, but to block the TLR4 system in humans. As an agonist, it is able to bridge between the two phosphate binding

sites of the two TLR4/MD-2 units. As an antagonist (the glucosamine backbone is in opposite orientation to that for LPS) it is buried more deeply into the cavity of MD-2 (4-5 Å) and thus can only connect to one phosphate binding site. Eritoran is not able to bridge between the two TLR4/MD-2 subunits of any of the analyzed species.

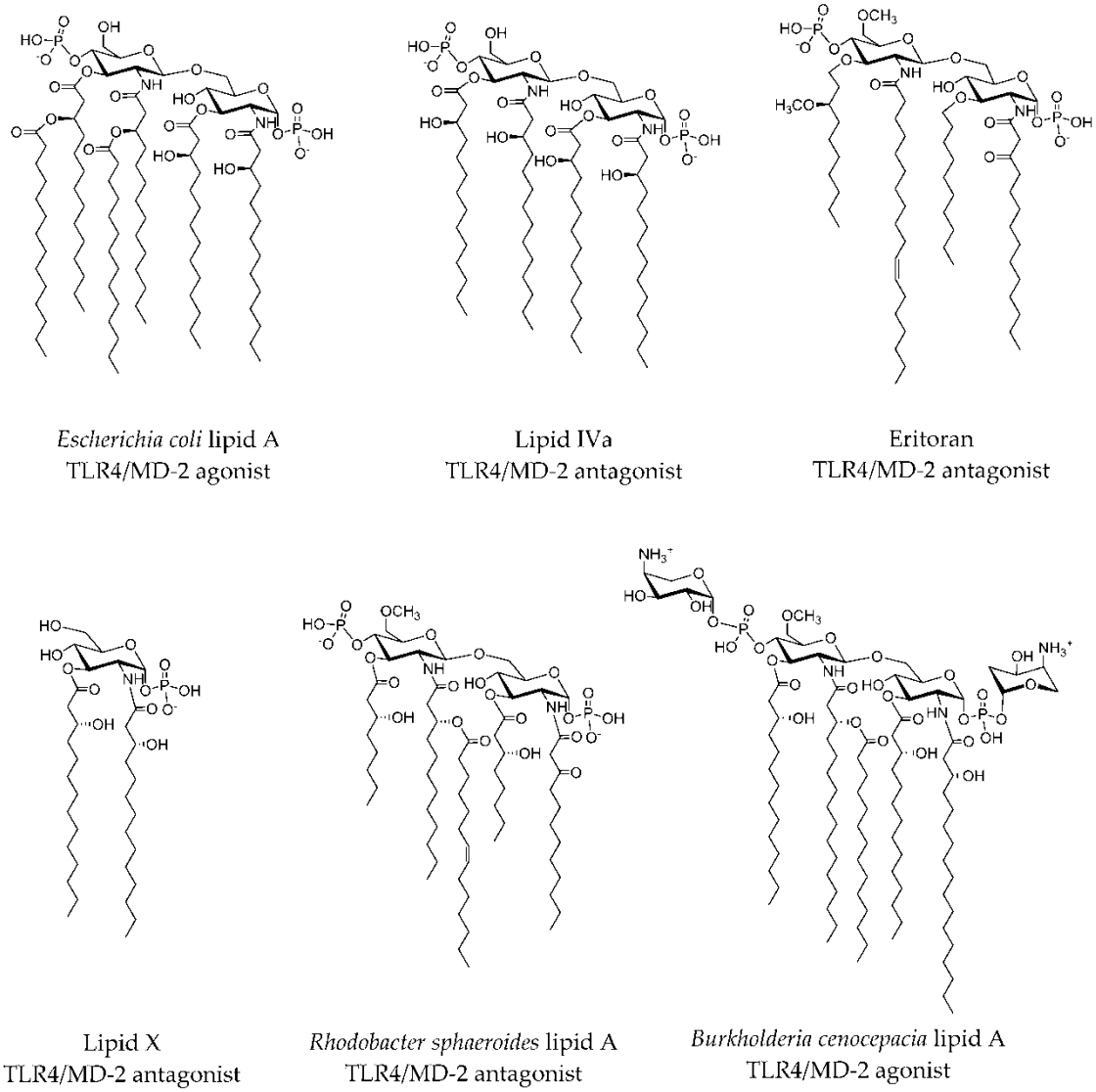


Figure 6: Lipid A and synthetic lipid A analogues with activity as TLR4 modulators. Activity is referred to hTMR4/MD-2. Extracted from [2].

1.2.1 TLR4 agonists and antagonists

TLR agonists have shown to improve anticancer vaccination protocols, and they are also the focal point for new vaccine development as non-infectious vaccines.[26] For example, the natural product monophosphoryl lipid A (MPLA,, a detoxified component of LPS from *Salmonella Minnesota* which contains the lipid A moiety that binds to TLR4/MD-2, is incorporated into several vaccines, including vaccines for Hepatitis B (Fendrix™), and cervical cancer (Cervarix™). It is also been applied in the immunotherapy against melanoma.[27]

Also synthetic TLR4 agonists have been designed and assayed. Compound E6020[28] have shown good adjuvant activity with antitumoral trastuzumab,[29] or enhancing vaccine efficacy.[30] Lipid A mimetics, such as the aminoalkyl glucosaminide phosphates, have been developed as TLR4 stimulants[31] with good adjuvant activity,[32] including the potent vaccine adjuvant RC-529,[33] and the bioisoster CRX-547, which has reduced toxicity in comparison to RC-529.[34] Small molecules pyrimido[5,4-b]indoles have shown to stimulate TLR4 and could potentially be used as adjuvants or immune modulators,[35] synthetic analogues of natural product euodenine A have exhibited potent and selective agonism towards TLR4,[36] and synthetic peptides to mimic the TLR4/LPS interaction have also been reported.[37]

The design of LPS mimetics with TLR4 antagonist activity is an emerging strategy for the treatment of sepsis,[38] combined with the challenge of obtaining good drug-like properties. Lipid IVa is an underacylated lipid A analogue with intriguing properties, being antagonist in human TLR4 but agonist in mouse. The tetraacylated synthetic compound Eritoran reached phase III in clinical trials, but failed to demonstrate sufficient efficacy in late stage human trials, although it has recently shown promising activity in preventing influenza induced acute lung injury, through a TLR4 antagonism mechanism.[39]

Simplified derivatives without the phosphate group have also been reported,[40] exploring the presence of a cation. Of merit, new glycolipids and benzylammonium lipids (for example, IAXO-102 and IAXO-103) are the first family derived from a monosaccharide core with effective TLR4 antagonist activity.[40b] Other synthetic lipid A analogues include, for example, compound D1,[41] and one lipid X mimetic[42] exhibiting a TLR4 antagonist mechanism by blocking the interaction of LPSs with both CD14 and MD-2 proteins.

Several small non LPS-like molecules with TLR4 antagonist activity have also been developed, such as ethyl 4-oxo-4-(oxazolidin-3-yl)-butenoate derivatives (OSL07), benzothiazole-based inhibitors, ethyl phenyl-sulfamoyl-cyclohexene-carboxylate derivatives (TAK-242 or resatorvid), and β-amino alcohol derivatives.[43] However, no successful progress was shown when reaching clinical phases (for example, in the case of compound OSL07). Examples of non-lipid TLR4 antagonists based on dendrimer architecture can also be found in the recent literature, showing that the presence of lipidic chains is not an absolute requirement for an MD-2 antagonist, and thus opening interesting opportunities for immunity modulation.[44]

1.2.2 Computational studies of TLR4 modulators

Several computational studies have been performed in order to clarify the binding mode of TLR4/MD-2 agonist and antagonist ligands. As introduced above, the unveiling of the molecular recognition process at atomic detail is one of the major challenges in TLR4/MD-2 modulation. Molecular modeling, docking studies and MD simulations have already provided relevant contributions about the ligand/receptor interactions with promising impact for rational drug design.

Synthetic LPS Mimetics

Inspired by the LPS structure, different ligands have been designed and synthesized.

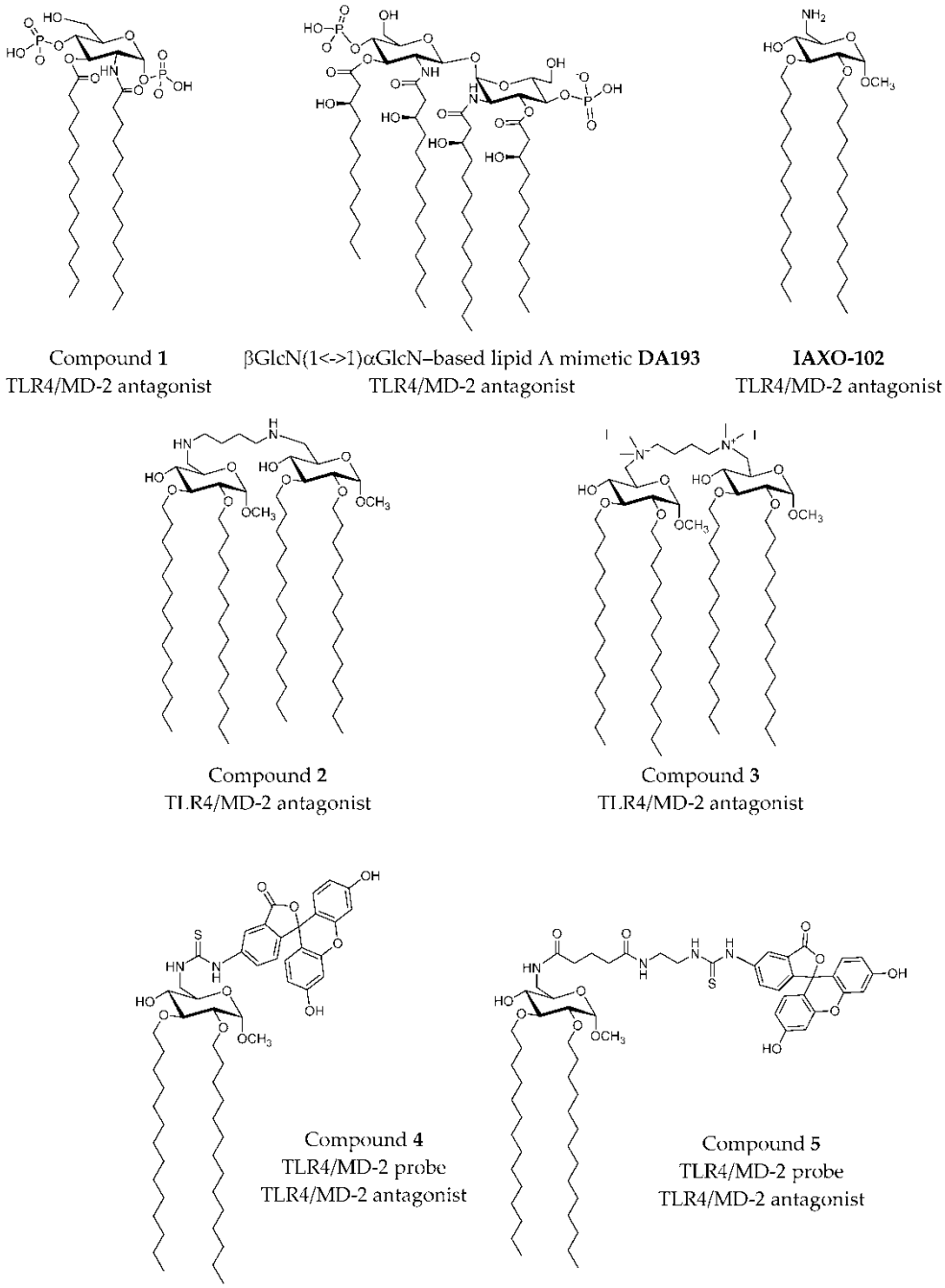


Figure 7: Synthetic LPS mimetics studied by computational approaches. Extracted from [26].

One of the first compounds to enter into clinical trials was Eritoran, a synthetic lipid A mimetic, potent TLR4 antagonist, which reached phase III clinical trials as an antisepsis agent, but failed since the study did not meet its primary endpoint of reduction in 28-day all-cause mortality in patients with severe sepsis[45]. Eritoran is a tetraacylated lipid A, the structural analogue of lipid A from RsLA, antagonist of human TLR4 and agonist of TLR4 from mouse and horse. In order to analyze the species-dependent activity of Eritoran, Scior *et al*. [46] built homology models by means of the SCWRL4 program.[47] Afterwards, docking of Eritoran was performed with AutoDock in order to determine the characteristics of the agonist/antagonist binding in the TLR4 structures from different species: human, mouse and horse. Some key amino acids were identified as relevant in species-specific binding: Lys58 (that corresponds to Asn in mouse and to Glu in horse), Lys388 (which is a Ser in mouse and a Lys in horse) and Gln436 (which is an Arg in mouse and a Gln in horse). The different pattern of interactions that are presented by these different residues impairs the TLR4-TLR4* bridging role of the ligand, thus preventing the effective dimerization and the agonist activity.

Modifications of the chemical structure of the lipid A scaffold have served as a starting point for the design of novel TLR4 modulators. One modification reported by Cighetti *et al*.[42] was the diphosphorylation of the scaffold of lipid X (Figure 7), a biosynthetic precursor of lipid A, leading to compound **1** , which has been found to be an antagonist in both human and mouse TLR4/MD-2. This compound was also shown to stimulate CD14 internalization in one-marrow-derived murine macrophages, thus demonstrating targeting of also CD14 in a TLR4-independent manner. In order to propose 3D models for the ligand recognition processes, computational studies were undertaken on both CD14 and MD-2 proteins. Docking calculations in MD-2 with AutoDock and AutoDock VINA,[48] followed by

MD simulations of the resulting complexes, led to the identification of two possible binding poses: the most stable one (in terms of predicted binding energy) allocated both FA chains inside the MD-2 binding pocket, mimicking the lipid IVa binding to MD-2 in the crystal structure (PDB ID 2E59). The MD-2/**1** complex is stabilized by hydrophobic interactions between its two FA chains and aliphatic and aromatic residues from the MD-2 pocket together with polar interactions at the rim of MD-2, involving mainly the phosphate groups and side chains from Ser118 and Arg96 residues and, in some cases, interactions between the amide CO or ester CO groups from compound **3** and the Ser120 OH group. This result was in agreement with NMR experiments performed by the authors that clearly showed FA chain-protein interactions. In a few cases, calculations predicted a second docked binding pose for compound **1** presenting only one FA chain inside the MD-2 hydrophobic pocket, while the second FA chain was lying over Ile124. Interestingly, in the agonist conformation, this residue has moved towards the inside of MD-2 and Phe126 occupies its place. This synchronism allows the agonist/antagonist switch. In the bound/unbound equilibrium, this alternative binding pose could co-exist with the first and most stable one. Cighetti *et al*.[42] also combined docking and MD simulations to propose a binding mode for compound **1** with CD14. CD14 is also characterized by having a wide lipophilic pocket, but with fewer polar residues at the rim. Compound **1** was predicted to bind with the saccharide moiety and the phosphate groups at the entrance of the CD14 hydrophobic cavity and with the FA chains inside the pocket, in agreement with the CD14 binding properties observed experimentally. In addition to its properties to prevent TLR4 signaling, Compound **1** has also been proposed as a promising hit as TLR4 modulator because of its favorable solubility properties and for its lack of toxicity according to the MTT tests.

Another strategy to mimic lipid A was the design of tetraacylated lipid A mimetics based on the βGlcN (1↔1) αGlcN scaffold analogue by substituting the β(1→6) with βα(1↔1) glycosidic linkage in order to confer rigidity to the molecule.[49] In particular, compound DA193 (Figure 7) resulted in being a dose-dependent antagonist in human and mouse, according to assays performed in HEK293 cells transiently transfected with membrane CD14 (*m*CD14)/*h*MD-2TLR4, HEK293 cells transfected with *h*MD-2TLR4 only and assays on human macrophage-like cell line (THP-1). In order to propose an atomistic understanding of the interactions between the ligand and the receptor, MD simulations were performed 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 the second one with the β-GlcN facing the Phe126 loop with an energy difference similar to that found for orientations of lipid A in the binding site of *h*MD-2. Dissociation constants, calculated from MD simulations of the MD-2/DA193 complex, estimated a binding to MD-2 20-fold stronger than lipid A and three-fold more than lipid IVa. It was concluded that the conformational rigidity of the βα(1↔1) diglucosamine backbone of these tetraacylated lipid A mimetics ensures strong binding to MD-2, in two possible binding poses, unlike the native lipid A structures.

The commercial TLR4 antagonist IAXO-102[50] has also served as inspiration for the rational design of TLR4 modulators and probes. Recently, the design and synthesis of IAXO-102-based TLR4 modulators were reported by Ciaramelli *et al*.[51] The design was based on a previous docked binding mode of IAXO-102 into MD-2 that showed the possibility to host two ligands simultaneously. A dimeric scaffold with two glycolipid units was designed by connecting both units through C4 diamino and di-ammonium linkers. Both compounds were confirmed to inhibit TLR4 activation and signaling in HEK-Blue™ cells expressing *h*TLR4 in a

concentration-dependent manner. Unfortunately, these compounds had a very poor solubility in aqueous solution.

The same IAXO-102 scaffold was used to design fluorescent probes[51]. 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. Normal mode analysis of CD14 was used to obtain motions and conformational changes, and docking calculations of both designed probes **4** and **5** were performed in three different conformations (Figure 7). Calculations predicted binding poses in which the fatty acid chains are buried inside the CD14 binding site (human and mouse) with the sugar located in the external portion. The thiourea linker and the fluorescein moiety established polar interactions with the hydrophilic rim without adopting a preferred pose. The best complexes were selected, according to their preferred docked poses, and were submitted to MD simulations and MM-GBSA analysis. In order to perform docking studies in the *h*TLR/MD-2, a hybrid *h*TLR4/MD-2 model in the antagonist conformation was modeled. As expected, predicted binding poses were similar to those found for CD-14, with the FA chains inside the pocket and the fluorescein moiety reaching TLR4 in the case of the longer probe. In fact, calculations of the SASA with CASTP[52] in both CD14 and MD-2 showed that both pockets have similar topologies and volumes. However, the presence of a lower number of polar residues in the rim of CD-14 allows it to recognize a wide range of microbial and cellular molecular determinants, such as lipopeptides to be transferred to TLR2. In contrast, MD-2 rim's polarity confers selectivity to the protein towards LPS.

Natural LPSs

Rhodobacter sphaeroides lipid A (RsLA,)[53] has five acyl chains, with one unsaturated and two shorter chains than *Escherichia coli* lipid *A (Figure 6).* It is an antagonist in

human and mouse, but an agonist in horse, although, intriguingly, 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 three 3D structures was undertaken. A homology model was built for horse and hamster TLR4/MD-2, with human and murine X-ray crystallographic structures as templates (PDB ID 3FXI and 2Z64).[54] The role of Arg385 had been proven in horse TLR4 complex activation by lipid IVa[55] through polar interactions between the guanidinium moiety and the phosphate group of lipid IVa. In fact, in other species, this residue is substituted by glycine in human and hamster and by an alanine in murine. The docked structure with AutoDock VINA of the horse TLR4/MD-2/RsLa complex closely resembled the pose of lipid IVa in murine crystal structure of TLR4/MD-2. On the contrary, the docked binding pose found in the hamster MD-2 was similar to the lipid IVa pose in the crystal structure from chicken (PDB ID 3MU3) and human. The difference between the species was mainly attributed to the different characteristic of each protein. By docking studies on *h*MD-2 with AutoDock, it has been observed that the longest chain of RsLPS could be accommodated in MD-2 by folding the chains itself as has been observed with the Eritoran fatty acid chains. The polar head (diglucosamine) is always exposed to the solvent.

Molecular modeling by Irvine *et al*. has also showed that the different human/horse TLR4 responses towards RsLA is related to two different amino acids, Gly384 and Ser441, in human TLR4 (Arg385 and Pro442 in horse). [56] Residue Arg385 in horse TLR4, although located around a 9 Å distance from the docked RsLA, could establish a long-range electrostatic interaction with a phosphate group of RsLA, while the Pro442 is situated near the dimerization interface with TLR4* and interacts with an FA chain of RsLA by van der Waals interactions. This hypothesis

was confirmed by experimental assays with transfected HEK293 cells with G384R/S441P *h*TLR4 with *eq*MD-2 and R385G/P442S eqTLR4 with *h*MD-2. It was observed that the R385G/P442S mutations in horse caused a complete loss of activity, and in human, the double mutant G384R/P441S TLR4 was unable to activate the signaling event. Since the double mutation did not recover the activity, other residues must be required. The docking of RsLA 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. RsLPS can adopt two orientations depending on the position of 1-PO_4 (oriented towards primary TLR4 in the case of horse and towards partner TLR4* in the case of human). This fact leads to different contacts between acyl chains of RsLPS and the hydrophobic pocket of MD-2. Moreover, superimposition of docked RsLA with X-ray crystallography poses of lipid A and lipid IVa showed that RsLA and lipid A acyl chains occupy more volume than lipid Iva, an, more importantly, the R2 chain of RsLA and lipid A protrudes from MD-2 and establishes interactions with the partner TLR4 in contrast to the R2 chain of lipid IVa, which is folded into the MD-2 pocket.

The severe pathogen *B. cenocepacia* LPS has been reported by Di Lorenzo *et al*. to strongly activate human TLR4/MD-2, despite the fact that its lipid A has only five acyl chains (Figure 6).[57] The Ara4N residues in lipid A have been shown to contribute to TLR4-lipid A interactions, and experiments in a mouse model of LPS-induced endotoxic shock confirmed the proinflammatory potential of *B. cenocepacia* penta-acylated lipid A. A combination of docking calculations and MD simulations, together with experimental mutagenesis of the TLR4/MD-2 interacting surfaces, suggested that the longer acyl chains allow reaching deeper regions inside the MD-2 pocket, thus compensating the absence of one FA chain and, at the

same time, allowing the exposure of the fifth FA chain on the surface of MD-2. This enables interactions with partner TLR4* and promotes its dimerization. The replacement of Val82 by Phe enhanced the inflammatory response, and it was related to the changes of van der Waals interactions into stronger CH–π interactions with the FA chain, longer than the corresponding one on *E. coli* LPS. The molecular model also showed that Ara4N residues provide additional polar interactions affecting the *B. cenocepacia* LPS binding to the TLR4/MD-2, and contribute to the anchoring of the lipid A into the receptor complex by 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 *Bordetella pertussis* lipid A, for example.[58] As described in Section 2.2, this model for the TLR4/MD-2/LPS_{BC} complex was used to generate a computational mutant TLR4/MD-2/LPS_{BC} complex (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.[23a] Altogether, these results provided a molecular model for the activation of the human TLR4/MD-2 complex by penta-acylated lipid A, which sheds some light onto the comprehension of the molecular recognition of LPS by TLR4/MD-2.

Non LPS-like modulators

The species-specific discrimination of TLR4 ligands by MD-2 is exemplified by taxanes, in particular paclitaxel, a proinflammatory murine TLR4/MD-2 ligand, which activates the subsequent inflammatory cytokine response (Figure 8).[59] Zimmer *et al*. demonstrated with different experiments that the activation of TLR4 by PTX requires the *m*MD-2 protein, being independent from TLR4 species.[60] This requirement is due to the electrostatic potential surfaces, hydrophobicity, binding

pocket size and the conformational gating of the 123–130 amino acids loop. *h*MD-2 and *m*MD-2 have a very large cavity volume that in principle allows lipid IVa, PTX and Eritoran to fit inside. The study of the electrostatic surfaces of *m*MD-2 (PDB ID 2Z64) and *h*MD-2 (PDB ID 2Z65) by means of SYBYL software[61] shows that the cavities for both structures are close to electroneutral, being *m*MD-2 more electronegative than *h*MD-2, especially in the Cys95–Cys105 loop, which is critical for the MD-2/TLR4 interaction. Furthermore, the electrostatic surface of *h*MD-2 displays three electropositive patches corresponding to Lys58, Lys122 and Lys125, which are absent on the *m*MD-2 surface.

Docking studies were performed with the help of the Glide program[62] by using the crystal structure of *h*MD-2 (PDB ID 2Z65) and *m*MD-2 (PDB ID 2Z64). 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 both aromatic groups. In addition, the Lys125 side chain establishes hydrophobic contact with the phenyl ring. Another key interaction (cation-π) is established between the phenyl group of PTX and the Lys122, which is the only different amino acid in the MD-2 species-conserved sequence Phe119–Gly123. In MD-2, the multiple interactions attract the Gly123–Lys130 loop so as to form a concave surface facing the docked PTX. The same loop in the mouse protein is oriented in the reverse direction. The presence of a Glu122 instead of the Lys122 in *m*MD-2 leads to a completely different binding pose, possibly due to the absence of the cation-π interaction.[60] Other work by Resman *et al*.[59c] proposed a similar binding mode for paclitaxel and the analogue docetaxel, on the basis of docking performed with AutoDock in *h*MD-2 (PDB ID 2E59). Also in this case, the most favorable docked binding poses of both taxanes oriented the benzoyl group towards the nearby region formed by Ile61, Phe76, Leu78, Phe119 and Phe151 of *h*MD-2

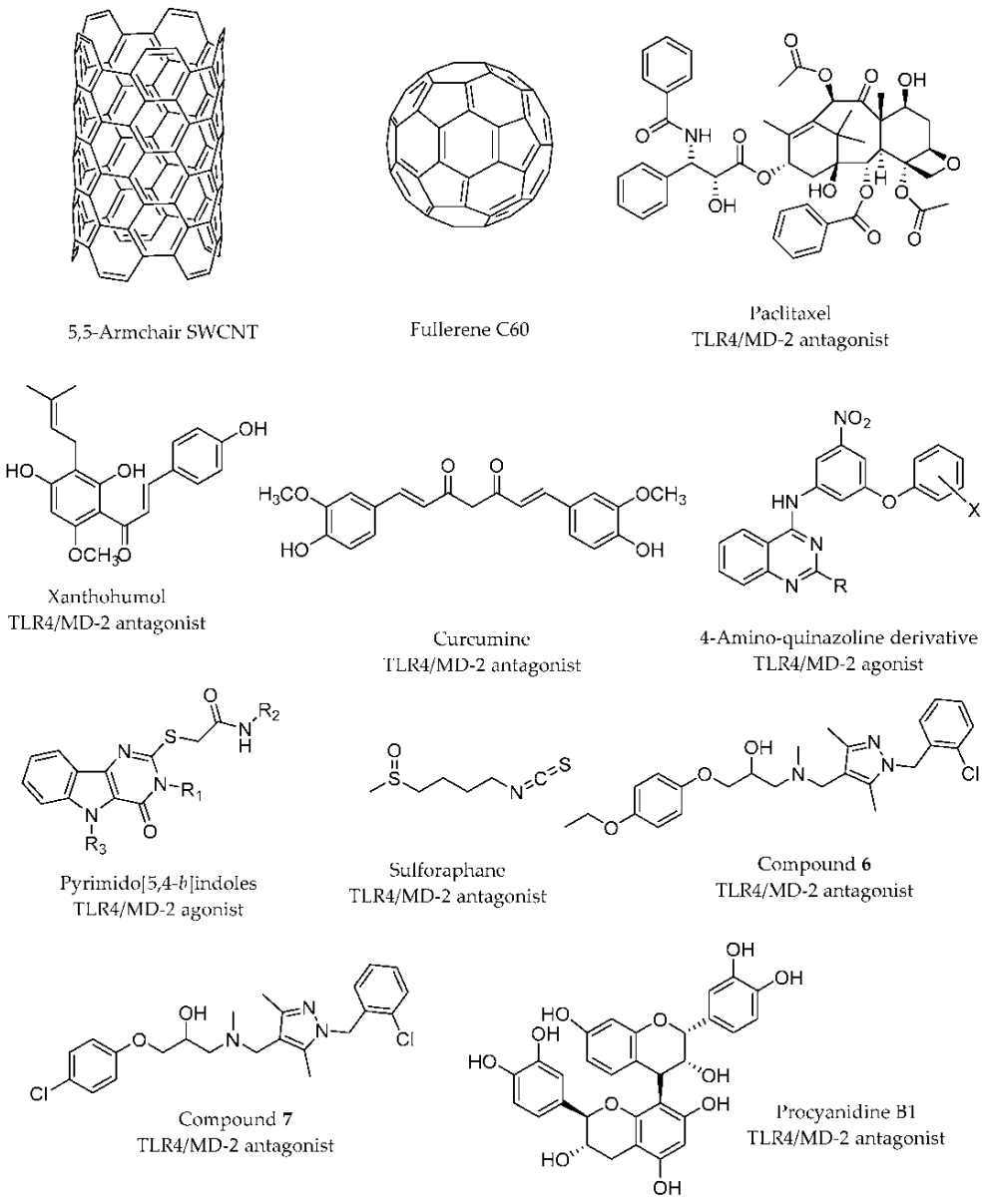


Figure 8: Non-LPS like TLR4 modulators studied by means of computational approaches. Extracted from [2].

A docked binding mode for a prenylated chalcone-type into the antagonist conformation of *h*MD-2 (PDB ID 2E59) has been proposed by Fu *et al*.[63] For this purpose, the Glide docking program was used, and the results highlighted the importance of the H-bonds between the OH groups present in the xanthohumol and residues Tyr102 and Arg90. Moreover, another H-bond between the OH of the phenolic group and Glu92 was identified from the docking studies, but this

interaction was rapidly broken during the subsequent MD simulation (50 ns), leading to a final MD-2/xanthohumol complex stabilized by the above-mentioned interactions. An analogue behavior was found for curcumin (Figure 8) after docking with AutoDock also in the same crystal structure of *h*MD-2 (PDB ID 2E59). The *h*MD-2/curcumin complex resulting from the docking was subjected to MD simulations leading to a stable complex with equivalent interactions with Tyr102 and Arg90. Accordingly, experimental studies with MD-2 mutants (MD-2R90A/Y102A) have pointed to a direct binding of curcumin to MD-2 in the same binding site as LPS. This ligand would occupy a large part of the hydrophobic pocket and form H-bonds with residues Arg90 and Tyr102, which were stable along the simulation trajectory. Analogously, the H-bond with Gly92 was broken during the simulation. In addition, MD simulations have revealed that the presence of the ligand stabilizes the complex. In particular, MD simulations of the apo-state and bound state of MD-2 have shown that, in the case of the apo-state, MD-2 suffers an important conformational change, reducing the volume of the cavity entrance, in agreement with other similar MD simulations performed on the apo MD-2,[64] whereas the bound MD-2 shows good stability.[65]

Cell-based high throughput screening (HTS) allowed the identification of novel chemical entities as potent NFκB activators as selective TLR4 ligands: substituted pyramid[5-4-*b*]indole derivatives[66] and 4-amino-quinazolines[67] From the former family, one hit compound was selected). A series of pyrimido[5,4-*b*]indole rings with carboxamides substituted with various alkyl, cycloalkyl, aromatic and heteroaromatic groups was synthesized and biologically tested in order to establish the SAR. One of the most active compounds was docked in the mouse TLR4/MD-2 system. The ligand was predicted to bind within the LPS-binding pocket forming H-bonds with residues Glu439(TLR4) and Arg90(MD-2), and

multiple hydrophobic interactions. This computational study supported that active compounds appeared to bind primarily to MD-2 in the TLR4/MD-2 complex.

From the second HTS, one 4-amino-quinazoline was identified with selective agonist activity for human TLR4/MD-2 rather than mouse.[67] The docking calculations of this 4-amino-quinazoline into TLR4/MD-2 showed that the ligand establishes hydrophobic interactions with Phe119-121-126 and Leu87 and makes H-bonds with the residues Gln436 and Glu439 of TLR4 and Arg90 of MD-2. Noteworthy is the interaction of the two polar nitro oxygens of the compound with the backbone nitrogens of Ile124 and Lys122 of the MD-2 protein. Moreover, the results from the computational study underlined the importance of the Lys122, which happens to be a glutamic acid in mouse. This could produce an electrostatic repulsion effect with the nitro group, thus justifying the decreased activity in *m*TLR4/MD-2.[67] Several analogues were synthesized to establish the basis for SAR, confirming the relevant role of the nitro group for the TLR binding and guiding further optimization of the lead compound.

It was shown by liquid chromatography-mass spectrometry analysis that the compound termed sulforaphane (SFN) forms a covalent bond with the residue Cys133 of *h*MD-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 (PDB ID 3FXI) and XA2 lipid chain of lipid IVa (*cf*. PDB ID 2E59). More precisely, 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 or settling inside the pocket.[68] The same

mechanism was reported for the caffeic acid phenethyl ester compound, using only experimental methods (thus, not reviewed herein).[69]

A series of compounds built by functionalizing pyrazole rings was reported by Bevan *et al*. [70] to inhibit TLR4 activation. Experimental studies indicated that two compounds were the lead inhibitors. Thus, 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 able to carry out its innate immunity role.

Polyphenol procyanidin B1 (Figure 8) has been shown to be able to regulate innate and adaptive immunity by, inter alia, impairing LPS-induced inflammatory responses in human monocytes.[71] In order to explain its mode of action at atomic level, the authors undertook experimental and docking studies.[72] They noted a high degree of similarity in terms of the interactions found in the predicted binding pose with the TLR4/MD-2 system when compared to the interactions established by LPS with TLR4/MD-2 in the crystal structure (PDB ID 3FXI). For instance, the phosphate group of LPS forms a hydrogen bond with Ser118 of MD-2 where procyanidin B1 is predicted to form a hydrogen bond with Ser120, which is in close proximity to Ser118. In turn, the binding mode proposed by the authors would suggest that procyanidin B1 impairs TLR4 signaling by successfully competing with LPS to bind to MD-2 inside the hydrophobic pocket.

1.2.3 Computational studies of the TLR4/MD2 mechanism

There are several studies in the literature focused on the extracellular domain of the TLR4/MD-2 complex, more specifically on its ability to recognize lipid A and lipid IVa. Garate *et al*.[73] report MD simulations of at least 11 ns of apo-MD-2, TLR4/MD-2 dimer, MD-2/lipid A, MD-2/lipid IVa and TLR4/MD-2/lipid A complexes; they conclude by highlighting the hydrophobicity of the MD-2 pocket and its ability to close promptly in an aqueous environment. According to the authors, the flexibility of the helix connecting MD-2 with TLR4 (helix H1) is essential for the observation of this behavior, and they propose a putative equilibrium between the open and the closed states of MD-2. MD-2 has been observed to close at a similar rate in the simulations where it was in the presence or in the absence of TLR4. However, in the former case, MD-2 was observed to fluctuate less due to the presence of TLR4, reducing the number of degrees of freedom. Another interesting conclusion by the authors is the key role that charged phosphates play in the early recognition of lipids with the corresponding impact on the formation of heterotetramers. 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 cooperativity in the binding process.

Evidence of the plasticity of MD-2 has also been observed by DeMarco *et al*. after several MD simulations performed in complex with variably-acylated lipid A molecules from *Escherichia coli* and *Neisseria meningitidis*. The results of these simulations (50 ns for each production run) led to the conclusion that the level of acylation of these ligands greatly influences the final architecture of the dimerization interface, leading to agonist or antagonist conformation of the TLR4/MD-2 system.

Paramo *et al*. performed long MD simulations of at least 100 ns on the TLR4/MD-2 system in complex with different ligands observing the Phe126 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. The dimerization interface between the two partners' heterodimers (TLR4/MD-2/TLR4*/MD-2*) was destabilized in agonist-free systems, especially due to the opening of the Phe126 switch, which disrupts the arrangement of nearby side chains from Leu87, Val82 and Met85 of MD-2. These simulations are in agreement with the NMR studies pointing at the re-orientation of the Phe126 aromatic side chain induced by the binding of hexa-acylated endotoxin. As can be observed, the lengths of the MD simulations range from relatively short to longer ones, depending on the origin (experimental vs. homology modeling) of the starting geometries, as well as the aim that is pursued: geometry optimization, study of the stability of the ligand-protein/protein-protein interactions, flexibility studies, binding free energy calculations, etc.

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*.[74] 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.

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.[53] 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[75] to compare the binding energy of a truncated MD-2 with the original one.[76] 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.*[77] *h*MD-2 has the ability to bind LPS in the absence of TLR4, while *m*MD-2 is responsive to LPS only when engaged in a complex with TLR4. Site-directed mutagenesis was applied to *h*MD-2 to mutate Val135 to its murine alanine counterpart. This single point mutation led to a mutant V135A *h*MD-2 lacking the ability to bind LPS. A series of 50-ns MD simulations of the WT *h*MD-2 and the V135A mutant *h*MD-2 in solution and in complex with TLR4 was performed to study the conformational changes. In the case of the WT *h*MD-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 trend was confirmed by

another simulation of MD-2 in complex with three myristic acids (as reported in PDB ID 2E26). The V135A mutant *h*MD-2 needed a much longer simulation time to adapt its shape to the three

1.2.4 Computational studies of the TLR4 intracellular domain

The intracellular domain of the TLR4 transmembrane protein contains a Toll/interleukin-1 receptor (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 (Figure 9) .As a major component of theses adaptors, the TIR domain is believed to play a central role in the recruitment processes.[14, 78]

The crystal structures of human TLR1 (PDB ID 1FYV) and TLR2 (PDB ID 1FYW) revealed the structural basis of the TIR domain[79] followed by the crystal structure of TLR10 TIR domain (PDB ID 2J67)[80] and the solution structure of MyD88 TIR domain resolved by NMR (PDB ID 2JS7 and 2Z5V).[81] 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.[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*.[84] built monomer models of TLR4, Mal and MyD88 using comparative modeling and loop refining techniques. They noted differences in the electrostatic surface potentials suggesting that adaptor binding is driven by electrostatic complementarity (Figure 10).

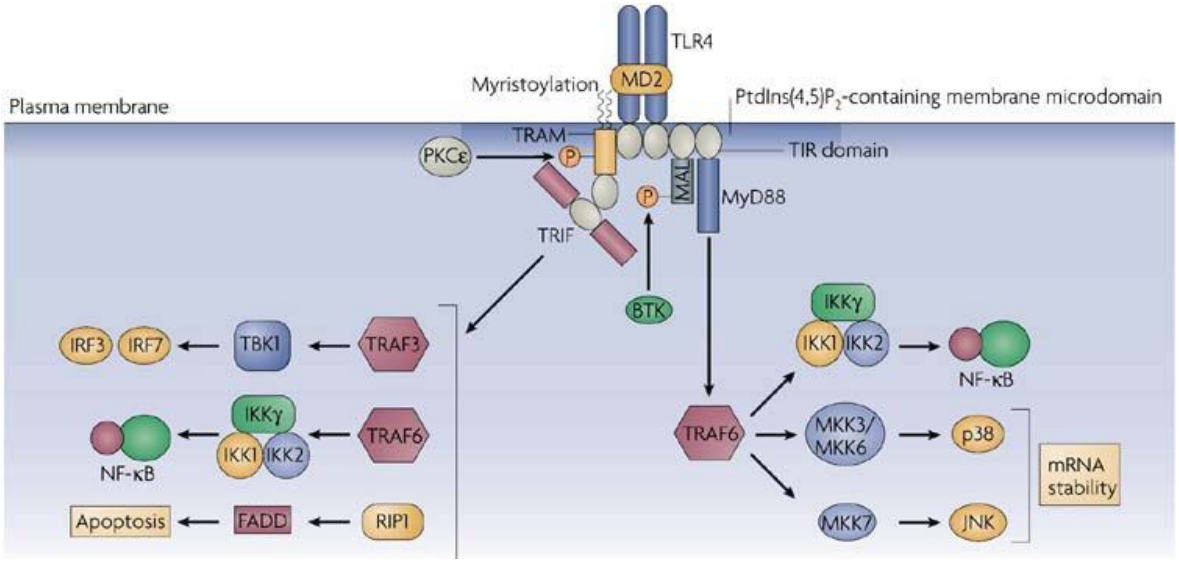


Figure 9: MyD88 and TRAM signalling pathways involved in TLR4 activation. Extracted from [85].

This point was also emphasized in a study by Kubarenko *et al*.[86] in which they compared the surface charges of TIR domains of the crystal structure of *h*TLR2 and of the models of *h*TLR3 and *h*TLR4 and noted that the surface charge distribution of the BB loop and the αC-helix present similarities in TLR2 and TLR4 and differ between TLR3 and TLR4. The authors considered that these findings could explain why TLR2 and TLR4 recruit MyD88, whereas TLR3 does not. In the computational study by Gong *et al*.,[87] it was highlighted that, whereas the BB-loop is highly conserved among TIR-domains, the APBS electrostatic surfaces differ. 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 hyporesponsive to LPS stimulation. The authors noted that their data do not suggest a direct role for this residue.

Dunne *et al*.[84] used a docking procedure based on hydrophobicity and geometry. Their results suggest that Mal and MyD88 bind at two distinct binding sites (non-overlapping): 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. 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*.[88] 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. 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*.[87] 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. 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. In another study, Basith *et al*.[89] 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-dependant signaling pathway. The authors reported a TLR4-TLR4 homodimer model,[88] 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.

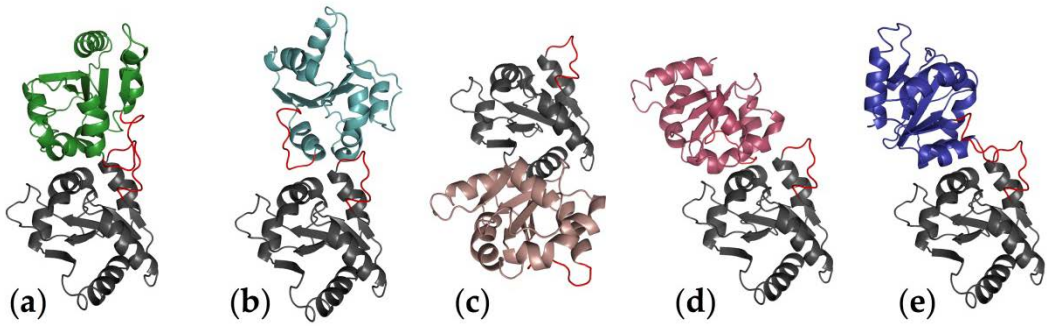


*Figure 10: Representation of the different ways the dimer is proposed by published computational strategies to be assembled in the literature by computational strategies. (**a**) First reported by Miguel et al.[88]; (**b**) reported by Gong et al.[87]; (**c**-**e**) reported by Guven-Maiorov et al.[90]*

In a later study, Bovijn *et al*.[91] reported a homology model constructed based on the crystal structure of the dimeric TLR10 TIR domain. This model is also in agreement with the first model reported by Miguel *et al*.[88] 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. Then, they describe that the binding site for TRAM and Mal is formed by the reunion of two sites (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*.[88] 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 binding site I and might be implicated in the interferon regulatory factor 3 (IRF-3) activation.

Singh *et al*.[92] 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. 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*.,[93] the authors used computational techniques to describe the architecture of the

signalosome of TLR4. They built three models of the intracellular part of the TLR4 protein. 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 to propose different binding modes for Mal.

1.3 Sugars and proteins

The fundamental dogma of molecular biology establishes that the information flux in a biological system starts in the DNA and ends in proteins, via the RNA. Thus, study fields that emerge from this affirmation are three: genomics (study of the DNA), transcriptomics (study of the RNA) and proteomics (study of proteins). However, this view is not complete inasmuch as step by step we realize, the two forgotten in the dogma, lipids and carbohydrates, play a fundamental role in the correct interpretation of the cell signaling. Due to the increasing importance these two cellular components, the lipidomics (study of lipids) and glycomics (study of carbohydrates), are getting, they are making their way to consolidate themselves as a study field, together with the other three, of the general interest in the scientific community.[94]

Carbohydrates are very abundant in nature and they key players in numerous biological roles. They are the fundamental source of metabolic energy. In fact they represent an important element in our diet; some insoluble carbohydrate polymers serve to lubricate skeletal joints and provide adhesion between cells and others serve as structural and protective elements in the cell walls of bacteria and plants and in the connective tissues and cell coats of animals. Sugars can exist in nature as single entities or as a forming part of glycoconjugates, essentially glycolipids and glycoproteins. Chemically, carbohydrates are polyhydroxy aldehydes or ketones,

and their chemical formula is usually $C_n(H_2O)_n$ (carbon "hydrates"), but some "natural" carbohydrates can also contain nitrogen, phosphorus, sulphur. Their chemical variability is enormous, due to the possibilities of branching at either hydroxyl group as well as the possible presence of pendant substituents, either neutral or charged.[95] In turn, this variability generates a huge possibility of constitutional isomers and a plethora of conformations. Moreover, their chemical properties in terms of the presence of polar and non-polar patches make them to exhibit peculiar stability and, acting as ligands for receptors, to be fairly adaptable to a variety of environments. They may adopt different shapes and display rather distinct conformational and dynamic properties. Glycan diversity depends on the monosaccharides (basic unit of saccharide) which are made by, on the linkages that connect monosaccharides (glycosidic linkage), and on other factors such as the *anomeric effects*, the orientation of all the torsional angles,[96] and the intramolecular hydrogen bonding between adjacent OH groups.[97] Therefore, it is of paramount importance to predict and characterize their conformational and structural properties in a systematic manner. Computational techniques have proven to be essential in this aspect.[98]

1.3.1 Sugars

Monosaccharides

Carbohydrates are made by small constituents, named monosaccharides (or simply sugars). Monosaccharides are composed of a single polyhydroxy aldehyde or ketone unit and may have different number of carbons normally from five to nine. The most abundant monosaccharide in nature is D-glucose. Usually the carbons of a monosaccharide have hydroxyl groups attached and they are normally stereogenic centres.

Monosaccharides of six carbons (called pyranoses) usually adopt chair conformations and the substituents can be found either in the axial or equatorial position. In pyranose sugars, if there is a hydroxyl group at the anomeric carbon (the hemiacetal/acetal carbon) the monosaccharide can adopt an α or a β configuration. Both α and β isomers are optically active and usually present similar physical properties. If the substituent on the anomeric carbon is electronegative (like a methoxy group) there is a preference of the substituent to adopt an axial configuration rather than an equatorial one. This phenomenon is called *anomeric effect*, which is countered by steric factors, reason why D-sugars present more stability as α isomer.[99] If the hydroxyl group is attached to another carbon which is not the anomeric one, leads to a different monosaccharide (epimer) with completely different properties. For example, glucose and mannose differ on the position of the OH group in position 2 (in equatorial position in glucose and in axial in mannose) and glucose and galactose in the OH of position 4 (in equatorial position in glucose and in axial in galactose) (Figure 11).

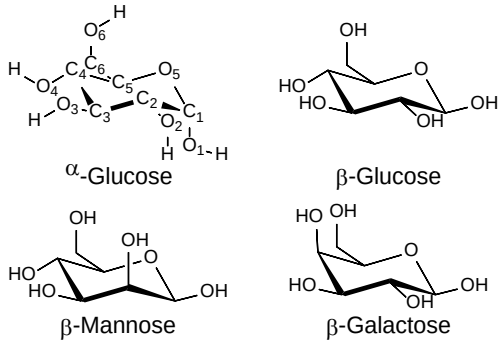


Figure 11: α and β configuration of glucose (top) and epimers of the β-glucose on C2 (mannose) and C4 (galactose)(bottom).

Moreover, simple monosaccharides may adopt distinct three-dimensional shapes. The most common shape for a pyranose is the chair conformation that minimizes steric congestion and includes the *anomeric effect*. However, also other shapes are

allowed, depending on the orientation, nature and number of substituents. The most prevalent alternative shapes for a pyranose are the boat, the skew and the envelope.

Regarding the conformation of the dihedral angle (ω) around C5-C6 in a pyranose ring (ω formed by O5-C5-C6-O6 according to the X-ray definition or by H5-C5-C6-O6 according to the NMR definition), a combination of solvation, steric, and stereoelectronic (*gauche effect*) effects drive the observed geometries. It has been monitored that for glucose-type sugars, a combination of *gauche-gauche* (*gg*) and *gauche-trans* (*gt*) rotamers coexists, while for galactose-type, ω prefers *gauche-trans* (*gt*) and *trans-gauche* (*tg*) geometries. This experimental behavior has also been validated using computational protocols.[100]

Disaccharides

Monosaccharides are attached among them through the so-called glycosidic linkage. The anomeric oxygen of one particular monosaccharide is attached to a secondary (or primary) hydroxyl group from the second sugar moiety to build one disaccharide (Figure 12). The glycosidic linkage is an acetal linkage with particular chemical properties and conformational features. For example, it is sensitive to strong acid and basic conditions.

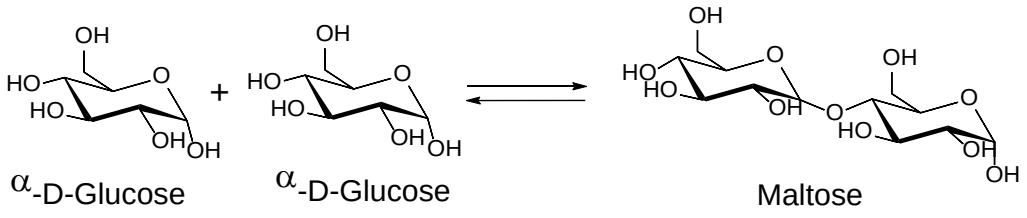


Figure 12: Formation of a Maltose unit from two glucose units.

If the substituent at the anomeric carbon presents lone pair electrons, for example an alkoxy group, retrodonation can occur from the exocyclic oxygen to the sigma

anti-bonding orbital of the intracyclic bond C-O. This effect is known as the *exo-anomeric effect*.[101] In this case, the lone-pair orbital has to be anti-periplanar to the antibonding orbital of C-O and, consequently, the alkyl substituent at the glycosidic oxygen adopts a *syn*-type orientation. The molecular orbital theory is able to predict also the modification of the bond length. In fact, for equatorially substituted pyranosides, the intracyclic C-O bond is longer than the exocyclic C-O bond.

Actually, these preferences lead to a *syn*-type conformation, although it has been observed that a minor proportion of *anti*- conformation may co-exist in β-glycosides (e.g. in 1→2, 1→3 and 1→4 linkages) provided that the contiguous OH-2 displays an equatorial orientation (i.e., glucose or galactose). The *anti*-type geometry is also favoured by the *exo-anomeric effect*. Combined modelling/Nuclear Magnetic Resonance (NMR) approaches are mandatory to characterize the glycosidic linkage of a particular disaccharide or oligo-saccharide (either α and β, either 1→2, 1→3, 1→4, or 1→6).[100, 102] All the particular features mentioned above for saccharides, besides their intrinsic dynamic properties require the synergy combination of rigorous experimental and theoretical protocols.[103]

Oligosaccharides and polysaccharides

Oligosaccharides are composed of a small number of monosaccharides. Their conformational properties, adaptability and recognition features have been, and still are, deeply analyzed.[104] Chemical methods have been employed to provide stable chemical analogues of oligosaccharides, with increased resistance to glycosyl hydrolases. Typically, either the glyosidic (inter-unit) or the endocyclic oxygen within the ring is substituted by other chemical element. Therefore, *C*-,[105] *S*-,[106] *Se*-,[107] and *N*-[108] glycosyl analogues have been synthesized to provide chemical probes for interaction studies or to prepare enzyme inhibitors. Given the different

chemical properties of these analogues, their conformational properties have been carefully analyzed and compared to those of their parent *O*-glycosides.[109] Computational studies have been directed towards the analysis of the flexibility of these new linked glycosides, and the preferences of the binding proteins to select the proper "binding pose" from the conformational assemble, not always being the most stable conformer in terms of potential energy.[110]

1.3.2 Proteins

Proteins that bind carbohydrates are called GBPs (Glycan Binding Protein) and in this group lectins, receptors, toxins, microbial adhesins, antibodies and enzymes, among others[111] are included. In the framework of cellular recognition, non-enzymatic proteins are the most important proteins that bind carbohydrates without changing its chemical structure. Between them, lectins are widely distributed in nature since they can be found in almost all living organisms. The recognition process between a lectin and a carbohydrate usually occurs through: A) hydrophobic interactions, such as CH-π interaction (between the sugar moiety and residues such as tryptophan, tyrosine and phenylalanine); B) a net of hydrogen bonds between the OH groups of the sugar and the polar groups of the protein (frequently, also water molecules are involved in this net); and C) metal coordination (usually with calcium or manganese) present in the binding site of the lectins (Figure 13).

These characteristic interactions occur in a specific region of the protein normally named Carbohydrate Recognition Domain (CRD). Unconventional sugars binding sites have been observed, as for the *C*-type lectin-like receptor 2 (CLEC-2).[112] The specificity of a lectin to recognize selectively some epitopes is due to the diversity of the CRD. Actually lectins have been classified based on their CRD, the sugar specificity and their expression patterns. In particular, human lectins are divided

into two classes: *C*-type (or calcium-dependent) lectins and S-type (sulfhydryl-dependent or calcium-independent) lectins. Their main functions are the cellular recognition, cellular adhesion regulation, as well as immunological functions such as recognition of sugars from pathogen organisms.[113] Lectins also play an essential role in the infection process. One of the most studied players in this event is the Dendritic-Cell Specific ICAM-3 Grabbing Nonintegrin (DC-SIGN). DC-SIGN is a *C*-type lectin that has been demonstrated to spread and evade the immune system by the HIV virus. Fittingly, this lectin seems to be involved in several infections by pathogens, such as *Ebola* virus and *Mycobacterium tuberculosis*.[114]

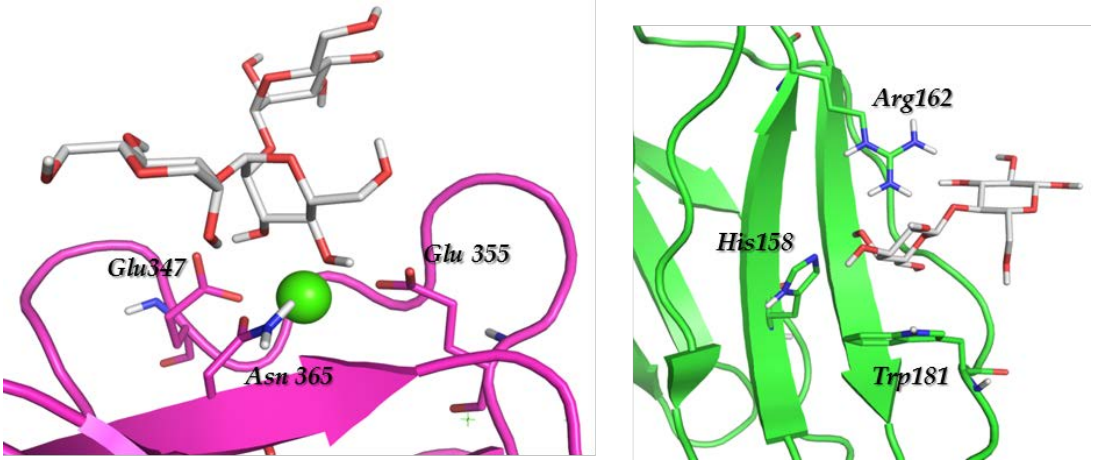


Figure 13: Examples of carbohydrate-protein interactions. Left) DC-SIGN in complex with a trimannoside, which coordinates through a calcium atom. Right) Galectin-3 in complex with lactose. Glucose moiety interacts with a Histidine and an Arginine through hydrogen bonds and with a Tryptophan through a CH-π interaction.

Plant lectins are accumulated in seeds or other tissues and mainly have a storage function but also they seem to play an important role in plant defence. Twelve families of lectins have been identified, although they commonly share some features. Plant lectins are usually found as dimers or tetramers, in which each monomer has a CRD with two metal binding sites for calcium and manganese. In the dimer there is a side-side protein-protein interaction forming 12-strandeed sheet. Plant lectins are very interesting because due to their easy availability are

extensively used as spy molecules due to their high glycans specificity. For example, it is very well known that Pea lectin and *Concanavalin A* selectively bind mannose or glucose, while *Maackia amurensis* lectin binds sialic acids. To date, more than 132 crystal structures of legume lectins from 18 different plants are available from the Protein Data Bank (PDB).[115]

Galectins

A special type of lectins evolutionary-conserved[116] is galectins (abbreviation of galactose binding lectins). Galectins constitute a family of β-D-galactoside binding proteins. They are localized in cytoplasm, nucleus, cell surface, and extracellular matrix,[117] and present a S-type sequence motif. To date 15 members of this family have been identified in mammals (gal-1 to -15), but only 10 of them are present in humans (gal-1, -2, -3, -4, -7, -8, -9, -10, -12, -13).[118] They are soluble proteins with a molecular weight around 14-36 kDa[119] and they lack of enzymatic activity.[96b] Among the physiological processes in which galectins have an important role, there are regulation of the immune response, cell cycle, cell growth and apoptosis.[120] Due to these functions, an alteration in these proteins can trigger various pathological processes, depending on which galectin is affected. Thus, it has been demonstrated that galectins are involved in processes like cancer, metastasis, inflammation, hypersensitivity and atherosclerosis, among other pathologies.[121] Therefore, in the last years the interest in these proteins as therapeutic targets has increased since the development of selective galectins modulators could improve the prognosis of many diseases.[122]

Based on the structural features, mammalian galectins have been classified in proto, tandem-repeat, and chimera types. Proto-type galectins (gal-1, -2, -5, -7, -10, -11, -13, -14, and -15) are usually dimers containing only one type of CRD (homodimers) non-covalently joined (Figure 13). Tandem-repeat-type galectins

(gal-4, -6, -8, -9, and -12) have two different CRDs, almost homologous, that are covalently joined. Chimera-type galectin (only gal-3) has a carboxyl-terminal CRD associated to an amino-terminal peptide which is rich in tyrosine, proline and glycine (collagen-like sequence),[123] forming a tail that intervenes in the oligomerization of this unique galectin.[124]

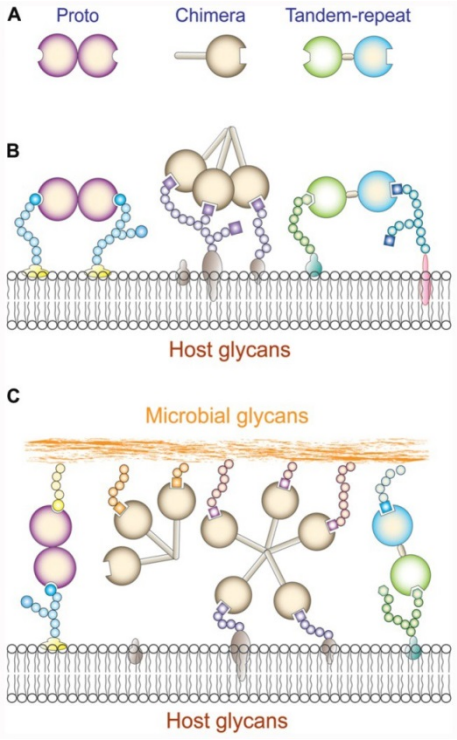


Figure 14: Schematic representation of the three galectin groups (prototype, tandem repeat and chimera) and their role in the recognition of microbial glycans. Extracted from [125].

The activity of galectins is ligand-concentration dependent and it can be both stimulatory and inhibitory.[126] These proteins recognize carbohydrates through the CRD, which consists of 130 amino acids approximately,[127] and it folds into a β-sandwich structure comprising two anti-parallel β-sheets (the F-sheet and S-sheet).[128] The CRD that is different among the galectins family, establishes a net of interactions selectively with the β-galactoside moiety through the formation of hydrogen bonds and hydrophobic interactions with tryptophan, histidine), and arginine residues.[129] Apart from the CRD, galectins differ in other aspects, such as

chain length or number of domains in their structure. To date, more than one hundred X-ray structures of human galectins are available at the PDB.[115] For the studies performed in this Thesis, we selected one or two X-ray crystallographic structures for each human galectin (PDB ID codes are shown in Table 1.2).

DC-SIGN

Dendritic Cell-Specific Intercellular adhesion molecule-3 grabbing non-integrin (DC-SIGN) is a type II membrane receptor that was characterized initially as a receptor for the HIV glycoprotein gp120. It was some years later when DC-SIGN was postulated as a cell-adhesion receptor, which has an important role in the infection by HIV.[130] DC-SIGN is expressed in dendritic cell, macrophages that infiltrate tumors and antigen presenting cells of tissues like the placenta.

The main function of DC-SIGN is the recognition of glycans through its CRD. This CRD is a C-type lectin like domain that binds fucose and mannose and has two Ca2+ coordination sites. Thus, DC-SIGN recognizes fucosylated glycans, like blood-type Lewis antigens, and high mannose glycans. The relatively broad specificity confers DC-SIGN to recognize a huge variety of ligands, both pathogenic (e.g. from HIV, *Micobacterium tuberculosis, Candida albicans* and Ebola virus) and self-glycoproteins, such as intercellular adhesion molecule and the Fc portion of immunoglobulins.[131] Furthermore, the supramolecular organization of DC-SIGN enhances its pathogen-recognition capacity. It is assembled as a tetramer through the interaction of its α-helical neck domain into a four-stranded bundle. This configuration achieves two goals: it projects the CRD about 35nm away from the membrane, and clusters the four CRDs together to facilitate the interactions with the ligand.[132]

DC-SIGN mediates the receptor-mediated endocytosis upon ligand binding. This internalization promotes and efficient antigen processing and presentation to

MHC class II.[133] Interestingly, DC-SIGN can also deliver its cargo to an uncharacterized cross-presentation route,[134] which makes this protein an interesting candidate for the development of cancer vaccines as activator of the CD8+ T cell responses.[135] Moreover, DC-SIGN-dependent cross-presentation is enhanced by the simultaneous triggering of TLRs. It has been demonstrated that the activation of TLR4 promotes the translocation of the cargo of activated DC-SIGN to the cytosol, leading to an increased CD8+ T cell activation.[136]

DC-SIGN also behaves as a signaling receptor. *In vitro* experiments have demonstrated that the triggering of DC-SIGN alone does not initiate changes in DC activation status or cytokine secretion, but can shape immune responses by modulating the signaling elicited by other pattern-recognition receptors. Interaction of DC-SIGN with mannose-containing ligands results in the recruitment of the upstream effector Rho guanine nucleotide exchange factor 12 (LARG) and Ras homolog family member A (RhoA). This recruitment yields to the increase in the transcription rate of genes like IL12 and IL6. On the contrary, fucose-containing DC-SIGN ligands suppress the production of pro inflammatory cytokines.[137] The capacity of DC-SIGN to discriminate among its multiple ligands to modulate the signaling of TLRs into a pro- or an anti-inflammatory response is unique among the C-type lectin family.[138] Mannose glycans are present in higher mammals only in the endoplasmic reticulum so, the exposure of these molecules on glycoproteins or in the extracellular space might be a sign for DC-SIGN of cell damage or invasion thus leading to the expression of proinflammatory cytokines.[139]

1.4 Objectives

The main objective of this thesis is to study by means of computational techniques the molecular recognition processes of receptors involved in the innate immunity. More concretely, on one hand, we will focus in two different types of lectins, Galectins and DC-SIGN, and on the other hand, we will focus in Toll-like receptor 4. The studies regarding these proteins will be carried out by addressing the following specific objectives along the following chapters:

Chapter 3. We propose the computational design of novel binders with improved affinity and selectivity towards galectins 1, 3 and 7, main players in tumoral events. By means of fragment-based virtual screening protocols, the vicinal pocket close to the carbohydrate recognition domain will be explored to identify appropriate moieties able to be selectively anchored to these pockets. We aim the design of modified lactose derivatives able to bind the galectins with improved affinity and with selectivity towards the different galectin subtypes 1, 3 and 7. Selected candidates will be synthesized by collaborators and their affinity and selectivity will be measured by biophysical techniques.

Chapter 4. We aim the design of new glycomimetics as DC-SIGN antagonists with high affinity and selectivity. A fragment-based virtual screening strategy will be used to explore adjacent pockets near the mannose binding site as accessible pockets to be anchored. Designed compounds will be docked and their binding mode studied by MD simulations. A selection of compounds will be synthesized to be tested and studied by biophysical techniques.

Chapter 5. This chapter will be devoted to the study of the molecular recognition process of several TLR4 modulators and their binding mode at atomic level. By means of docking and MD simulation, we will propose binding modes for three

reported peptides: peptides RS01 and RS09 with agonist activity, and MDMP peptide with antagonist activity. Moreover, we will study and propose the binding mode of cardiolipin to TLR4 to explain the synergistic effect of this compound in combination with LPS. Finally, the binding mode of a tetracylated and pentacylated *Bacteroides vulgatus* LPS accounting for the observed weak agonist activity will be proposed.

Chapter 6. We aim the discovery of non LPS-like compounds with activity as TLR4 modulators. We will develop a virtual screening strategy to be applied from public, commercial and *in-house* libraries, followed by re-docking and biological assays, in order to identify novel TLR4 modulators with a non LPS-related chemical structure.

Chapter 7. We will deepen in the knowledge of the structure of the full TLR4/MD-2 heterodimer system in order to gain insights into its functioning for drug design purposes. To do so, we will build a complete heterodimer including the extracellular, transmembrane and intracellular domains, including the membrane environment. We will model the dimerization of the transmembrane and intracellular domains. Also the binding mode of the MAL intracellular adaptor protein will be studied

Bibliography

1. Iwasaki, A.; Medzhitov, R., Regulation of adaptive immunity by the innate immune system. *Science* **2010,** *327*, 291–295.
2. (a) Monie, T. P.; Bryant, C. E.; Gay, N. J., Activating immunity: lessons from the TLRs and NLRs. *Trends Biochem. Sci.* **2009,** *34* (11), 553-561; (b) Rakoff-Nahoum, S.; Medzhitov, R., Toll-like receptors and cancer. *Nat. Rev. Cancer* **2009,** *9*, 57-63.
3. www.nobelprize.org/nobel_prizes/medicine/laureates/2011/ The Official Website of the Nobel Prize.
4. Hennessy, E. J.; Parker, A. E.; O'Neill, A. J., Targeting of Toll-like receptors: Emerging therapeutics? *Nat. Rev. Drug Discov.* **2010,** *9*, 293–307.
5. Medzhitov, R.; Preston-Hurlburt, P.; Janeway, C. A., A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **1997,** *388*, 394–397.
6. Janeway, C. A., Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **1989,** *54*, 1–13.
7. 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*.
8. (a) Tajalli-Nezhad, S.; Karimian, M.; Beyer, C.; Atlasi, M. A.; Azami Tameh, A., The regulatory role of Toll-like receptors after ischemic stroke: neurosteroids as TLR modulators with the focus on TLR2/4. *Cellular and molecular life sciences : CMLS* **2018**; (b) Jurewicz, M.; Takakura, A.; Augello, A.; Movahedi Naini, S.; Ichimura, T.; Zandi-Nejad, K.; Abdi, R., Ischemic injury enhances dendritic cell immunogenicity via TLR4 and NF-kappa B activation. *J. Immunol.* **2010,** *184* (6), 2939-48; (c) Montecucco, F.; Braunersreuther, V.; Burger, F.; Lenglet, S.; Pelli, G.; Carbone, F.; Fraga-Silva, R.; Stergiopulos, N.; Monaco, C.; Mueller, C.; Pagano, S.; Dallegri, F.; Mach, F.; Vuilleumier, N., Anti-apoA-1 auto-antibodies increase mouse atherosclerotic plaque vulnerability, myocardial necrosis and mortality triggering TLR2 and TLR4. *Thrombosis and haemostasis* **2015,** *114* (2), 410-22.
9. (a) Zolezzi, J. M.; Inestrosa, N. C., Wnt/TLR Dialog in Neuroinflammation, Relevance in Alzheimer's Disease. *Front. Immunol.* **2017,** *8*, 187; (b) Xie, X.; Shi, X.; Liu, M., The Roles of TLR Gene Polymorphisms in Atherosclerosis: A Systematic Review and Meta-Analysis of 35,317 Subjects. *Scandinavian journal of immunology* **2017,** *86* (1), 50-58.
10. O'Neill, L. A., Therapeutic targeting of Toll-like receptors for inflammatory and infectious diseases. *Curr. Opin. Pharmacol.* **2003,** *3* (4), 396-403.
11. Kaisho, T.; Akira, S., Toll-like receptor function and signaling. *J. Allergy Clin.* **2006,** *117*, 979-987.
12. Kobe, B.; Kajava, A. V., The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* **2001,** *11*, 725–732.
13. (a) Jin, M. S.; Kim, S. E.; Heo, J. Y.; Lee, M. E.; Kim, H. M.; Paik, S. G.; Lee, H.; Lee, J. O., Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **2007,** *130*, 1071–1082; (b) Farhat, K.; Riekenberg, S.; Heine, H.; Debarry, J.; Lang, R.; Mages, J.; Buwitt-Beckmann, U.; Roschmann, K.; Jung, G.; Wiesmuller, K. H.;

Ulmer, A. J., Heterodimerization of TLR2 with TLR1 or TLR6 expands the ligand spectrum but does not lead to differential signaling. *Journal of leukocyte biology* **2008,** *83* (3), 692-701.
14. O'Neill, L. A.; Bowie, A. G., The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Reviews: Immunology* **2007,** *7* (5), 353-64.
15. 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.
16. Botos, I.; Segal, D. M.; Davies, D. R., The Structural Biology of Toll-like Receptors. *Structure* **2011,** *19* (4), 447-459.
17. 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.
18. (a) Akira, S.; Uematsu, S.; Takeuchi, O., Pathogen recognition and innate immunity. *Cell* **2006,** *124* (4), 783-801; (b) Kawai, T.; Akira, S., Signaling to NF-κB by Toll-like receptors. *Trends in molecular medicine* **2007,** *13* (11), 460-469; (c) Lu, Y.-C.; Yeh, W.-C.; Ohashi, P. S., LPS/TLR4 signal transduction pathway. *Cytokine* **2008,** *42* (2), 145-151.
19. 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.
20. Ohto, U.; Fukase, K.; Miyake, K.; Satow, Y., Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **2007,** *316*, 1632-4.
21. 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.
22. Park, B. S.; Lee, J.-O., Recognition of lipopolysaccharide pattern by TLR4 complexes. *Exp. Mol. Med.* **2013,** *45* (12), e66.
23. (a) 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; (b) 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), 13440-53.
24. 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.; Lee, J. O., Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* **2007,** *130* (5), 906-17.
25. (a) Yu, L.; Phillips, R. L.; Zhang, D. S.; 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; (b) Akira, S.; Takeda, K., Toll-like receptor signalling. *Nat. Rev. Immunol.* **2004,** *4*, 499-511.
26. Baxevanis, C. N.; Voutsas, I. F.; Tsitsilonis, O. E., Toll-like receptor agonists: current status and future perspective on their utility as adjuvants in improving antic ancer vaccination strategies. *Immunotherapy* **2013,** *5* (5), 497-511.

27. (a) Garçon, N.; Chomez, P.; Van Mechelen, M., GlaxoSmithKline Adjuvant Systems in vaccines: concepts, achievements and perspectives. *Expert Rev. Vaccines* **2007,** *6* (5), 723-39; (b) Casella, C. R.; Mitchell, T. C., Inefficient TLR4/MD-2 heterotetramerization by monophosphoryl lipid A. *PloS one* **2013,** *8* (4); (c) Thoelen, S.; De Clercq, N.; Tornieporth, N., A prophylactic hepatitis B vaccine with a novel adjuvant system. *Vaccine* **2001,** *19* (17-19), 2400-3; (d) Mitchell, M. S.; Kan-Mitchell, J.; Kempf, R. A.; Harel, W.; Shau, H. Y.; Lind, S., Active specific immunotherapy for melanoma: phase I trial of allogeneic lysates and a novel adjuvant. *Cancer Res.* **1988,** *48* (20), 5883-93.
28. Przetak, M.; Chow, J.; Cheng, H.; Rose, J.; Hawkins, L. D.; Ishizaka, S. T., Novel synthetic LPS receptor agonists boost systemic and mucosal antibody responses in mice. *Vaccine* **2003,** *21* (9-10), 961-70.
29. Wang, S.; Astsaturov, I. A.; Bingham, C. A.; McCarthy, K. M.; von Mehren, M.; Xu, W.; Alpaugh, R. K.; Tang, Y.; Littlefield, B. A.; Hawkins, L. D., Effective antibody therapy induces host-protective antitumor immunity that is augmented by TLR4 agonist treatment. *Cancer Immunol. Immunother.* **2012,** *61* (1), 49-61.
30. Morefield, G. L.; Hawkins, L. D.; Ishizaka, S. T.; Kissner, T. L.; Ulrich, R. G., Synthetic Toll-like receptor 4 agonist enhances vaccine efficacy in an experimental model of toxic shock syndrome. *Clin. Vaccine Immunol.* **2007,** *14* (11), 1499-1504.
31. Stöver, A. G.; Correia, J. D. S.; Evans, J. T.; Cluff, C. W.; Elliott, M. W.; Jeffery, E. W.; Johnson, D. A.; Lacy, M. J.; Baldridge, J. R.; Probst, P., Structure-activity relationship of synthetic toll-like receptor 4 agonists. *J. Biol. Chem.* **2004,** *279* (6), 4440-4449.
32. Johnson, D. A., Synthetic TLR4-active glycolipids as vaccine adjuvants and stand-alone immunotherapeutics. *Curr. Top. Med. Chem.* **2008,** *8* (2), 64-79.
33. Johnson, D. A.; Gregory Sowell, C.; Johnson, C. L.; Livesay, M. T.; Keegan, D. S.; Rhodes, M. J.; Terry Ulrich, J.; 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.
34. Bowen, W. S.; Minns, L. A.; Johnson, D. A.; Mitchell, T. C.; Hutton, M. M.; Evans, J. T., Selective TRIF-dependent signaling by a synthetic toll-like receptor 4 agonist. *Sci. Signal.* **2012,** *5* (211), ra13.
35. 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.
36. Neve, J. E.; Wijesekera, H. P.; Duffy, S.; Jenkins, I. D.; Ripper, J. A.; Teague, S. J.; Campitelli, M.; Garavelas, A.; Nikolakopoulos, G.; Le, P. V., Euodenine A: A Small-Molecule Agonist of Human TLR4. *J. Med. Chem.* **2014,** *57* (4), 1252-1275.
37. Shanmugam, A.; Rajoria, S.; George, A. L.; Mittelman, A.; Suriano, R.; Tiwari, R. K., Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PloS one* **2012,** *7* (2), e30839.
38. Wittebole, X.; Castanares-Zapatero, D.; Laterre, P.-F., Toll-like receptor 4 modulation as a strategy to treat sepsis. *Mediators Inflamm.* **2010,** *2010*.

39. 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), 498-502.
40. (a) Peri, F.; Marinzi, C.; Barath, M.; Granucci, F.; Urbano, M.; Nicotra, F., Synthesis and biological evaluation of novel lipid A antagonists. *Bioorg. Med. Chem.* **2006,** *14* (1), 190-9; (b) Piazza, M.; Rossini, C.; Della Fiorentina, S.; Pozzi, C.; Comelli, F.; Bettoni, I.; Fusi, P.; Costa, B.; Peri, F., Glycolipids and benzylammonium lipids as novel antisepsis agents: synthesis and biological characterization. *J. Med. Chem.* **2009,** *52* (4), 1209-13; (c) Peri, F.; Granucci, F.; Costa, B.; Zanoni, I.; Marinzi, C.; Nicotra, F., Inhibition of lipid a stimulated activation of human dendritic cells and macrophages by amino and hydroxylamino monosaccharides. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (18), 3308-12.
41. Piazza, M.; Calabrese, V.; Damore, G.; Cighetti, R.; Gioannini, T.; Weiss, J.; Peri, F., A Synthetic Lipid A Mimetic Modulates Human TLR4 Activity. *ChemMedChem* **2012,** *7* (2), 213-7.
42. 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.
43. (a) Park, S.-J.; Kang, S. H.; Kang, Y. K.; Eom, Y.-B.; Koh, K. O.; Kim, D. Y.; Youn, H.-S., Inhibition of homodimerization of toll-like receptor 4 by 4-oxo-4-(2-oxo-oxazolidin-3-yl)-but-2-enoic acid ethyl ester. *Int. Immunopharmacol.* **2011,** *11* (1), 19-22; (b) Jin, G. H.; Li, H.; An, S.; Ryu, J.-H.; Jeon, R., Design, synthesis and activity of benzothiazole-based inhibitors of NO production in LPS-activated macrophages. *Bioorg. Med. Chem. Lett.* **2010,** *20* (21), 6199-6202; (c) Kawamoto, T.; Ii, M.; Kitazaki, T.; Iizawa, Y.; Kimura, H., TAK-242 selectively suppresses Toll-like receptor 4-signaling mediated by the intracellular domain. *Eur. J. Pharmacol.* **2008,** *584* (1), 40-48; (d) Chavez, S. A.; Martinko, A. J.; Lau, C.; Pham, M. N.; Cheng, K.; Bevan, D. E.; Mollnes, T. E.; Yin, H., Development of β-Amino Alcohol Derivatives That Inhibit Toll-like Receptor 4 Mediated Inflammatory Response as Potential Antiseptics. *J. Med. Chem.* **2011,** *54* (13), 4659–4669.
44. (a) Barata, T. S.; Teo, I.; Brocchini, S.; Zloh, M.; Shaunak, S., Partially glycosylated dendrimers block MD-2 and prevent TLR4-MD-2-LPS complex mediated cytokine responses. *PLoS Comput. Biol.* **2011,** *7* (6), e1002095; (b) Barata, T.; Teo, I.; Lalwani, S.; Simanek, E.; Zloh, M.; Shaunak, S., Computational design principles for bioactive dendrimer based constructs as antagonists of the TLR4-MD-2-LPS complex. *Biomaterials* **2011,** *32* (33), 8702-8711.
45. Barochia, A.; Solomon, S.; Cui, X.; Natanson, C.; Eichacker, P. Q., Eritoran tetrasodium (E5564) treatment for sepsis: review of preclinical and clinical studies. *Expert opinion on drug metabolism & toxicology* **2011,** *7* (4), 479-94.
46. 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), 1-11.

47. Krivov, G. G.; Shapovalov, M. V.; Dunbrack, R. L., Jr., Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **2009,** *77* (4), 778-95.
48. 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), 455-61.
49. Garate, J. A.; Stöckl, J.; Fernández-Alonso, M. d. C.; Artner, D.; Haegman, M.; Oostenbrink, C.; Jiménez-Barbero, J.; Beyaert, R.; Heine, H.; Kosma, P., Anti-endotoxic activity and structural basis for human MD-2· TLR4 antagonism of tetraacylated lipid A mimetics based on βGlcN (1↔ 1) αGlcN scaffold. *Innate Immun.* **2015,** *21* (5), 490-503.
50. 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), 2050-6.
51. 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**, doi: 10.111 1/cbdd.12749.
52. 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* (Web Server issue), W116-8.
53. 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.
54. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2014,** *47*, 5.6.1-32.
55. 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.
56. 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.
57. Klett, J.; Nunez-Salgado, A.; Dos Santos, H. G.; Cortes-Cabrera, A.; Perona, A.; Gil-Redondo, R.; 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), 3395-408.
58. 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.
59. (a) 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), 2251-4; (b) Manthey, C. L.; Qureshi, N.; Stutz, P. L.; Vogel, S. N., Lipopolysaccharide antagonists block taxol-induced signaling in murine macrophages. *J. Exp. Med.* **1993,** *178* (2), 695-702; (c) 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.
60. 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.
61. 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*, 2251-2254.
62. 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.; Kim, J.-H., The procyanidin trimer C1 inhibits LPS-induced MAPK and NF-κB signaling through TLR4 in macrophages. *Int. Immunopharmacol.* **2013,** *15* (2), 450-456.
63. 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 design, development and therapy* **2016,** *10*, 455.
64. (a) 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; (b) 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), 2309-18.
65. 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.
66. Chan, M.; Hayashi, T.; Mathewson, R. D.; Nour, A.; Hayashi, Y.; Yao, S.; Tawatao, R. I.; Crain, B.; Tsigelny, I. F.; Kouznetsova, V. L.; Messer, K.; Pu, M.; Corr, M.; Carson, D. A.; Cottam, H. B., Identification of substituted pyrimido[5,4-b]indoles as selective Toll-like receptor 4 ligands. *J. Med. Chem.* **2013,** *56* (11), 4206-23.
67. 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. *Biorg. Med. Chem. Lett.* **2014,** *24* (21), 4931-4938.
68. 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.
69. 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. *British journal of pharmacology* **2013,** *168* (8), 1933-1945.
70. 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.

71. (a) 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.; Kim, J. H., The procyanidin trimer C1 inhibits LPS-induced MAPK and NF-kappaB signaling through TLR4 in macrophages. *Int. Immunopharmacol.* **2013,** *15* (2), 450-6; (b) 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 Radic. Res.* **2011,** *45* (5), 611-9; (c) 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), 1263-80.

72. 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.

73. Garate, J. A.; Oostenbrink, C., Lipid A from lipopolysaccharide recognition: structure, dynamics and cooperativity by molecular dynamics simulations. *Proteins* **2013,** *81* (4), 658-74.

74. de Aguiar, C.; Costa, M. G.; Verli, H., Dynamics on human Toll-like receptor 4 complexation to MD-2: The coreceptor stabilizing function. *Proteins: Structure, Function, and Bioinformatics* **2015,** *83* (2), 373-382.

75. 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.

76. 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 : a European journal of chemical biology* **2009,** *10* (4), 645-649.

77. 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.

78. Narayanan, K. B.; Park, H. H., Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis* **2015,** *20* (2), 196-209.

79. 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), 111-115.

80. 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.

81. 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.

82. (a) 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; (b) 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. *Proceedings of the National Academy of Sciences of the United States of America* **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. 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*, 353–364.

86. Kubarenko, A.; Frank, M.; Weber, A., Structure–function relationships of Toll-like receptor domains through homology modelling and molecular dynamics. *Biochem. Soc. Trans.* **2007,** *35* (6), 1515-1518.

87. 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. *Journal of structural biology* **2010,** *169* (3), 323-330.

88. 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.

89. 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.

90. 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.

91. 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.

92. 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.

93. 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.

94. Hart, G. W.; Copeland, R. J., Glycomics hits the big time. *Cell* **2010,** *143* (5), 672-6.

95. (a) Gabius, H.-J.; Siebert, H. C.; André, S.; Jiménez-Barbero, J.; Rudiger, H., Chemical biology of the sugar code. *ChemBioChem* **2004,** *5* (6), 740-764; (b) Gabius, H.-J.; André, S.; Kaltner, H.; Siebert, H. C., The sugar code: functional lectinomics. *Biochim. Biophys. Acta* **2002,** *1572* (2-3), 165-177.

96. (a) Cummings, R. D., The repertoire of glycan determinants in the human glycome. *Mol. Biosyst.* **2009,** *5* (10), 1087-1104; (b) Gabius, H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D., From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* **2011,** *36* (6), 298-313.
97. Mallajosyula, S. S.; MacKerell, A. D., Jr., Influence of solvent and intramolecular hydrogen bonding on the conformational properties of O-linked glycopeptides. *J. Phys. Chem. B* **2011,** *115* (38), 11215-11229.
98. Fadda, E.; Woods, R. J., Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov. Today* **2010,** *15* (15-16), 596-609.
99. Lii, J. H.; Chen, K. H.; Durkin, K. A.; Allinger, N. L., Alcohols, ethers, carbohydrates, and related compounds. II. The anomeric effect. *J. Comput. Chem.* **2003,** *24* (12), 1473-89.
100. Kirschner, K. N.; Woods, R. J., Solvent interactions determine carbohydrate conformation. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (19), 10541-10545.
101. Alonso, E. R.; Pena, I.; Cabezas, C.; Alonso, J. L., Structural Expression of Exo-Anomeric Effect. *The journal of physical chemistry letters* **2016,** *7* (5), 845-50.
102. Dabrowski, J.; Kozar, T.; Grosskurth, H.; Nifant'ev, N. E., Conformational mobility of oligosaccharides: experimental evidence for the existence of an "Anti" conformer of the Gal-beta-1-3-Glc-beta-1-OMe disaccharide. *J. Am. Chem. Soc.* **1995,** *117* (20), 5534-5539.
103. Asensio, J. L.; Jiménez-Barbero, J., The use of the AMBER force field in conformational analysis of carbohydrate molecules: determination of the solution conformation of methyl alpha-lactoside by NMR spectroscopy, assisted by molecular mechanics and dynamics calculations. *Biopolymers* **1995,** *35* (1), 55-73.
104. Widmalm, G., A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr. Res.* **2013,** *378*, 123-132.
105. Vidal, P.; Roldos, V.; Fernandez-Alonso Mdel, C.; Vauzeilles, B.; Bleriot, Y.; Cañada, F. J.; André, S.; Gabius, H.-J.; Jiménez-Barbero, J.; Espinosa, J. F.; Martin-Santamaria, S., Conformational selection in glycomimetics: human galectin-1 only recognizes syn-Psi-type conformations of beta-1,3-linked lactose and its C-glycosyl derivative. *Chem. Eur. J.* **2013,** *19* (43), 14581-14590.
106. Martin-Santamaria, S.; André, S.; Buzamet, E.; Caraballo, R.; Fernandez-Cureses, G.; Morando, M.; Ribeiro, J. P.; Ramirez-Gualito, K.; de Pascual-Teresa, B.; Cañada, F. J.; Menendez, M.; Ramstrom, O.; Jiménez-Barbero, J.; Solis, D.; Gabius, H.-J., Symmetric dithiodigalactoside: strategic combination of binding studies and detection of selectivity between a plant toxin and human lectins. *Org. Biomol. Chem.* **2011,** *9* (15), 5445-5455.
107. Suzuki, T.; Makyio, H.; Ando, H.; Komura, N.; Menjo, M.; Yamada, Y.; Imamura, A.; Ishida, H.; Wakatsuki, S.; Kato, R.; Kiso, M., Expanded potential of seleno-carbohydrates as a molecular tool for X-ray structural determination of a carbohydrate-protein complex with single/multi-wavelength anomalous dispersion phasing. *Biorg. Med. Chem.* **2014,** *22* (7), 2090-2101.
108. Sanchez-Fernandez, E. M.; Risquez-Cuadro, R.; Aguilar-Moncayo, M.; Garcia-Moreno, M. I.; Ortiz Mellet, C.; Garcia Fernandez, J. M., Generalized anomeric effect in

gem-diamines: stereoselective synthesis of alpha-N-linked disaccharide mimics. *Org. Lett.* **2009,** *11* (15), 3306-3309.
109. Asensio, J. L.; Siebert, H. C.; von Der Lieth, C. W.; Laynez, J.; Bruix, M.; Soedjanaamadja, U. M.; Beintema, J. J.; Cañada, F. J.; Gabius, H.-J.; Jiménez-Barbero, J., NMR investigations of protein-carbohydrate interactions: studies on the relevance of Trp/Tyr variations in lectin binding sites as deduced from titration microcalorimetry and NMR studies on hevein domains. Determination of the NMR structure of the complex between pseudohevein and N,N',N''-triacetylchitotriose. *Proteins* **2000,** *40* (2), 218-236.
110. (a) Espinosa, J.-F.; Cañada, F. J.; Asensio, J. L.; Martín-Pastor, M.; Dietrich, H.; Martín-Lomas, M.; Schmidt, R. R.; Jiménez-Barbero, J., Experimental evidence of conformational differences between C-glycosides and O-glycosides in solution and in the protein-bound State: the C-lactose/O-lactose case. *J. Am. Chem. Soc.* **1996,** *118* (44), 10862-10871; (b) Garcia-Herrero, A.; Montero, E.; Munoz, J. L.; Espinosa, J. F.; Vian, A.; Garcia, J. L.; Asensio, J. L.; Cañada, F. J.; Jiménez-Barbero, J., Conformational selection of glycomimetics at enzyme catalytic sites: experimental demonstration of the binding of distinct high-energy distorted conformations of C-, S-, and O-glycosides by E. Coli beta-galactosidases. *J. Am. Chem. Soc.* **2002,** *124* (17), 4804-4810; (c) Gonzalez, L.; Asensio, J. L.; Ariosa-Alvarez, A.; Verez-Bencomo, V.; Jiménez-Barbero, J., Solution conformation and dynamics of the trisaccharide fragments of the O-antigen of Vibrio cholerae O1, serotypes Inaba and Ogawa. *Carbohydr. Res.* **1999,** *321* (1-2), 88-95; (d) Montero, E.; Vallmitjana, M.; Perez-Pons, J. A.; Querol, E.; Jiménez-Barbero, J.; Cañada, F. J., NMR studies of the conformation of thiocellobiose bound to a beta-glucosidase from Streptomyces sp. *FEBS Lett.* **1998,** *421* (3), 243-248.
111. Garner, O. B.; Baum, L. G., Galectin-glycan lattices regulate cell-surface glycoprotein organization and signalling. *Biochem. Soc. Trans.* **2008,** *36* (Pt 6), 1472-1477.
112. Fukuhara, H.; Furukawa, A.; Maenaka, K., New binding face of C-type lectin-like domains. *Structure* **2014,** *22* (12), 1694-1696.
113. Sharon, N.; Lis, H., Lectins: cell-agglutinating and sugar-specific proteins. *Science* **1972,** *177* (4053), 949-959.
114. (a) van Kooyk, Y.; Geijtenbeek, T. B., DC-SIGN: escape mechanism for pathogens. *Nat. Rev. Immunol.* **2003,** *3* (9), 697-709; (b) Jones, D.; Metzger, H.; Schatz, A.; Waksman, S. A., Control of gram-negative bacteria in experimental animals by streptomycin. *Science* **1944,** *100* (2588), 103-105.
115. http://www.sciences.univ-nantes.fr/elnemo/.
116. Kuwabara, N.; Hu, D.; Tateno, H.; Makyio, H.; Hirabayashi, J.; Kato, R., Conformational change of a unique sequence in a fungal galectin from Agrocybe cylindracea controls glycan ligand-binding specificity. *FEBS Lett.* **2013,** *587* (22), 3620-3625.
117. (a) Raz, A.; Nakahara, S., Biological modulation by lectins and their ligands in tumor progression and metastasis. *Anticancer Agents Med. Chem.* **2008,** *8* (1), 22-36; (b) Lannoo, N.; Van Damme, E. J., Nucleocytoplasmic plant lectins. *BBA. General subjects* **2010,** *1800* (2), 190-201.

118. Yoshioka, K.; Sato, Y.; Murakami, T.; Tanaka, M.; Niwa, O., One-step detection of galectins on hybrid monolayer surface with protruding lactoside. *Anal. Chem.* **2010,** *82* (4), 1175-1178.
119. Murakami, T.; Yoshioka, K.; Sato, Y.; Tanaka, M.; Niwa, O.; Yabuki, S., Synthesis and galectin-binding activities of mercaptododecyl glycosides containing a terminal β-galactosyl group. *Biorg. Med. Chem. Lett.* **2011,** *21* (4), 1265-1269.
120. (a) Yang, R.-Y.; Hsu, D. K.; LIu, F.-T., Expression of galectin-3 modulates T-cell growth and apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **1996,** *93* (13), 6737-6742; (b) Liu, F.-T.; Patterson, R. J.; Wang, J. L., Intracellular functions of galectins. *BBA. General subjects* **2002,** *1572* (2), 263-273; (c) Yang, R.-Y.; Rabinovich, G. A.; Liu, F.-T., Galectins: structure, function and therapeutic potential. *Expert Rev. Mol. Med.* **2008,** *10*, e17; (d) Delacour, D.; Koch, A.; Jacob, R., The role of galectins in protein trafficking. *Traffic* **2009,** *10* (10), 1405-1413; (e) Sato, S.; St-Pierre, C.; Bhaumik, P.; Nieminen, J., Galectins in innate immunity: dual functions of host soluble β-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). *Immunol. Rev.* **2009,** *230* (1), 172-187; (f) Elola, M. T.; Ferragut, F.; Cárdenas, D. V.; Nugnes, L. G.; Gentilini, L.; Laderach, D.; Troncoso, M. F.; Compagno, D.; Wolfenstein-Todel, C.; Rabinovich, G. A., Expression, localization and function of galectin-8, a tandem-repeat lectin, in human tumors. *Histol. Histopathol.* **2014,** *29* (9), 1093-1105.
121. (a) Ochieng, J.; Green, B.; Evans, S.; James, O.; Warfield, P., Modulation of the biological functions of galectin-3 by matrix metalloproteinases. *BBA. General subjects* **1998,** *1379* (1), 97-106; (b) Vasta, G. R., Roles of galectins in infection. *Nat. Rev. Microbiol.* **2009,** *7* (6), 424-438; (c) Tuğçe, Ç., Immunhistochemical expression of galectin-3 in cancer: a review of the literature. *Turk. Patoloji Derg.* **2012,** *28* (1), 1-10; (d) Radosavljevic, G.; Volarevic, V.; Jovanovic, I.; Milovanovic, M.; Pejnovic, N.; Arsenijevic, N.; Hsu, D. K.; Lukic, M. L., The roles of Galectin-3 in autoimmunity and tumor progression. *Immunol. Res.* **2012,** *52* (1-2), 100-110; (e) Fujihara, S.; Mori, H.; Kobara, H.; Rafiq, K.; Niki, T.; Hirashima, M.; Masaki, T., Galectin-9 in cancer therapy. *Recent Pat. Endocr. Metab. Immune Drug Discov.* **2013,** *7* (2), 130-137; (f) Heusschen, R.; Griffioen, A. W.; Thijssen, V. L., Galectin-9 in tumor biology: a jack of multiple trades. *BBA Reviews on cancer* **2013,** *1836* (1), 177-185; (g) Shin, T., The pleiotropic effects of galectin-3 in neuroinflammation: a review. *Acta Histochem.* **2013,** *115* (5), 407-411; (h) Duray, A.; De Maesschalck, T.; Decaestecker, C.; Remmelink, M.; Chantrain, G.; Neiveyans, J.; Horoi, M.; Leroy, X.; Gabius, H.-J.; Saussez, S., Galectin fingerprinting in naso-sinusal diseases. *Oncol. Rep.* **2014,** *32* (1), 23-32; (i) Kopitz, J.; Vértesy, S.; André, S.; Fiedler, S.; Schnölzer, M.; Gabius, H.-J., Human chimera-type galectin-3: defining the critical tail length for high-affinity glycoprotein/cell surface binding and functional competition with galectin-1 in neuroblastoma cell growth regulation. *Biochimie* **2014,** *104*, 90-99; (j) Thijssen, V. L.; Griffioen, A. W., Galectin-1 and-9 in angiogenesis: a sweet couple. *Glycobiology* **2014**, cwu048.
122. (a) Ito, K.; Stannard, K.; Gabutero, E.; Clark, A. M.; Neo, S.-Y.; Onturk, S.; Blanchard, H.; Ralph, S. J., Galectin-1 as a potent target for cancer therapy: role in the tumor microenvironment. *Cancer Metastasis Rev.* **2012,** *31* (3-4), 763-778; (b) St-Pierre, C.; Ouellet, M.; Giguère, D.; Ohtake, R.; Roy, R.; Sato, S.; Tremblay, M. J., Galectin-1-specific

inhibitors as a new class of compounds to treat HIV-1 infection. *Antimicrob. Agents Chemother.* **2012,** *56* (1), 154-162; (c) Zucchetti, M.; Bonezzi, K.; Frapolli, R.; Sala, F.; Borsotti, P.; Zangarini, M.; Cvitkovic, E.; Noel, K.; Ubezio, P.; Giavazzi, R., Pharmacokinetics and antineoplastic activity of galectin-1-targeting OTX008 in combination with sunitinib. *Cancer Chemother. Pharmacol.* **2013,** *72* (4), 879-887; (d) Astorgues-Xerri, L.; Riveiro, M. E.; Tijeras-Raballand, A.; Serova, M.; Neuzillet, C.; Albert, S.; Raymond, E.; Faivre, S., Unraveling galectin-1 as a novel therapeutic target for cancer. *Cancer Treat. Rev.* **2014,** *40* (2), 307-319; (e) Funasaka, T.; Raz, A.; Nangia-Makker, P. In *Nuclear transport of galectin-3 and its therapeutic implications*, Seminars in Cancer Biology, Elsevier: 2014; pp 30-38.

123. André, S.; Kaltner, H.; Manning, J. C.; Murphy, P. V.; Gabius, H.-J., Lectins: getting familiar with translators of the sugar code. *Molecules* **2015,** *20* (2), 1788-1823.

124. Ahmad, N.; Gabius, H.-J.; André, S.; Kaltner, H.; Sabesan, S.; Roy, R.; Liu, B.; Macaluso, F.; Brewer, C. F., Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. *J. Biol. Chem.* **2004,** *279* (12), 10841-10847.

125. Vasta, G. R.; Ahmed, H.; Nita-Lazar, M.; Banerjee, A.; Pasek, M.; Shridhar, S.; Guha, P.; Fernandez-Robledo, J. A., Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. *Front. Immunol.* **2012,** *3*, 199.

126. Thijssen, V. L.; Barkan, B.; Shoji, H.; Aries, I. M.; Mathieu, V.; Deltour, L.; Hackeng, T. M.; Kiss, R.; Kloog, Y.; Poirier, F., Tumor cells secrete galectin-1 to enhance endothelial cell activity. *Cancer Res.* **2010,** *70* (15), 6216-6224.

127. Collins, P. M.; Bum-Erdene, K.; Yu, X.; Blanchard, H., Galectin-3 interactions with glycosphingolipids. *J. Mol. Biol.* **2014,** *426* (7), 1439-1451.

128. Salomonsson, E.; Carlsson, M. C.; Osla, V.; Hendus-Altenburger, R.; Kahl-Knutson, B.; Öberg, C. T.; Sundin, A.; Nilsson, R.; Nordberg-Karlsson, E.; Nilsson, U. J., Mutational tuning of galectin-3 specificity and biological function. *J. Biol. Chem.* **2010,** *285* (45), 35079-35091.

129. Bourne, Y.; Bolgiano, B.; Liao, D.-I.; Strecker, G.; Cantau, P.; Herzberg, O.; Feizi, T.; Cambillau, C., Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. *Nat. Struct. Mol. Biol.* **1994,** *1* (12), 863-870.

130. Jin, W.; Li, C.; Du, T.; Hu, K.; Huang, X.; Hu, Q., DC-SIGN plays a stronger role than DCIR in mediating HIV-1 capture and transfer. *Virology* **2014,** *458-459*, 83-92.

131. Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I., Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001,** *294* (5549), 2163-6.

132. de Bakker, B. I.; de Lange, F.; Cambi, A.; Korterik, J. P.; van Dijk, E. M.; van Hulst, N. F.; Figdor, C. G.; Garcia-Parajo, M. F., Nanoscale organization of the pathogen receptor DC-SIGN mapped by single-molecule high-resolution fluorescence microscopy. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2007,** *8* (10), 1473-80.

133. Engering, A.; Geijtenbeek, T. B.; van Vliet, S. J.; Wijers, M.; van Liempt, E.; Demaurex, N.; Lanzavecchia, A.; Fransen, J.; Figdor, C. G.; Piguet, V.; van Kooyk, Y., The

dendritic cell-specific adhesion receptor DC-SIGN internalizes antigen for presentation to T cells. *J. Immunol.* **2002,** *168* (5), 2118-26.
134. Tacken, P. J.; de Vries, I. J.; Gijzen, K.; Joosten, B.; Wu, D.; Rother, R. P.; Faas, S. J.; Punt, C. J.; Torensma, R.; Adema, G. J.; Figdor, C. G., Effective induction of naive and recall T-cell responses by targeting antigen to human dendritic cells via a humanized anti-DC-SIGN antibody. *Blood* **2005,** *106* (4), 1278-85.
135. Unger, W. W.; van Beelen, A. J.; Bruijns, S. C.; Joshi, M.; Fehres, C. M.; van Bloois, L.; Verstege, M. I.; Ambrosini, M.; Kalay, H.; Nazmi, K.; Bolscher, J. G.; Hooijberg, E.; de Gruijl, T. D.; Storm, G.; van Kooyk, Y., Glycan-modified liposomes boost CD4+ and CD8+ T-cell responses by targeting DC-SIGN on dendritic cells. *Journal of controlled release : official journal of the Controlled Release Society* **2012,** *160* (1), 88-95.
136. Horrevorts, S. K.; Duinkerken, S.; Bloem, K.; Secades, P.; Kalay, H.; Musters, R. J.; van Vliet, S. J.; Garcia-Vallejo, J. J.; van Kooyk, Y., Toll-Like Receptor 4 Triggering Promotes Cytosolic Routing of DC-SIGN-Targeted Antigens for Presentation on MHC Class I. *Front. Immunol.* **2018,** *9*, 1231.
137. Gringhuis, S. I.; den Dunnen, J.; Litjens, M.; van der Vlist, M.; Geijtenbeek, T. B., Carbohydrate-specific signaling through the DC-SIGN signalosome tailors immunity to Mycobacterium tuberculosis, HIV-1 and Helicobacter pylori. *Nature immunology* **2009,** *10* (10), 1081-8.
138. Feng, D.; Wang, Y.; Liu, Y.; Wu, L.; Li, X.; Chen, Y.; Chen, Y.; Chen, Y.; Xu, C.; Yang, K.; Zhou, T., DC-SIGN reacts with TLR-4 and regulates inflammatory cytokine expression via NF-kappaB activation in renal tubular epithelial cells during acute renal injury. *Clinical and experimental immunology* **2018,** *191* (1), 107-115.
139. van Liempt, E.; Bank, C. M.; Mehta, P.; Garcia-Vallejo, J. J.; Kawar, Z. S.; Geyer, R.; Alvarez, R. A.; Cummings, R. D.; Kooyk, Y.; van Die, I., Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS Lett.* **2006,** *580* (26), 6123-31.

CHAPTER 2:

Materials and Methods

OVERVIEW

The rise of computational methods and their wider application in research has been very helpful to solve biological problems, since they allows to explain molecular recognition events, facilitate the design of novel molecules with higher affinity and guide the synthesis in the optimization of new molecules.[1] Docking techniques predict the possible binding poses at a concrete binding site of a given target, and provides a first estimation of the binding energy. The estimation of the strength of the intermolecular interaction between the ligand and its biological target is predicted through molecular mechanics (MM) and molecular dynamics (MD). These methods are also used to predict the conformation of the small molecules and to model conformational changes in the target that may occur when the binding occurs. In the case that the 3D structure of the target is not available (it has not been elucidated structurally), homology modelling will allow us to predict it, using templates that share common features with it. Semi-empirical, *ab initio* quantum mechanics methods and density functional theory (DFT) are often used to provide optimized parameters for the molecular mechanics calculations and also to estimate the electronic properties (electrostatic potential, polarizability, etc.) of the drug candidate that will influence binding affinity.

This chapter aims to describe the different computational techniques that have been used is this thesis.

1.1 Docking

The interactions 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 produced signal (e.g., agonism or antagonism). Molecular docking is a

computational procedure that aims to predict the preferred orientation of a ligand with a macromolecular target (also called receptor) in which they are bound to each other to form a stable complex. Docking is a useful tool to predict bioactive conformations, identify binding sites inside a given receptor, unveil essential ligand-receptor interactions, and to screen vast databases of potential ligands.[2]

Several docking programs are available in the context of molecular recognition and drug design. We can cite AutoDock[3], AutoDock Vina[4], DOCK[5], FlexX[6], GLIDE[7], ICM[8], PhDOCK[9], and Surflex[10]). They have been extensively tested and compared.[11] Although each docking program operates slightly differently, a docking calculation is basically characterized by two main steps: in the first step, a conformational search is performed, to predict possible conformations of the small molecule (ligand); in the subsequent second step, for the different binding poses, the ligand binding energy is calculated by applying a scoring function, so the predicted ligand-receptor complexes are scored and ranked. Therefore, the main differences among the existing docking programs rely on how they perform the computational search and which scoring function they use in order to rank the docked poses.[12]

Regarding the conformational search, the accuracy and the computational cost of a docking calculation depend also on the number of rotatable bonds present in the ligand. Each program has a threshold of the maximum number of allowed rotatable bonds, due to the exponentially growth of the computational cost that causes a higher number of this feature.

Three types of search algorithm can be applied: shape matching, systematic search and stochastic or non-deterministic algorithms. The more time-efficient algorithms, in terms of accuracy and computational time, are the stochastic algorithms.

In the stochastic algorithm, random changes in the ligand are executed. This change will be accepted or rejected according to probabilistic criteria. Four types of stochastic algorithms are known: Monte Carlo (MC) methods, Evolutionary Algorithms (EA), Tabu search methods and Swarm Optimization (SO) methods.

In MC methods, Boltzmann probability function is applied:

$$P = exp\left[\frac{-(E_1 - E_0)}{k_B T}\right] \qquad \textbf{[2-1]}$$

E_0 and E_1 represent the energy score before and after the change of the conformer, respectively, k_B the Boltzmann constant and T the absolute temperature of the system.

EA algorithms search for the correct ligand binding mode using ideas based on genetics and evolutionary processes in biological systems. A "natural protocol" is applied, which consists in the first generation of individuals and the evaluation of the fitness. According to the Darwin theory, the best-fit individuals for reproduction are selected, and after crossover and mutations new individuals are generated. Thus, the optimization is based on a selection process that mimics biological evolution. One of the most popular is the Genetic Algorithm (GA), in which an analogy with evolutionary selection is applied. Basically, the translation, orientation, and conformation of a small molecule describe the gene, each state (conformation) corresponds to the genotype, and the atomic coordinates correspond to the phenotype. The best solutions will survive and will reproduce (with crossover and mutation), whereas the worse ones will be discarded. Other genetic algorithms have been proposed in order to represent more complex situations.[13] For example, an improvement of the genetic algorithm can be obtained by introducing a local search (LS) method.[14] A LS method performs energy minimization, without requiring gradient information about the local energy

landscape. The hybrid method between GA and LS is named Lamarckian Genetic Algorithm (LGA). Another hybrid algorithm is the simulated annealing, in which a GA and LS methods are applied depending on the temperature. The advantage of this method is the possibility to overcome energy barriers separating energetic valleys by means of high temperature.

Regarding the second step, the scoring and ranking of the different predicted binding poses are performed by means of a scoring function. The available scoring functions can be of three types:

- **Force Field (FF) based scoring functions**,[15] where a classic force field is employed to compute individual interaction terms such as van der Waals and electrostatic energies and stretching, bending, and torsional energies. Several disadvantages are related to this scoring function. One of the disadvantages is the solvent effect, which is finally defined by a distance-dependent dielectric constant. Another problem related to the FF scoring function is how to treat the water. The Poisson-Boltzmann/Surface Area (PB/SA) model and the Generalized-Born/Surface Area (GB/SA) use implicit solvent models to overcome this problem. But the more severe challenge still remains the entropic effect.

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \qquad \textbf{[2-2]}$$

In equation [2-2], r_{ij} represents the distance between an atom i of the protein and an atom j of the ligand. Aij and Bij are the van der Walls parameters, and q_i and q_j are the atomic charges of the atom i and j respectively. $\varepsilon(r_{ij})$ reflects the screening effect of water on electrostatic interactions and it is usually set to $4r_{ij}$.

- **Empirical scoring function**[16] calculates the overall binding free energy, by summing up a set of weighted empirical energy terms, including hydrogen bond (H-bond) and hydrophobic interactions. Compared to the FF scoring functions, it is more computationally efficient. Its applicability depends on the training set (*i.e.* known experimental data for a set of molecules) and the fitting to known binding affinities. Glide Score is one of the examples of empirical scoring function.

$$\Delta G = \sum_i W_i \Delta G_i \qquad \textbf{[2-3]}$$

ΔG_i stands for the individual empirical energy terms and the corresponding coefficients W_i.

- **Knowledge-Based (KB) scoring functions**[17] are based on the sum of distance-dependent statistical potentials between the ligand and the target. The structural information of the complex formed by the protein and the ligand is needed.

$$\omega(r) = -k_B T ln\left[\frac{\rho(r)}{\rho^*(r)}\right] \qquad \textbf{[2-4]}$$

the scoring function is dependent on the density of the protein-ligand atom pair ρ at the distance r in the training set, and the pair density in a reference state ρ^* where there are no interatomic interactions at the absolute temperature T. kB is the Boltzmann constant.

In this thesis, three different docking programs were used: AutoDock, AutoDock Vina, and GLIDE.

1.2.1 Autodock

AutoDock[18] is an automated procedure for predicting the interaction of ligands with bio-macromolecular targets developed by Olson and coworkers in 1996.[19] It

works with autodocktools and autogrid, accessory programs that pre-calculate 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 and the macromolecule. This pre-calculation greatly reduce the time needed for the docking calculation.

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, were known.[18] The force field evaluates the binding in two steps. The ligand and protein start in an unbound conformation. In the first step, intramolecular energetics are estimated for the transition from the unbound state of both ligand and protein to the conformation of the ligand and protein in the bound state. The second step evaluates then the intermolecular energetics of combining the ligand and protein in their bound conformation. The force field includes six pair-wise evaluations (V) and an estimation of the conformational entropy lost upon binding (ΔS_{conf}):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}\right) \quad \textbf{[2-5]}$$

where L refers to the ligand and P refers to the protein in a ligand-protein docking calculation.[18]

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

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

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.[18]

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 $V_{bound}^{L-L} - V_{unbound}^{L-L} = 0$.[18]

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. 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.

1.2.2 Autodock Vina

AutoDock Vina is an open-source program to perform molecular docking. It was designed and implemented by Trott *et al.* at the Scripps Research Institute.[4] 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.[4]

Vina scoring function was mostly inspired by X-score, and, like X-score, was tuned using the PDBbind.[20] However, some terms are different from X-score, and, in tuning the scoring function, Vina's developers went beyond linear regression.[4] As optimization algorithm, Vina uses the Iterated Local Search global optimizer[21] similar to that by Abagyan *et al*.[8] 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.[22] Vina uses the Broyden-Fletcher-Goldfarb-Shanno (BFGS)[23] method for the local optimization, which is described as an efficient quasi-Newton method.[4]

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. This fact allows Vina to take advantage of multithreading.[4] AutoDock Vina tends to be faster than AutoDock 4 by two orders of magnitude.

1.2.3 GLIDE

Glide is a commercial docking program provided by Schrödinger.[7] It uses a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor. It has a systematic method to treat ligand flexibility, combined with an exhaustive search algorithm.

The predicting binding affinity and the ranking of the binding pose is performed using GlideScore function based on ChemScore function:

$$\Delta G_{bind} = C_o + C_{lipo}\sum f(r_{lr}) + C_{hbond}\sum g(\Delta r)h(\Delta\alpha) + C_{metal}\sum f(r_{lm}) + C_{rotb}H_{rotb} \quad \textbf{[2-7]}$$

The Glide protocol is intuitive and relies on 3 steps: 1) ligand and protein preparation; 2) the receptor grid generation, and 3) the docking process. Before launching the docking step, Glide has to generate a grid that represents the shape and the properties of the receptor, using several different sets of fields that provide progressively more accurate scoring of ligand poses. The grid permits to dock only the relevant region of the receptor and, thus, saving time calculations.

Regarding the last point, the full docking VS workflow includes 3 docking stages: HTVS (High Throughput Virtual Screening), SP (Standard Precision) and XP (eXtra Precision).[24] The first stage performs HTVS docking. It is intended for rapid screening of very large number of ligands and it has much more restricted conformational sampling than SP docking. The second stage performs SP docking. It is appropriate for screening ligands of unknown quality in large numbers. The third stage is the XP docking and scoring.[25] It is a more powerful and discriminating procedure using an implementation of a modified and expanded

version of the ChemScore scoring function, called GlideScore (equation [2-8]) and categorized as an empirical scoring function.

Glide can be used to perform VS with an accurate binding mode prediction. Furthermore, Glide exhibits excellent docking accuracy and high enrichment across a diverse range of receptor types.

$$\begin{aligned}\Delta G_{bind} = \; & C_{lipo-lipo} \sum f(r_{lr}) \\ & + \; C_{hbond-neut-neut} \sum g(\Delta r) h(\Delta \alpha) \\ & + C_{hbond-neut-charged} \sum g(\Delta r) h(\Delta \alpha) \\ & + C_{hbond-charged-charged} \sum g(\Delta r) h(\Delta \alpha) \\ & + C_{max-metal-ion} \sum f(r_{lm}) + C_{rotb} H_{rotb} \\ & + C_{polar-phob} V_{polar-phob} + C_{coul} E_{coul} + C_{vdW} E_{vdW} \\ & + solvation\ terms\end{aligned} \qquad \textbf{[2-8]}$$

The lipophilic-lipophilic term and H-bonding term are defined as in ChemScore. But the H-bonding term is separated into different weights, depending on whether one of the nature of the donor and acceptor *i.e.* whether they are charged or not. The metal-ligand interaction term also uses the same functional form as in ChemScore but varies in three principal ways: a) this term considers only interactions with anionic acceptor atoms; b) it counts just the single best interaction when two or more bounded metal are found; and c) the net charge on the metal ion is assessed in the unbound apo–protein. If the net charge is positive, the preference for anionic ligand is incorporated into the scoring function. However, if net charge is neutral, the preference is suppressed. The sixth term in the function rewards situations in which a polar, but non H-bonding, atom is found in a hydrophobic region.

The major components in Glide Score equation (2-9) are the contributions from the Coulomb and van der Waals energies between the ligand and the receptor.

$$XPGlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penaltyd} \quad \textbf{[2-9]}$$

Another major component is the introduction of a solvation model and the incorporation of the solvation effects. Glide docks explicit waters into the active site for each energetically competitive ligand pose and utilize empirical scoring terms that assess the exposure of various groups to the explicit waters.

$$E_{bind} = E_{hyd\_enclosure} + E_{hb\_nn\_motif} + E_{hb\_cc\_motif} + E_{PI} + E_{hb\_pair} + E_{phobic\_pair} \quad \textbf{[2-10]}$$

where, $E_{hyd\_enclosure}$ represents hydrophobic enclosure; $E_{hb\_nn\_motif}$ represents the special neutral-neutral H-bond motifs; $E_{hb\_cc\_motif}$ represents special charged H-bond motifs, E_{PI} π-stacking and π-cation interactions; and E_{phobic_pair} hydrophobic atom-atom pair energy term. E_{hb_pair} term is same as defined in ChemScore scoring function,

$$E_{penalty} = E_{desolv} + E_{ligand\_strain} \quad \textbf{[2-11]}$$

where E_{desolv} represents desolvation penalties, and E_{ligand_strain} contac penalties (penalizing strain energy).

2.1 Simulation of Biomolecules

Various computational techniques are currently available to describe and predict the behavior of molecular biosystems on a wide distribution of length and time scales.[26] Each method is appropriate for a particular range of length and time scales.

1.2.4 Quantum Mechanics Methods

Quantum mechanics (QM) is the fundamental theory of physics explaining the behavior of atoms and subatomic particles.[27] Mathematically, the changes over time of such a system are described by the Schrödinger equations:

$$i\hbar \frac{\partial}{\partial t} |\psi(\mathbf{r},t)\rangle = \hat{H} |\psi(\mathbf{r},t)\rangle \qquad \textbf{[2-12]}$$

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, **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 (equation [2-12]) 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:

$$\hat{H} = \hat{T} + \hat{V} \qquad \textbf{[2-13]}$$

  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,[28] is a computational QM modeling method to investigate the electronic structure principally at 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 in 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[29] functional is the most used.

In this thesis, Gaussian program was used to perform all quantum mechanics calculations *i.e.* to parametrize new units for further MD simulation. Gaussian is a general purpose computational chemistry software package initially released in 1970 by Pople and coworkers[30]. It implements packages related to molecular mechanics (Amber), semi-empirical quantum chemistry methods (Austin Model 1,

PM3), built-in density functional methods (DFT, B3LYP and other hybrid functionals), self-consistent field methods (HF), etc.

1.2.5 Molecular Mechanics and Molecular Dynamics Simulations

The 3D structure at atomic level of a biological macromolecule can be obtained through a number of methods among which X-ray crystallography is the prevalent one. In the particular case of a protein, the X-ray crystallographic structure brings excellent information about the spatial organization of each of the atoms (excluding hydrogens) of the amino acids sequence composing this protein at a given crystallized state.[31] 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, binding sites and 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).[32] Although NMR can give information about the molecules in solution and the different conformations of a target,[33] the motion of a given atom (or a whole structure) through time can only be predicted and observed with molecular dynamics methods.[34]

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. However, such a precise description of an atomic system is very costly to be computed in systems with 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.[35]

$$V(r) = \sum_{bonds} k_l\,(l - l_o)^2 + \sum_{angles} k_\theta\,(\theta - \theta_o)^2 + \sum_{torsions} \frac{1}{2} V_n\,[1 + \cos(n\omega - \gamma)] + \sum_{\substack{nonbond \\ pairs}} \left[\frac{A_{ij}}{{r_{ij}}^{12}} - \frac{C_{ij}}{{r_{ij}}^{6}} + \frac{q_j q_i}{4\pi\varepsilon_0 r_{ij}}\right] \quad \textbf{[2-14]}$$

In molecular mechanics, the potential energy of the system is given by equation [2-14]. 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). Extra terms can be added, for example, one term describing hydrogen bonds. In more details, the first term of Equation [2-14] 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 the case that rotation would need 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 1*) A graphical representation of these components is given in *Figure 2*.

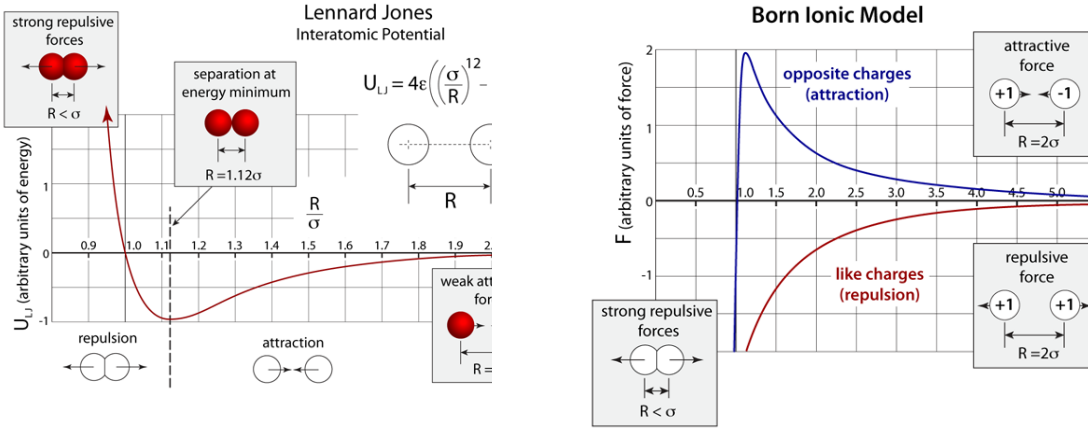


Figure 1: Left) graph of the Lennard-Jones potential function in which regions of repulsion, on the left side of the graph, and attraction, on the right side of the graph are explicitly described. At lowest temperature distances between atoms tend toward the energy minimum. Right) the born ionic-model 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 adapted from [36]).

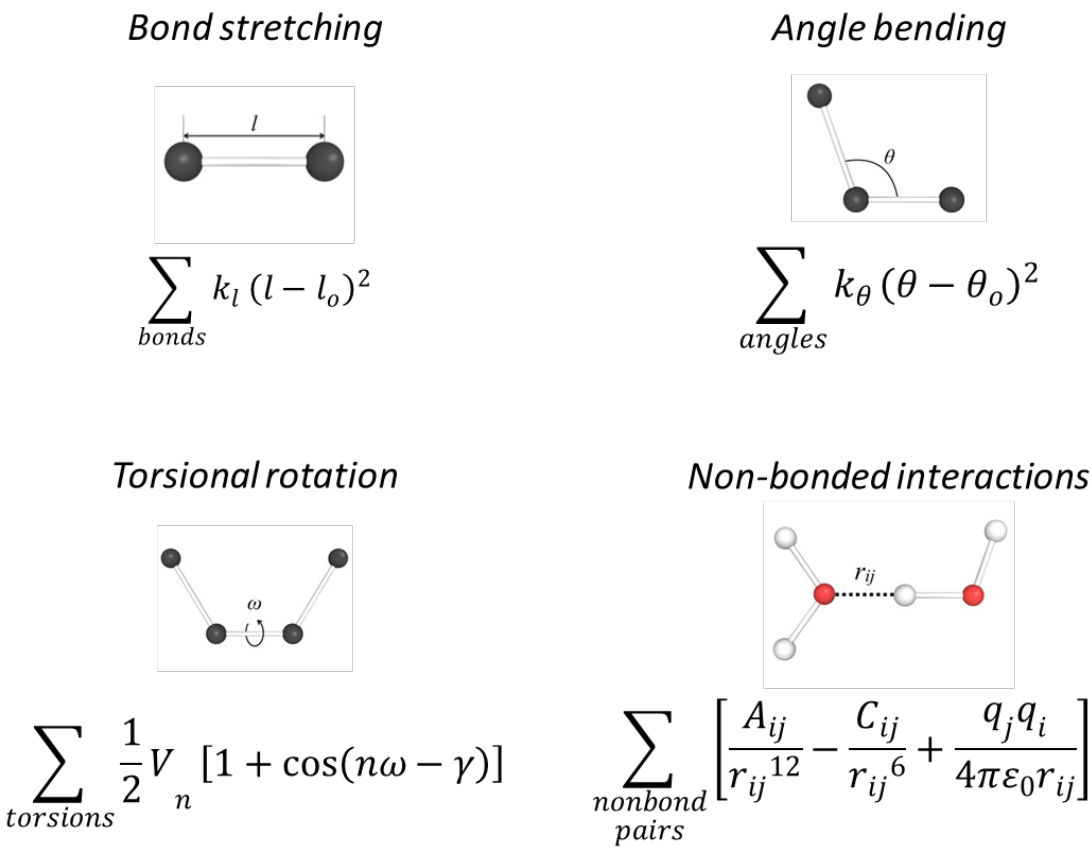


Figure 2: Schematic representation of the molecular mechanics potential function components. Atoms are symbolized by black, white, and red spheres and covalent bonds by white sticks.

In molecular dynamics, successive configurations of the system are generated by integrating Newton's equation of motion (equation [2-15]). The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

$$\mathbf{F} = m\mathbf{a} = m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = m\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2} \qquad \textbf{[2-15]}$$

The force **F** acting 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 (**v)** with respect to time $\left(\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right)$ and the second derivative of the position (**q**) 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 [2-16])

$$\mathbf{q}(t+\Delta t) = \mathbf{q}(t) + \frac{\mathrm{d}\mathbf{q}(\mathrm{t})}{\mathrm{d}t}\Delta\mathrm{t} + \frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2}\frac{\Delta t^2}{2} + \cdots \qquad \textbf{[2-16]}$$

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 [2-16] in which the force F is obtained by differentiating the energy of the system.[1]

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.[37]

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 atom 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.

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.[38]

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[39]), introducing fictitious particles in an extended system (e.g. Nosé-Hoover[40]) 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[41]) and introducing a piston. In addition, the pressure regulation of a simulation needs to be handled accordingly to the system under simulation. For example, 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.

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.

Amber ff14SB. The Amber ff14SB force field[42] improved protein secondary structure balance and dynamics from earlier force fields like ff99, 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.[43] 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.[44] 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 reproduced their benchmarks, but also improved secondary structure content in small peptides and reproduction of NMR $\chi 1$ scalar coupling measurements for proteins in solution.[44]

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 was removed.[45] GLYCAM06 is a consistent and transferable parameter set for modeling carbohydrates, and glycoconjugates.[46] 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. The GLYCAM force field family, especially, GLYCAM06, has been extensively employed in simulations of biomolecules by the larger scientific community.[47] 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.

GAFF is the 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.[48]

Lipid 14, the Amber lipid force field. In 2014, Lipid14[49] 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[50] with a number of refinements inspired by GAFFlipid. 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 simulation.[51] Furthermore, Lipid14 was recently expanded to include cholesterol parameters.[52] Lipid14 has been designed to be fully compatible with the other pairwise-additive protein, nucleic acid, carbohydrate, and small molecule Amber force fields, such as **ff14SB** mentioned above.

Molecular Dynamics Protocol

The simulations reported here, unless specified, are mainly performed with the following 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·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·A 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·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

3.1 Virtual Screening

In the context of drug discovery, virtual screening (VS) techniques have already proved to make hit identification more goal-oriented. Virtual screening allows us to dock a huge amount of molecules (sometimes more than 100,000) on a target structure and to discriminate which of these molecules could bind better. Thus, the costs of performing this kind of screening is relatively low than doing it experimentally. This computational approach has been subjected to extensive attention and revision over the years, from the early perspective of being an emerging method,[53] until the current time where new challenges are faced.[54] The availability of the target 3D coordinates is mandatory in order to perform a VS protocol. Prior knowledge about the ligand binding site may help in the identification of proper binders although, in some approaches, the search for novel binding pockets can be an additional interesting and challenging element in the drug discovery process.

Now a day, a number of public and commercial drug-like small molecules databases can be found. Each database has unique features that may make them more adequate for a particular VS project, and still a relevant number of unique compounds is found within each database that makes it worth considering more than one library if possible. Duplicate analysis of screening libraries comprising the

ones cited below showed that 40%–50% of the totality of structure subjected to analysis were structures exclusive to one supplier library Many of them are freely available and may possess desirable characteristics such as "drug-likeness", being the most popular the "Lipinski Rule of Five".[55] Others collect chemical structures from natural products or approved drugs. Here, a brief description of both databases used in this thesis is given:

- **Zinc Database.** ZINC (recursive acronym for Zinc Is Not Commercial) is a public access database of commercially available compounds developed in the Department of Pharmaceutical Chemistry at the University of California, San Francisco.[56] It contains a constantly growing number of three-dimensional (3D) structures ready-to-dock from catalogues of major compound vendors with annotated relevant protonation and tautomeric states, and properties such as size, calculated logP, number of rotatable bonds, etc. Each molecule also contains purchasability and vendor information, making this ZINC's focus on docking and purchasability the main distinctive characteristic from other databases. In its latest version, ZINC 16 comprises over 120 million purchasable "drug-like" compounds together with information regarding target and biological activity, related scaffolds and bioactive and biogenic compounds (Tanimoto index of 0.6).[57] It also offers other features such as the possibility to define target-focused libraries and to download subsets of a physical property space ("fragment-like", lead-like, drug like subsets).
- **Maybridge database.** Maybridge Screening Hit Discovery collection (over 53,000 compounds) is a commercial library of small hit-like and lead-like organic compounds of high diversity (Tanimoto Clustering at 0.9), that covers ca. 87% of the 400,000 theoretical drug pharmacophores with general compliance with the Lipinsky rule of five and of good ADMET properties.[58] The

HitCreator TM Collection (selection of 14,400 of Maybridge screening compounds) aims to represent the diversity of the main collection covering the drug-like chemical space. Maybridge also offers a fragment library (30,000 fragments), a hit-to-lead building block collection, and a Ro3 2500 diversity fragment library (2500 fragments) with a Tanimoto similarity index of 0.66 (based on standard Daylight fingerprinting), assured solubility, optimized for SPR and Ro3 compliant.

1.2.6 Database Processing

Database processing constitutes a fundamental step in VS approaches. It is crucial to generate the proper chemical library, with the adequate geometries, ionization states and conformations of the chemical entities. A good database processing will assure a rigorous and well-conducted virtual screening, as well as it will decrease the required computational time and it will avoid identification of unsuitable drug candidates.

To prepare the databases and include decoys, different softwares were used in this Thesis:

- LigPrep,[59] a software by Schrödinger LLC, is a collection of tools designed to prepare high quality 3D structures for large numbers of drug-like molecules, starting from 2D or 3D structures. LigPrep starts by converting the input structure files to Maestro format. LigPrep process consists of a series of steps that perform conversions and apply corrections to the structures. It also generates variations on the structures (for example, takes into account different tautomers). Finally, it will optimize the geometry of the ligands. LigPrep produces a single low-energy 3D structure with defined chiralities for each processed input structure and it can also produce a number of structures, from each starting geometry, varying ionization states, tautomeric forms, stereoisomers, and ring conformations. Additionally,

LigPrep offers the option to eliminate molecules from the collection to be screened using various criteria including molecular weight or quantity and types of functional groups composing the molecule.

- AutoDockTools[60] is a widely accessible software that has a graphical interface implemented within the Python Molecular Viewer. The graphical interphase facilitates the formatting of input molecule files, with a set of methods that guide the user through protonation, calculation of charges, and specification of rotatable bonds in the ligand and the protein. As a brief outline of the preparation process, the ligand and the protein are both loaded. AutoDock atom types are assigned, polar hydrogens are added, charges are calculated, and nonpolar hydrogens are merged with the heavier atoms to which they are attached. If the input ligand file presents no charges, ADT will compute Gasteiger charges. Regarding the ligand preparation for the virtual screening, it is important to treat flexibility of the ligands. For this purpose, ligand flexibility is assigned in several steps. First, a root atom is chosen, which will act as the fixed position during coordinate transformation in the docking simulation. To find the optimal atom, the number of atoms in each branch is evaluated, and the root atom that minimizes the size of the largest branch is chosen. However, the ligand flexibility can be limited. As a limitation, each step in ADT has to be launched manually, one by one, as well as the preparation of each ligand. However, it is possible, with simple scripts, to do it automatically.

1.2.7 Docking Tools for Virtual Screening

FLAP

The new molecular modeling tool FLAP (Fingerprints for Ligands and Proteins) provides a common reference framework for comparing molecules, using GRID Molecular Interaction Fields.[61] The MIFs describe the type, strength and direction

of the molecular interactions between two biological partners. These MIFs are then condensed into discrete pharmacophoric points representing favorable and unfavorable interactions using a weighted energy-based and space coverage function. Using these discrete points, all four-point quadruplets are generated, and the resulting pharmacophore quadruplet fingerprint describes the target of interest. In addition to the fingerprints, the GRID MIFs are retained. The targets are then aligned by matching quadruplets in Cartesian space and a field similarity is computed using the pre-calculated MIFs. Hence, the fingerprints are used to find matching pharmacophoric regions and the entire fields are used to score the match.[62]

FLAP consists of a graphical user interface, and several command-line programs which execute the various tasks. FLAP is based on four probes only, H, O, N1, and DRY, which characterize the shape, hydrogen-bond acceptor, hydrogen-bond donor, and hydrophobic interactions, respectively.

We used this program to perform structure-based and ligand- based virtual screening. The aim of running structure-based VS is that the 3D structure of a receptor binding site is known, and can be used as a target for FLAP. When performing alignments, the MIFs of a drug candidate are compared with the receptor MIFs. These similarities are used to rank the potential interactions of drug candidates with the receptor. Notice that the structure-based enrichments are in general lower than the ligand-based enrichments. This is most likely due to the fact that a structure-based search uses a larger cavity definition (there is more noise). Moreover, a ligand-based approach is based on the fact that active compounds should be chemically very similar, biasing the search results to favor a particular chemical species. The structure-based approach has the advantage that ligands

binding different subpockets can be identified, and should not be as closely tied to the chemotype of a ligand template.

Virtual screening employing a Ligand-Based Approach is the best way to select potential new candidate drugs from a library of compounds with known three-dimensional structure but unknown activity against the biological target (the so-called "decoys"). Having access to a series of compounds with known activity on a specific biotarget, FLAP is able to align the molecules from the database (decoys) to one specific and chosen active compound (template). It then computes the GRID based - MIFs similarity) between decoys and the template assigning a score that can be used to rank the most similar compounds The assumption is that the higher the similarity with the template, the higher the probability of similar mechanism of action at the receptor site. Once scores are produced for each molecule of the dataset, it can be necessary to evaluate how well the known active compounds are recognized by FLAP through the use of an Enrichment Plot or a ROC curve.

GLIDE

The full docking VS workflow in GLIDE includes 3 docking stages: HTVS, SP (Standard Precision) and XP (Extra Precision). The first stage performs High Throughput Virtual Screening (HTVS) docking. It is intended for rapid screening of very large number of ligands and has much more restricted conformational sampling than SP docking. The second stage performs SP docking. It is appropriate for screening ligands of unknown quality in large numbers. The third stage is the XP docking and scoring. It is a more powerful and discriminating procedure using an implementation of a modified and expanded version of the ChemScore scoring function, called GlideScore and categorized as an empirical scoring function. Glide can be used to perform virtual screening, accurate binding mode precision and

furthermore, Glide exhibits excellent docking accuracy and high enrichment across a diverse range of receptor types.

2 Homology Modelling

The availability of the 3D structure of a given protein is a limiting step in the drug design and development process. Usually, the 3D structures of the macromolecules are elucidated by NMR, X-ray crystallography or electron microscopy techniques, and they are deposited in the public database PDB.[63] When the 3D structure for a protein of interest has not been resolved by experimental techniques, homology modelling is very powerful to predict the structure. Enormous efforts have been devoted to develop user-friendly servers, and secondary and tertiary structure prediction platforms have risen in the recent years to make structure prediction more accessible to a wide range of investigators, with the additional attractive of being usually freely available. All these servers are based on the Anfinsen´s dogma relating the structure of a given protein, and thus its function, to its amino acid sequence, and on the principle that the native structure of the protein is the one in which the free energy is the lowest.[64] In particular, homology modelling methods use a homologous protein of known structure as a template to determine the structure of the problem protein. The principal steps in a homology modelling protocol are template selection, alignment, backbone and side-chain prediction and structure optimization. The target protein is aligned to the template sequence, which can be found online using BLAST (Basic Local Alignment Search Tool). BLAST scoring estimates the quality of the sequence alignment using the BLOSUM62 (BLOck SUbstitution Matrix).[65] In the score assignment, different parameters are taken into account such as the percentage of identity between the searched protein or target and the template (a good model has minimum 30% of the percentage identity) and the percentage of query cover and the backbone atoms

are mapped onto the three dimensional template structure. The non-conserved side chains orientations are predicted. Finally, a force field is used to optimize the structure, which basically includes the removal of steric clashes and the optimization of hydrogen bond network.

Some examples of homology modelling web servers are presented here. ROSETTA[66] is a fragment-based method that combines native-like structures of unrelated proteins to search the most probable 3D structure for a given amino acid sequence. The scoring function includes biological information, like solvation and residue-pair interactions, and the 3D structures are generated by combining fragments of known structures based on a large number of sequences with known structure present in the Protein Data Bank (PDB). Another example is the I-TASSER server,[67] an online resource for automated protein structure prediction that uses basically the information stored in the PDB to perform the alignment of related sequences. Then, the protein of study is divided into aligned and unaligned regions, these regions are connected, and a Monte-Carlo (MC) simulation is performed for the assembling and refinement. I-TASSER combines different methods to add backbone atoms, side chains, and to refine the final structure. SWISS-MODEL[68] is also a popular homology modelling web server that presents three different modelling protocols, depending on the degree of similarity between the input sequence and the template, allowing to predict structures with a very low percentage of similarity. Another homology modelling software that uses comparative methods is MODELLER.[69]

4.1 Other Programs

2.2.1 PyMol

PyMOL molecular graphics and modeling package[70] 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

2.2.2 Maestro

Maestro[71] 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

2.2.3 Antechamber

Antechamber[72] 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[30] at the Hartree-Fock level (HF/6-31G* Pop=MK iop(6/33=2) iop(6/42=6)) from the DFT B3LYP optimized structure, then derived and formatted for Ambertools15 and Amber 14 with Antechamber assigning the general AMBER force field (GAFF)[48] atom types. Subsequent modifications to the atom types were made when thought necessary, e.g. to take advantage of the GLYCAM06[45] force field should a saccharide fragment of a bigger molecule be already described by the force field.

2.2.4 LeaP

LEaP[73], 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.[74] 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, one needs to define the parameters necessary to describe the interface between the two force fields. To do that, modifications were introduced within GAFF, GLYCAM06, and AMBERff14. LEaP was used to create force field template for new residues. 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. 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.
2. Gupta, M.; Sharma, R.; Kumar, A., Docking techniques in pharmacology: How much promising? *Computational biology and chemistry* **2018,** *76*, 210-217.
3. 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.
4. 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.
5. 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.
6. Kramer, B.; Rarey, M.; Lengauer, T., Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins* **1999,** *37* (2), 228-241.
7. 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.
8. 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.
9. 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.
10. Jain, A. N., Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003,** *46* (4), 499-511.
11. (a) 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; (b) 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.
12. (a) Pagadala, N. S.; Syed, K.; Tuszynski, J., Software for molecular docking: a review. *Biophysical reviews* **2017,** *9* (2), 91-102; (b) Li, J.; Fu, A.; Zhang, L., An

Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. *Interdisciplinary sciences, computational life sciences* **2019**.
13. Michalewicz, Z., Evolutionary programming and genetic programming. In *Genetic algorithms+ data structures= evolution programs*, Springer: 1996; pp 283-287.
14. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998,** *19* (14), 1639-1662.
15. Genheden, S.; Ryde, U., The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015,** *10* (5), 449-461.
16. (a) Böhm, H.-J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994,** *8* (3), 243-256; (b) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997,** *11* (5), 425-445.
17. (a) Verkhivker, G.; Appelt, K.; Freer, S.; Villafranca, J., Empirical free energy calculations of ligand-protein crystallographic complexes. I. Knowledge-based ligand-protein interaction potentials applied to the prediction of human immunodeficiency virus 1 protease binding affinity. *Protein Eng.* **1995,** *8* (7), 677-691; (b) Gohlke, H.; Hendlich, M.; Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000,** *295* (2), 337-356.
18. Forli, W. E. H.; Halliday, S.; Belew, R.; Olson, A. J., AutoDock Version 4.2.
19. Goodsell, D. S.; Morris, G. M.; Olson, A. J., Automated docking of flexible ligands: applications of AutoDock. *Journal of molecular recognition : JMR* **1996,** *9* (1), 1-5.
20. (a) 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; (b) 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.
21. (a) Baxter, J., Local optima avoidance in depot location. *J. Oper. Res. Soc.* **1981,** *32* (9), 815-819; (b) Blum, C.; Roli, A., Hybrid metaheuristics: an introduction. In *Hybrid Metaheuristics*, Springer: 2008; pp 1-30.
22. Chen, Y.; Roux, B., Generalized Metropolis acceptance criterion for hybrid non-equilibrium molecular dynamics-Monte Carlo simulations. *J. Chem. Phys.* **2015,** *142* (2), 024101.
23. Nocedal, J.; Wright, s., Numerical Optimization, Chapter 10, Nonlinear Least-Squares Problems, eds. Glynn, P. & Robinson, sM. springer, New York: 1999.

24. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004,** *47* (7), 1750-9.
25. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** *49* (21), 6177-96.
26. van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J., Biomolecular simulation and modelling: status, progress and prospects. *Journal of the Royal Society, Interface* **2008,** *5 Suppl 3*, S173-90.
27. Griffiths, D. J., *Introduction to quantum mechanics*. 2nd edition ed.; Cambridge University Press: Cambridge, 1982.
28. (a) Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Phys. Rev.* **1964,** *136* (3B), B864-B871; (b) Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** *140* (4A), A1133-A1138.
29. Stephens, P.; Devlin, F.; Ashvar, C.; Chabalowski, C.; Frisch, M., Theoretical calculation of vibrational circular dichroism spectra. *Faraday Discuss.* **1994,** *99*, 103-119.
30. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. B.01*, Wallingford, CT, 2016.
31. Drenth, J., *Principles of protein X-ray crystallography*. Springer Science & Business Media: 2007.
32. (a) Bai, X.-C.; McMullan, G.; Scheres, S. H., How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci* **2015,** *40* (1), 49-57; (b) 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.

33. Stark, J. L.; Powers, R., Application of NMR and molecular docking in structure-based drug discovery. *Top. Curr. Chem.* **2012,** *326*, 1-34.
34. Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018,** *99* (6), 1129-1143.
35. Vanommeslaeghe, K.; Guvench, O.; MacKerell, A. D., Jr., Molecular mechanics. *Current pharmaceutical design* **2014,** *20* (20), 3281-92.
36. http://atomsinmotion.com/book/chapter5/md (accessed May 2019).
37. (a) 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; (b) Verlet, L., Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **1967,** *159* (1), 98-103.
38. 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.
39. 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.
40. Nosé, S., A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys* **1984,** *81* (1), 511-519.
41. (a) Parrinello, M.; Rahman, A., Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **1980,** *45* (14), 1196-1199; (b) Parrinello, M.; Rahman, A., Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182-7190.
42. 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.
43. 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.
44. 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.
45. 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.
46. (a) 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; (b)

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.

47. (a) 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; (b) 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; (c) 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.

48. 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.

49. Dickson, C. J.; Madej, B. D.; Skjevik, A. 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.

50. Skjevik, A. 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), 11124-36.

51. (a) Skjevik, A. 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. *Phys. Chem. Chem. Phys.* **2016,** *18* (15), 10573-84; (b) Skjevik, A. 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. *Chemical communications (Cambridge, England)* **2015,** *51* (21), 4402-5.

52. 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), 12424-35.

53. Lyne, P. D., Structure-based virtual screening: an overview. *Drug Discov. Today* **2002,** *7* (20), 1047-55.

54. (a) Haga, J. H.; Ichikawa, K.; Date, S., Virtual Screening Techniques and Current Computational Infrastructures. *Curr. Pharm. Des.* **2016,** *22* (23), 3576-84; (b) Yan, X.; Liao, C.; Liu, Z.; Hagler, A. T.; Gu, Q.; Xu, J., Chemical Structure Similarity Search for Ligand-based Virtual Screening: Methods and Computational Resources. *Curr. Drug Targets* **2016,** *17* (14), 1580-1585; (c) Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W., State-of-the-art strategies for targeting protein-

protein interactions by small-molecule inhibitors. *Chem. Soc. Rev.* **2015,** *44* (22), 8238-59; (d) Pérez-Regidor, L.; Zarioh, M.; Ortega, L.; Martín-Santamaría, S., Virtual screening approaches towards the discovery of Toll-Like receptor modulators. *International journal of molecular sciences* **2016,** *17* (9), 1508.

55. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced drug delivery reviews* **2001,** *46* (1-3), 3-26.

56. http://zinc15.docking.org ZINC 2016. (accessed March 2019).

57. Willett, P., Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **2006,** *11* (23–24), 1046-1053.

58. Maybridge database. www.maybridge.com (accessed March 2019).

59. ***Schrödinger Realease 2019-1:*** *LigPrep*, Schrödinger, LLC: New York, NY, 2019.

60. 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.

61. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S., A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** *47* (2), 279-94.

62. (a) Cross, S.; Baroni, M.; Goracci, L.; Cruciani, G., GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *J. Chem. Inf. Model.* **2012,** *52* (10), 2587-98; (b) Cross, S.; Ortuso, F.; Baroni, M.; Costa, G.; Distinto, S.; Moraca, F.; Alcaro, S.; Cruciani, G., GRID-based three-dimensional pharmacophores II: PharmBench, a benchmark data set for evaluating pharmacophore elucidation methods. *J. Chem. Inf. Model.* **2012,** *52* (10), 2599-608.

63. http://www.sciences.univ-nantes.fr/elnemo/.

64. Anfinsen, C. B., Principles that govern the folding of protein chains. *Science* **1973,** *181* (4096), 223-30.

65. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997,** *25* (17), 3389-402.

66. Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D., Protein structure prediction using Rosetta. *Methods Enzymol.* **2004,** *383*, 66-93.

67. Roy, A.; Kucukural, A.; Zhang, Y., I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **2010,** *5* (4), 725-38.

68. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **2006,** *22* (2), 195-201.
69. Webb, B.; Sali, A., Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinformatics* **2016,** *54*, 5.6.1-5.6.37.
70. Schrödinger, L., PyMOL(TM) Molecular Graphics System, Version 1.6.0.0.
71. ***Schrödinger Release 2019-1:*** *Maestro*, New York, NY, 2019.
72. 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.
73. 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**.
74. Case, D.; Betz, R.; Botello-Smith, W.; Cerutti, D.; Cheatham III, T.; Darden, T.; Duke, R.; Giese, T.; Gohlke, H.; Goetz, A., Amber 2016.

CHAPTER 3:

Design of lactose-based galectin binders with high affinity profile

3.1 Introduction

Galectins are β-galactoside binding proteins involved in many physiological and pathological processes. In mammals, they exert a wide range of biological functions and play an important role in immune and inflammatory responses, and in the cell cycle regulation, among other processes. Also, galectins have been identified as key players in cancer and immune pathologies pointing towards these proteins as important targets for the development of drugs for cancer therapy and immunotherapy.[1] Current research confirms the involvement of galectins in tumor progression and indicates that galectins could serve as therapeutic targets, and have prognostic value.[2] In particular, galectins -1, -3, and -7 are adhesion/growth-regulatory galectins whose functions have been shown to be correlated among them in several tumoral processes.[3] Galectin-3 (gal-3) participates in the regulation of inflammation,[4] and mediates important tumor-related functions (tumorigenesis, cell adhesion, proliferation, differentiation, angiogenesis, and metastasis), being involved in the development and malignancy of several types of tumor.[5] Therefore, synthetic gal-3 binders have been pursued for the understanding of the molecular regulation of gal-3 expression and for the development of new antitumoral therapeutic strategies.[6] Galectin-1 (gal-1) also acts in cellular adhesion, mobility and invasion, tumor-induced angiogenesis, and apoptosis. Although the mechanisms of these processes are not yet well understood, and less is known about gal-1 pathways compared to gal-3,[7] the overexpression of gal-1 in cancer progression indicates a significant role for gal-1.[8] In fact, both galectins, -1 and -3, seem to be main players in cancer biology as it can be deduced from significant studies. Regarding Galectin-7 (gal-7), it is one of the galectins least studied. The available information about altered gal-7 expression between normal and tumor types is also mostly limited to single reports. In addition, some findings resemble

the differences in gal-3 expression with respect to cancer subtype and cellular localization. These observations further exemplify that galectin localization might be used to distinguish between histological subtypes.[1a]

The possibility of differential targeting of galectins has been sought as an interesting aim in the glycomimetics field. Selective galectin binders could help in cellular assays to unravel the mechanisms regulated by galectins and their different (and sometimes opposite roles). Moreover, selective galectin ligands with improved affinity would also be interesting molecules with therapeutic applications. Although it is known that galectins bind carbohydrates through a carbohydrate recognition domain (CRD), a deeper knowledge of the interactions that take place and the search of new binding pockets to anchorage putative modulators, are key factors for the development of galectin modulators.[6a] Different strategies have been followed to obtain glycomimetics targeting galectins with improved affinity and selectivity. Given that galectins specifically recognize β-galactosides, an evident starting point has been the modification of the galactose moiety to produce different monosaccharides. In particular, changes on the anomeric position and position 3 of the galactose ring have been performed. Nevertheless, the most used strategy has pursued the modification of disaccharide scaffolds, mainly the lactose. Chemical modifications include: the substitution of the glycosidic linkage in order to improve the glycolytic stability; the insertion of relatively long linkers on position 1 in order to have multivalent systems; the introduction of substituents in position 3 of the galactose moiety and/or positions 2 and 3 of the glucose moiety; and the use of other disaccharide scaffolds, such as the digalactose[9]. To date, a large number of gal-1 and gal-3 inhibitors have been synthesized, although it is scarce the information about selective gal-7 inhibitors. In general terms, there are three types of described gal-1 and gal-3 that can be

classified as monosaccharide, disaccharide and non-carbohydrate inhibitors. Most of monosaccharide derived inhibitors are based on galactose, since it is the natural moiety galectins can bind, and on mannose or talose, since they are epimers of galactose in carbons 2 and 4. Substitutions on carbon 1 or 3 of these glycans, in order to maintain the nature of the galectin-carbohydrate interaction, are the most common substitutions in the galectin inhibitors design field. The vast majority of these substitutions are joined to the desired glycan through an ether, amide or triazol linkers. Disaccharide inhibitors are based mostly on lactose and LacNac, due to the higher affinity compared to monosaccharide ligands, but also on seleno and diseleno glycans, given that the use of these compounds can improve the bioavailability of the resulting binder. For both mono and disaccharide inhibitors, the modifications on different carbons of the sugar are based on aromatic substituents, due to the presence of different arginine residues in the adjacent pockets, in the case of gal-3, that can interact with this type of substituents through a cation-π interaction. Different modifications of these types have been explored, based in benzene, naphthalene, quinolones, steroids, etc. Simultaneous modifications have also been explored in both mono and disaccharide inhibitors, combining C3 and C1 modifications, in order to reach to more pockets and thus increase the affinity. In the case of non-carbohydrate ligands, there have been described only peptide and peptidomimetics as inhibitors of galectin-1, although their binding mode has not been elucidated.

In rational drug design it is evident that structural information is of paramount importance.[10] The first X-ray structure of Gal-3/Lactose complex was the starting point of the rational drug design of galectin inhibitors because the groove extending from galactose O3' was perfect spot to accommodate new chemical entities and enhance affinity and selectivity. Another example is , the knowledge of

the conformational change of Arg144 in gal-3, in presence of aromatic ring establishing a cation-π interaction, has permitted the synthesis of several gal-3 inhibitors.[11] However, to the best of our knowledge, there are no computer-aided studies for the design of galectin ligands with improved selectivity and affinity. In particular we are interested in the identification of novel scaffolds able to lead to second generation of lactose derivatives with different selective binding properties towards human galectins -1, -3, and -7 that would be helpful for the study of galectin mechanism and would be interesting as putative galectin modulators. We here report the design, synthesis and biophysical studies of the binding properties of a series of lactose derivatives with improved selectivity and affinity towards galectins 1-, 3-, and 7-. The computer-aided design of the galectin ligands has followed two steps: i) a fragment-based virtual screening to identify the best moieties able to be accommodated in the adjacent pocket to the CRD of each galectin, and ii) a computational study of the designed lactose derivatives to select those candidates for synthesis with better predicted binding properties. Several compounds have been synthesized and studied by ITC and NMR techniques. Our studies have led to the design of novel ligands with improved predicted binding affinity, and, putative selectivity towards gal-1, -3 and -7. These studies can assist in the further development of synthetic glycans with potential therapeutic applications.

3.2 Results

Design of selective galectin ligands. From the study of galectins is very easy to underlines how CRD could be very similar among protein of the same family. They shared some common features such as the presence of an arginine, a histidine and a tryptophan. So we start to explore the adjacent binding site of the CRD which is already known to have some conformational changes depending on the

ligand interacting with them, such as the conformational change of the Arg144 in the galectin 3 in the presence of a ligand with an aromatic portion.[12] So, we focus the attention in particular on galectin -1, -3 and -7 which seem to be very interesting as therapeutical targets. The necessity to aim to a specific galectin is mandatory. Given that β-galactose is the molecular pattern characteristic of the specific recognition of all galectins, this moiety is usually maintained in the reported synthetic glycomimetics aiming to target galectins. Moreover the common natural ligand for galectins is a lactose or lactosamine. Therefore, we focused our attention in the adjacent pocket of the CRD in order to explore the possibility to identify specific moieties with good binding properties. We explore this region through binding site pocket finder software (SiteMap) (Figure 1).[13]

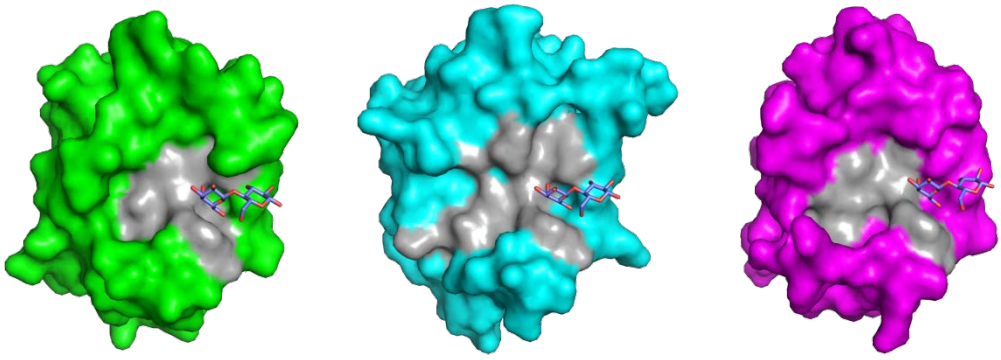


Figure 1: Surface representation of gal-1 (PDB ID 1GWZ, green), gal-3 (PDB ID 3ZKJ, cyan) and gal-7 (PDB ID 4GAL, magenta).

Protocol for the design

We follow a protocol summarized in Figure 2 that can be divided into the following steps.

Virtual screening of fragments in the vicinal pocket.

The Maybridge fragment library was screened in the vicinal pocket of all available galectins. We aimed to identify the best anchored fragments in this pocket for each

of the seeked galectines -1, -3 and -7. Additionally, pocket X of galectins 2, 4, 8, 9, and 10, were also screened to discard fragments with high affinity towards these "unwanted" galectins, focusing only on those fragments with the highest possible affinity and selectivity towards the X pocket of galectins 1, 3 and 7. The Maybridge fragment library provides structural diversity and drug-like profile, including experimental solubility.[14]

Selection of the best fragments for each galectin X pocket

More than 300 fragments were filtered as the best possible fragments in term of binding energies, and analysed to finally yield 25, 32, and 30 selected fragments for each galectin (1, 3, and 7 respectively).

Docking of the designed lactose derivatives

Based on the selected fragments, 106 different OMe-lactose derivatives were docked into the carbohydrate binding domain of the eight galectins. Analysis of the docking results allowed the final selection of 25 compounds as candidates to be synthesized.

Synthesis and biophysical binding studies

The synthesis of some of them was performed in collaboration with Prof. Menéndez (UCM, Madrid, Spain) and Prof. Oscarson (UCD, Dublin, Irlanda). Biophysical assays were performed by us (ITC, at the CIB-CSIC with collaboration of Prof. Margarita Menéndez from IQFR-CSIC) and in collaboration with Prof. Jiménez-Barbero (NMR at CIC bioGUNE).

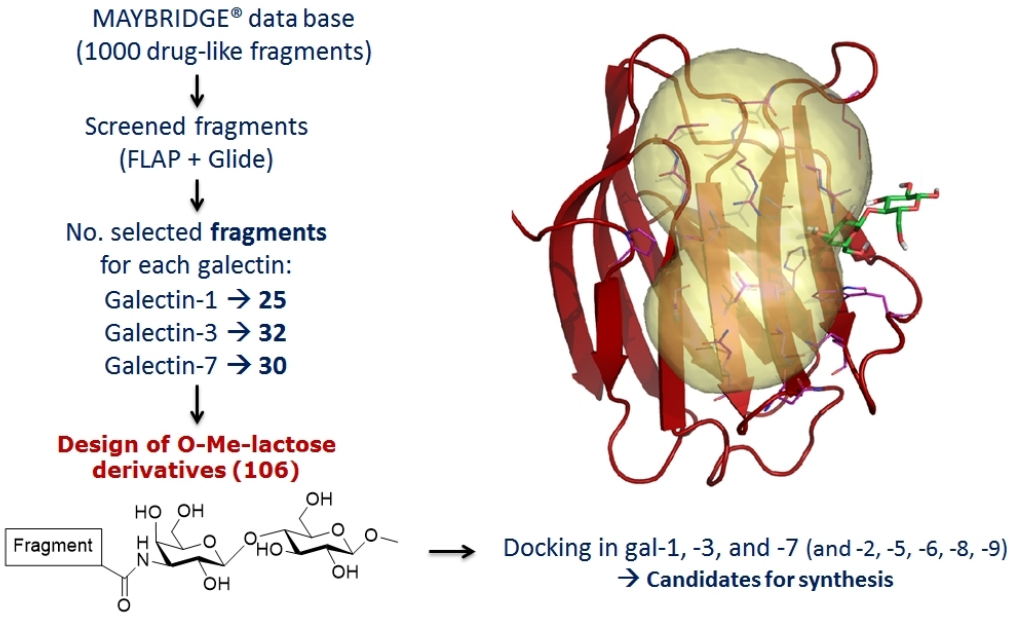


Figure 2: Left: Scheme of the protocol applied for the virtual screening on each galectin. Right: Cartoon representation of gal-3, bound to lactose. Yellow sphere represents the merge of two pockets used to perform virtual screening with the Maybridge database.

Virtual screening of fragment library. Keeping the lactose as in the crystal binding pose, we performed a structure-based virtual screening in the adjacent region using two different protocols (FLAP and Glide). In particular, we used the area delimited by Ser29-Lys63-Tyr119 for gal-1, Arg144-Lys176-Lys233 in gal-3 and Lys6-Arg31-Lys64 for gal-7. The surface area of these pockets was approximately 1200 $Å^3$. The fragment library used in this work was the Maybridge data base, which contains more than one thousands drug-like fragments. The choice of the fragments was dictate from the best scoring marks.

From FLAP virtual screening, two amino acids per galectin were chosen to determine the centre of two different pockets that were merged together, in order to study the complete adjacent pocket. Pocket extension and thickness were 8 and 26 for all the cases, respectively, and high accuracy performance option was chosen. Best fragments, i.e. fragments with a GlobScore of more than 0,9 (26 for

gal-1, 32 for gal-3, and 30 for gal-7), were considered to build the methyl-lactose (OMe-Lac) based ligands. In the case of Glide, VS protocol comprises 3 levels of performance: High-Throughput Virtual Screening (HTVS), Standard Precision (SP) docking and Extra Precision (XP) docking. In every step, only 30% of the fragments were kept. By applying this protocol, 26 fragments were asked to Glide to be selected for each galectin -1, -3 and -7, starting from the Maybridge database. An average of 25% of the fragments was repeated in more than one galectin by both FLAP and Glide VS protocols.

In total, taking into account both results from FLAP and Glide, and alternative binding poses and possible linkage connections with the lactose moiety, more than 300 possible fragment poses were analyzed in order to build the full OMe-Lac ligands based on these fragments. Only fragments with the appropriate orientation and position were considered, i.e. binding pose of the fragment near the lactose binding site and the linkage position oriented towards position 3 of galactose residue (Figure 3). Finally, 106 fragments were selected to build OMe-Lac based ligands.

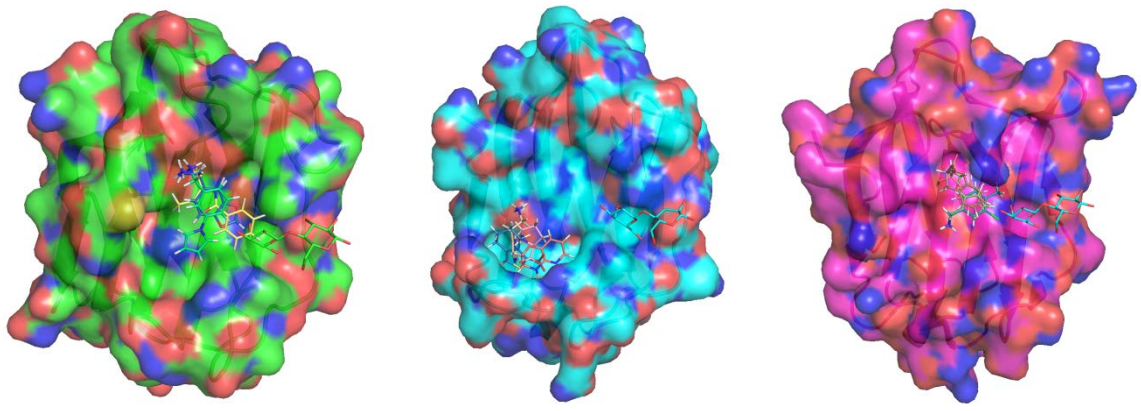


Figure 3: Representative results of the fragment-based virtual screening in gal-1 (PDB ID 1GWZ, green), gal-3 (PDB ID 3ZKJ, cyan) and gal-7 (PDB ID 4GAL, magenta).

Docking of the designed methyl-lactose derivatives

We designed the ligands based on the OMe-Lac scaffold and more than one hundred (106) structures were built using the best binding fragments. The three-dimensional structures were then prepared (minimized and charged) to docking purpose. To validate the protocol and to estimate an improvement of affinity we firstly perform docking studies by means of AutoDock4 with OMe-Lac. Then the new 106 ligands were docked in gal-1 (PDB ID 1GZW), gal-3 (PDB ID 3ZSJ and 1KJR) and gal-7 (PDB ID 4GAL). Moreover, the designed compounds were submitted to the same docking protocol with the other human galectins (gal-2 PDB ID 5DG2, gal-4 PDB ID 4XZP, gal-8 PDB ID 4BMB, gal-9 PDB ID 3NV4 and gal-10 PDB ID 1G86) to discard those ligands with putative affinity towards other galectins. Overall, we found that the designed compounds presented worst theoretical binding energies for galectins different t gal-1, 3-, 1-, and 7-.

Focusing on galectin -1, -3 and -7 we compared the predicted docked poses for each designed compound with each galectin. For each compound the lowest energy lactose-like binding pose was taken into consideration. The choice of best candidates for synthesis was dictated by different criteria. Among lowest-energy binding poses we chose those with an improvement of energy in comparison with the predicted binding energy of the OMe-Lac for the same galectin and with a worsening of the binding energy for the other galectins, comparing with the OMe-Lac predicted binding energy. Among the best candidates, we chose compounds J013, J228 and J683 given the synthetic accessibility of the corresponding moiety to be condensed to the amino-lactose (Figure 4). Reference compounds L2 and L3[12] and J009 were synthesized, whose fragments were commercial available. The compounds were modelled and the docking calculations were performed. These compounds were used as a "control".

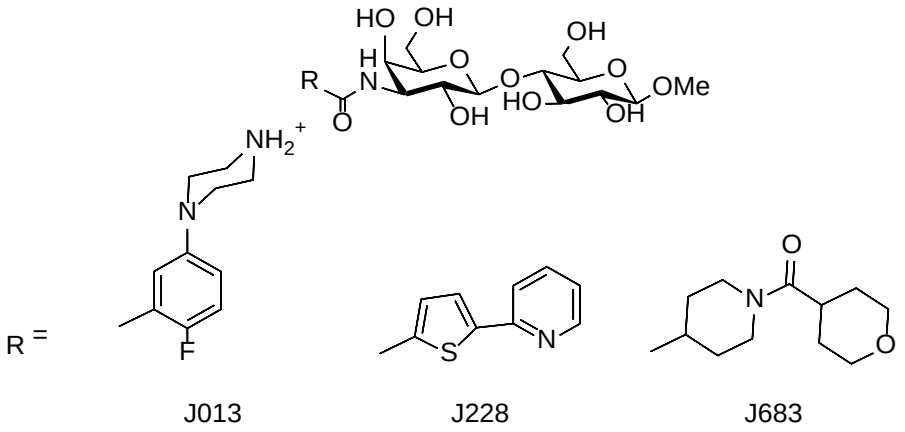


Figure 4: 2D representation of three of the best candidate compounds obtained by crossing the two VS protocols (FLAP and Glide).

According to the docking study the compound J013 show an improvement of the affinity due to the ionic interaction between the ammonium group with Arg123 and Arg125 (Gal-1). The docking study of the compound J013 with galectin-1 shows a considerable improvement of the predicted affinity due to the ionic interaction between the ammonium group present in the piperazine and the carboxylic group of Asp124 and Asp126 (Figure 5). The presence of the lactose moiety keeps the "classical" interactions with the protein. In particular the stacking interaction between the galactose ring and the Trp69, and the hydrogen bonds established by the polar groups of the glucose and galactose with His53, Arg49, Glu72, Asn47, His45, and Asn62.

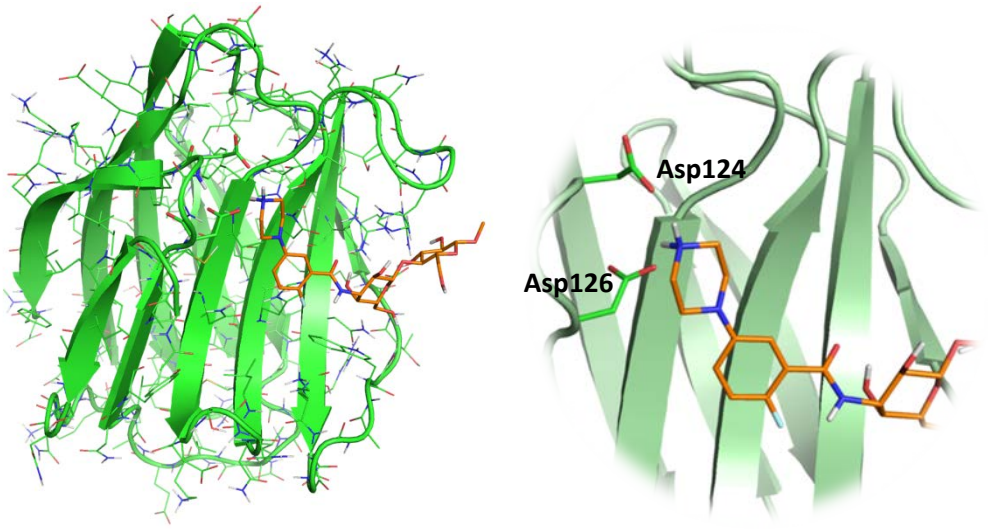


Figure 5: Docked pose of compound J013 into galectin-1.

These interactions in fact, are conserved during all the MD simulation and they strongly contribute to the binding, as possible to note from the MMPBSA calculation (Figure 6).

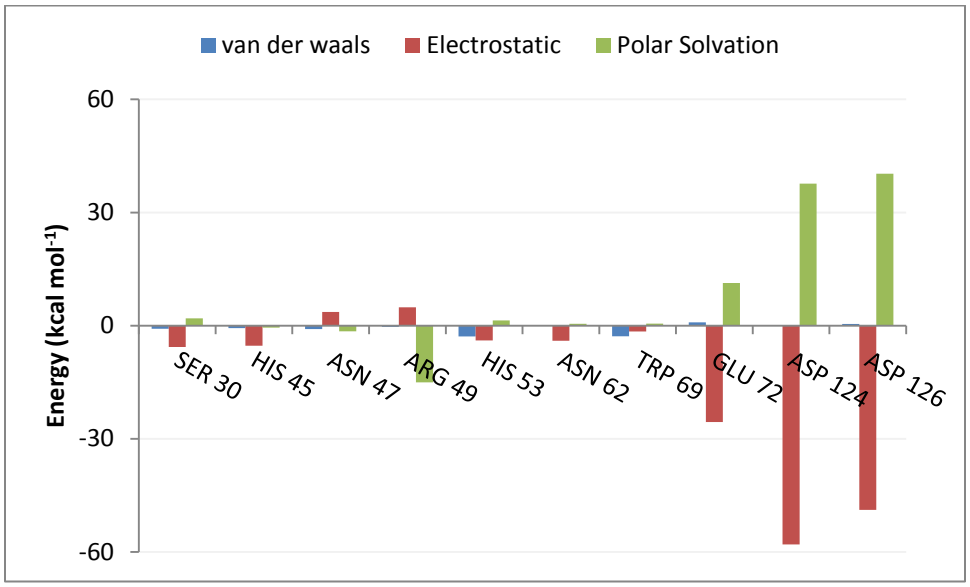


Figure 6: Representation of the contributions to the free binding energy of the gal-1/J013 complex based on MMPBSA calculations.

The compound J228 shows selectivity for the galectin-3 due to the cationic-π interaction with the Arg144, which suffers a conformational change due to the presence of the aromatic portion in the ligand. The docking study of the compound J228 shows a selectivity of this ligand for the galectin-3 in particular for the presence of an aromatic portion in the new fragment that establishes a cationic-π interaction with the Arg144, which suffers a conformational change due to the presence of the aromatic portion in the ligand (Figure 7). This kind of behaviour was already observed with previous ligand that presented an aromatic portion attached to the position 3 of the galactose.[12] Actually all the dockings were performed in the two different crystal conformation of the Arg144 and the predicted binding energies were compared.

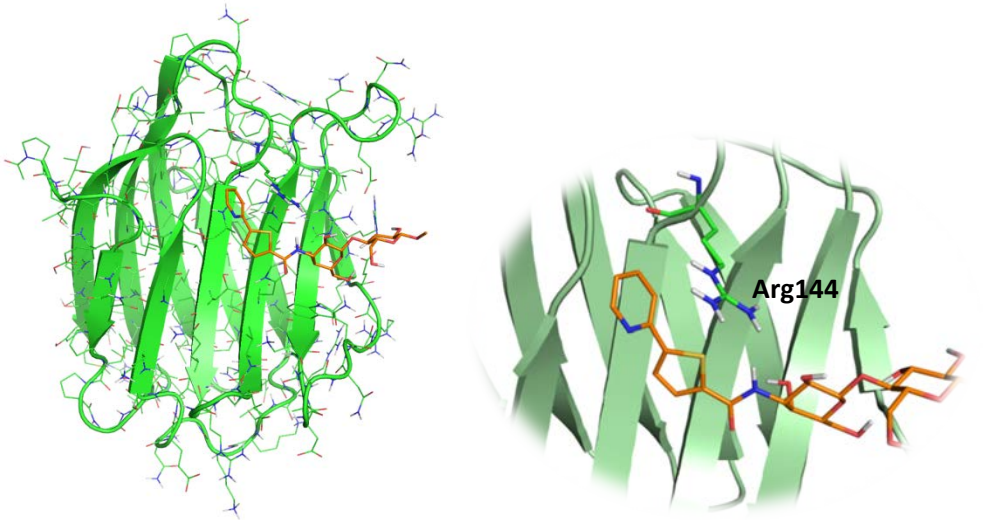


Figure 7: Docked pose of compound J228 into galectin-3.

Apart of the new interactions established by the new fragment, the lactose moiety still conserves the interactions observed in the crystal structures. In particular, the stacking interaction between the galactose moiety and Trp181 and the polar interactions between the OH groups of the sugars and the side chains of Arg162,

His158, Glu184. Also from the MMPBSA calculation is evident how these interactions are important for the binding (Figure 8).

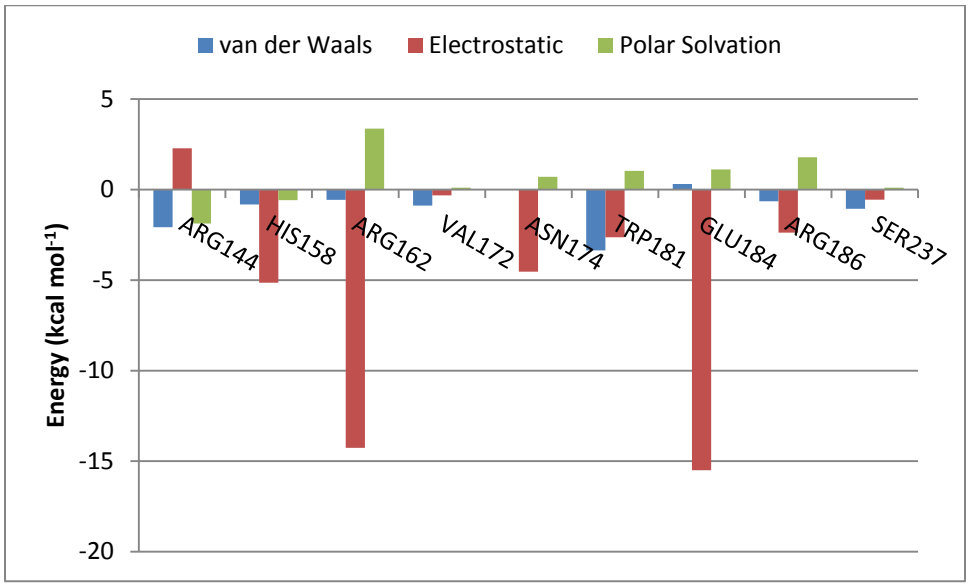


Figure 8: Representation of the contributions to the free binding energy of the gal-3/J228 complex based on MMPBSA calculations.

The compound J683 shown an improvement of the affinity due to the ability to achieve a polar portion delimited by three asparagines (3, 36, 52) and the Arg31 in the adjacent pocket (Gal-7). The compound J683 shown an improvement of the affinity due to the ability to achieve a polar portion delimited by three asparagine (23, 36, 52, Figure 9) and the Arg32 in the adjacent pocket (Gal-7). Nevertheless also the lactose moiety interacts with the principal binding pocket through CH-π staking with the Trp70 and polar interactions between the OH groups of the galactose and glucose with the polar groups of the side chain of the Arg32, Arg54, His50 and Glu73. These residues mainly contribute to the binding as shown from the MMPBSA calculation (Figure 10).

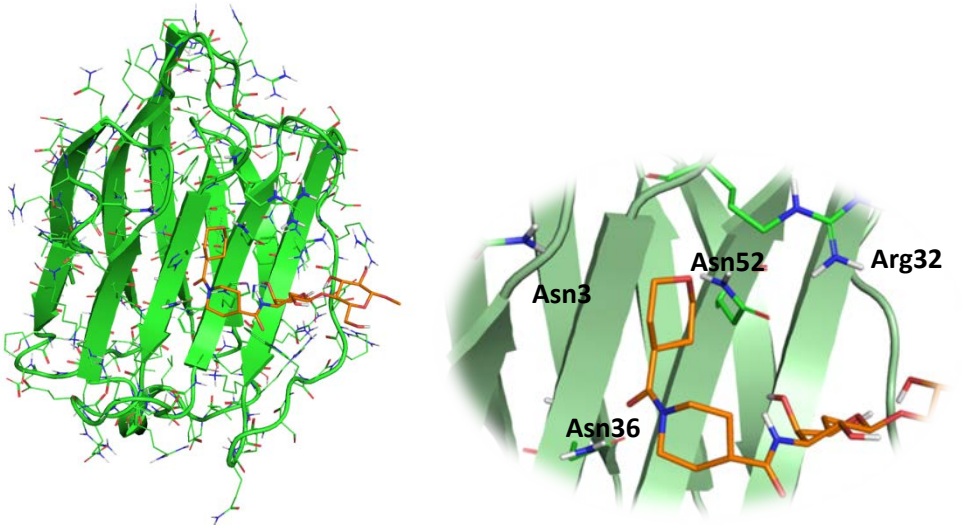


Figure 9: Docked pose of compound J683 into galectin-7.

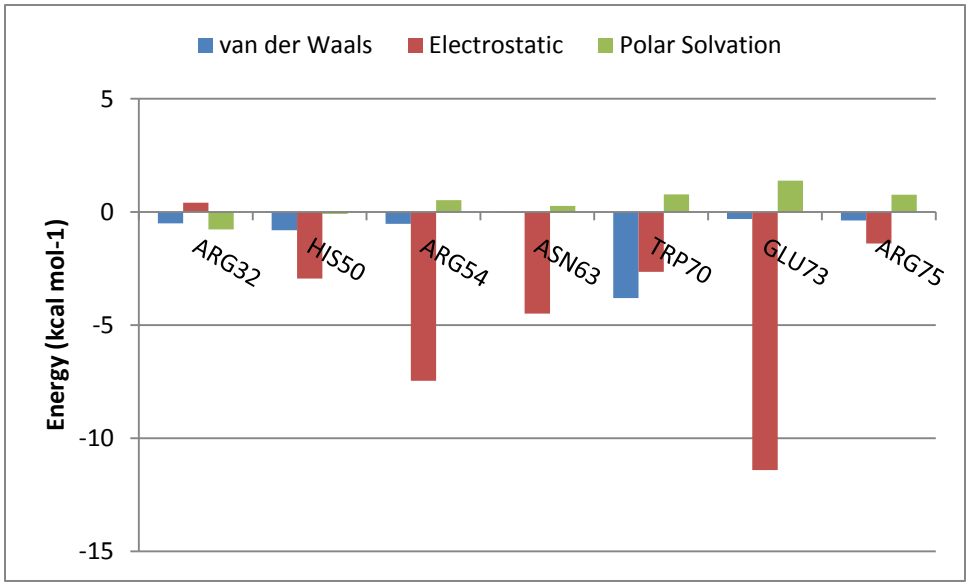


Figure 10: Representation of the contributions to the free binding energy of the gal-7/J683 complex based on MMPBSA calculations.

NMR and ITC results

For compound J013 a difference in terms of binding has been register. In fact, depending on method use to evaluate the affinity the results were different. From the STD experiments (data not shown), this compound shows a very good affinity, 4.5 times better than lactose, whilst the measure by ITC shows an affinity improved of 1.5 times in comparison with lactose (data not shown). Also the selectivity profile changes depending on the technique. In fact with NMR no major differences among the three galectins were observed. On the contrary, ITC measurements show that J013is very selective towards galectin-3. This result could be explained by a CH-pi interaction between the phenyl group and the Arg144 side chain. In fact, the STD results clearly show the involvement of the fragment moiety in the binding. According to the docking calculations, the compound J228 shows an improvement in the affinity/selectivity profile for galectin 3. This improvement is basically due to the presence of the tiophene ring which can establish a CH-pi interaction with Arg144 side chain. The involvement of the fragment in the binding is clearly shown in STD experiment (data not shown). Nevertheless the binding towards galectins 1, and 7 is still very good, although this molecule binds preferentially to galectin 3 (not very selective). Regarding compound J683, this ligand does not show selectivity but, interestingly, shows an improved affinity towards galectin 7 in comparison to lactose. This is an interesting property not gained for any reported galectin 7 modulator to the best of our knowledge.

3.3 Conclusions

Selective modulators of galectins 1, 3 and 7 have been computationally designed based on a lactose core from a fragment-based virtual screening protocol. NMR and ITC experiments show results clearly indicating improved affinity and

selectivity for some of them in comparison with the lactose reference compound. These novel compounds can be useful in cellular assays for the understanding of the cellular roles, especially in cancer, of these important galectins.

3.4 Material and methods

Computational Methods.

Protein Preparation. More than 150 X-ray crystallographic structures of human galectins are available at the Protein Data Bank[15]: 27 structures for gal-1, 4 for gal-2, 40 for gal-3, 12 for gal-4, 15 for gal-7, 28 for gal-8, 15 for gal-9, and 4 for gal-10. For our studies we used the X-ray structures with the following PDB IDs: 1GZW for gal-1, 5DG2 for gal-2, 3ZSJ and 1KJR for gal-3, 4XZP for gal-4, 4GAL for gal-7, 4BMB for gal-8, 3NV4 for gal-9, and 1G86 for gal-10. In the case of gal-3, two different structures were used in order to take into consideration the conformational change of the Arg144. As shown in Figure 11, the presence of a ligand with an aryl group in position 3 of the galactose moiety induces a change in the conformation of the Arg144 side chain and the establishment of CH-π interaction, increasing the affinity of the ligand towards the protein.[28]

The proteins were prepared using the Protein Preparation Wizard tool included in Maestro.[16] Water molecules, co-factors of crystallization and ligands were removed, missing atoms and cap termini were added, side chains and loops were filled by Prime, and hydrogens were added with Epik[17] at physiological pH. For the preparation of the proteins, the protonation of the main histidine in the CRD considered as HID. The final structure of the proteins was minimized with OPLS 2005 force field as implemented in Maestro with implicit solvent (water). The final minimized structures were used for docking purposes.

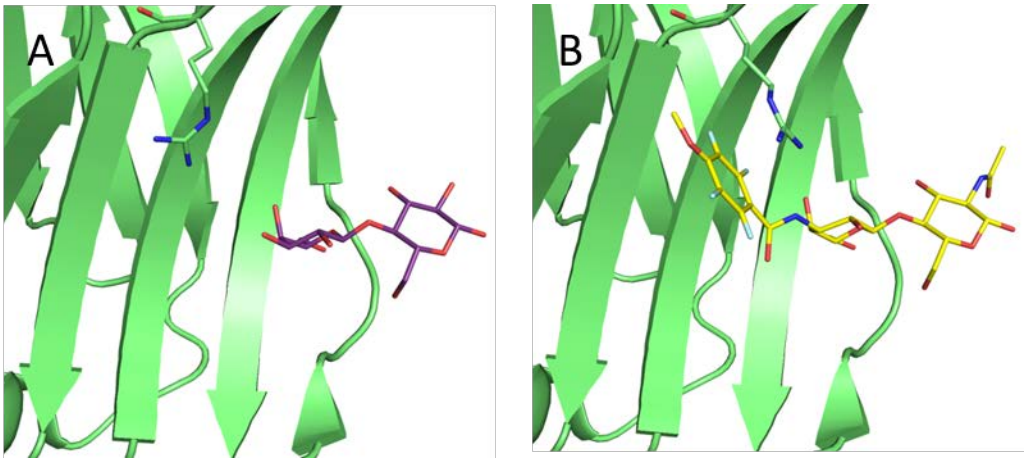


Figure 11: X-ray crystal structures of gal-3 (in green) with different ligands. Left: gal-3 complexed with lactose (represented in purple sticks, PDB ID 3ZSJ). Right: gal-3 complexed with L3 (represented in yellow sticks, PDB ID 1KJR). Arg144 in both crystal structures is shown in sticks.

Virtual Screening of Maybridge database. To perform the virtual screening in galectins -1, -3 and -7 Maybridge database was imported. Maybrige database consists of 1002 different drug-like fragments of different chemical composition and properties. Additional entries were added for the L2 benzene fragment, L3 tetrafluorated benzene fragment and a naphthalene fragment as reference fragments with reported experimental data.[12] The protonation state of every fragment at physiological pH was calculated by Epik. Two different virtual screening protocols were applied by means of two VS programs: FLAP[18] and Glide.[19]

VS Protocol with FLAP program. The pocket point radius was 2.0 Å and high accuracy performance was chosen. To explore only the adjacent accessory pocket of each galectin, we took in consideration the "lactose space" setting as macromolecule the complex of each galectin with lactose. Table 1 shows the two aminoacids chosen as the centre of the pocket of each galectin. The pocket extension and thickness were 8 and 26 for all the cases, respectively.

Protein	Galectin-1	Galectin-3	Galectin-7
Centric aminoacids	Val31/Asn33	Arg144/Asp148	His34/Arg32

Table 1: Chosen aminoacids for the pocket centre to perform virtual screening in galectin -1, -3 and -7

The VS protocol was validated using Chembl database.[20] For the validation protocol we use the gal-3. In Chembl database we found 62 known as active gal-3 binders and 48 as inactives. So, we performed a VS in the pocket centred in the Trp181 with extension and thickness of 15 and 5, respectively.

Building of the designed compounds. To build the designed compounds, OMe-Lac (Galβ1-4Glcβ-OMe) scaffold was used (Figure 12). The OMe-Lac moiety was obtained by means of GLYCAM[21] carbohydrate builder tool. The selected fragments from VS studies (32, Glob-Prob score up to 0.9 for FLAP + Glide) were attached through an amidic bond to the position 3 of the galactose. These 106 new ligands were then minimized with MMFFs force field with Macromodel[22] and Gasteiger charges were added with AutoDockTools.[23]

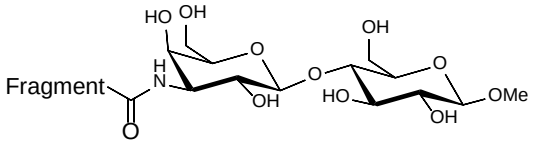


Figure 12: General structure of the designed compounds

Validation of the docking protocols. Different docking protocols were applied (described below): AutoDock4,[23] VINA[24] and Glide.[19] In order to validate the docking protocol, we docked lactose in gal-3, obtaining results in full agreement with the X-ray crystallographic binding pose with the three docking programs

AutoDock4 (RMSD = 0.727 Å), Vina (RMSD = 0.906 Å) and Glide (RMSD = 0.686 Å).

Docking with AutoDock4. The same grid box coordinates and size for all the ligands were used in the same protein. The box size chosen was big enough to ensure that all ligands could be docked. A three-dimensional grid was defined (68 x 60 x 40 Å̊) centred on the centroid of the key residues from the CRD (Trp, His and Arg). In particular, Trp181, His158 and Arg162 were chosen for gal-3, Trp68, His44 and Arg48 for gal-1, and Trp69, His49 and Arg53 for gal-7. The grid spacing was 0.375 Å̊, and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock4 were also used. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 1.0 Å̊. The clusters were ranked by the lowest energy representative of each cluster.

Docking with VINA. The same gridboxes used with Autodock4 were employed also for VINA protocol. Twenty conformers for each ligand with each protein were generated.

Docking with Glide. The ligands were prepared by means of LigPrep tool in Maestro. Ligand minimization was performed with OPLS 2005 force field. The specified chiralities were retained as well as the ionization state. The same grid boxes in terms of size and centre were generated. Van der Waals scaling factor and partial charge cutoff were kept as default, which is 1.0 and 0.25, respectively. Unlike AutoDock4 and VINA, first a number of conformers were generated; the best poses were kept and posteriorly docked into the proteins. Standard precision Glide Docking was performed, maintaining all the default parameters except the partial charge cutoff of van der Waals radii, which was set in 0.5, to enhance the

contribution of the Trp interaction to the global binding energy. The output file was generated writing 5 poses per ligand conformation and the RMSD for the clustering was set in less than 2 Å. Calculations with Glide were only able to reproduce the crystallographic pose for lactose when two crystallographic water molecules were kept: the one close to Glu184 and Arg186, and the one close to Arg144 and Arg162. van der Waals scaling factor and partial charge cutoff were kept as default, which are 1.0 and 0.25, respectively. The other parameters were kept as default as well. The key residues from the CRD were used to generate the box centre position, and box size was 20x26x20 grid point, being the spacing between points of 1 Å. The output file was generated writing 5 poses per ligand conformation and the RMSD for the clustering was set in less than 2 Å.

MD simulation. The best docked solutions for each ligand were considered as starting geometries to perform the MD simulations. GLYCAM06, gaff, and ff14SB were used as force fields. All MD simulations were carried out by using the sander module in AMBER14. Counterions (Na^+ and Cl^-) were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 50 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an

ensemble of 5000 structures for each complex, which were finally analysed. Finally, the relative binding free energies of the ligand-galectin complexes were calculated by the MMPB/GBSA approach provided in the Amber14 suite. In the preparation of the MMPB/GBSA calculations, 500 snapshots were collected from the 50 ns MD simulations. From the trajectory file all counterions and water molecules were stripped and a salt concentration of 0.1 M. Poisson-Boltzmann (PB) model and three Generalized Born (GB) models were used to calculate the solvation free energy

Bibliography

1. (a) Thijssen, V. L.; Heusschen, R.; Caers, J.; Griffioen, A. W., Galectin expression in cancer diagnosis and prognosis: a systematic review. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2015,** *1855* (2), 235-247; (b) Liu, F. T.; Rabinovich, G. A., Galectins as modulators of tumour progression. *Nature Reviews Cancer* **2005,** *5* (1), 29-41.
2. Diaz-Alvarez, L.; Ortega, E., The Many Roles of Galectin-3, a Multifaceted Molecule, in Innate Immune Responses against Pathogens. *Mediators Inflamm* **2017,** *2017*, 9247574.
3. (a) Saussez, S.; Cludts, S.; Capouillez, A.; Mortuaire, G.; Smetana Jr, K.; Kaltner, H.; André, S.; Leroy, X.; Gabius, H.-J.; Decaestecker, C., Identification of matrix metalloproteinase-9 as an independent prognostic marker in laryngeal and hypopharyngeal cancer with opposite correlations to adhesion/growth-regulatory galectins-1 and-7. *International journal of oncology* **2009,** *34* (2), 433-439; (b) Grassadonia, A.; Tinari, N.; Iurisci, I.; Piccolo, E.; Cumashi, A.; Innominato, P.; D'Egidio, M.; Natoli, C.; Piantelli, M.; Iacobelli, S., 90K (Mac-2 BP) and galectins in tumor progression and metastasis. *Glycoconjugate Journal* **2004,** *19* (7-9), 551-556.
4. Henderson, N. C.; Sethi, T., The regulation of inflammation by galectin-3. *Immunological reviews* **2009,** *230* (1), 160-171.
5. Newlaczyl, A. U.; Yu, L. G., Galectin-3 - a jack-of-all-trades in cancer. *Cancer letters* **2011,** *313* (2), 123-128.
6. (a) Campo, V. L.; Marchiori, M. F.; Rodrigues, L. C.; Dias-Baruffi, M., Synthetic glycoconjugates inhibitors of tumor-related galectin-3: an update. *Glycoconjugate journal* **2016,** *33* (6), 853-876; (b) Wang, L.; Guo, X. L., Molecular regulation of galectin-3 expression and therapeutic implication in cancer progression. *Biomed Pharmacother* **2016,** *78*, 165-171.
7. Hughes, R. C., Galectins as modulators of cell adhesion. *Biochimie* **2001,** *83* (7), 667-676.
8. Cousin, J. M.; Cloninger, M. J., The Role of Galectin-1 in Cancer Progression, and Synthetic Multivalent Systems for the Study of Galectin-1. *International journal of molecular sciences* **2016,** *17* (9), E1566.
9. (a) Blanchard, H.; Bum-Erdene, K.; Bohari, M. H.; Yu, X., Galectin-1 inhibitors and their potential therapeutic applications: a patent review. *Expert Opinion on Therapeutic Patents* **2016,** *26* (5), 537-554; (b) Blanchard, H.; Yu, X.; Collins, P. M.; Bum-Erdene, K., Galectin-3 inhibitors: a patent review (2008–present). *Expert Opinion on Therapeutic Patents* **2014,** *24* (10), 1053-1065.
10. Delaine, T.; Cumpstey, I.; Ingrassia, L.; Mercier, M. L.; Okechukwu, P.; Leffler, H.; Kiss, R.; Nilsson, U. J., Galectin-inhibitory thiodigalactoside ester derivatives have antimigratory effects in cultured lung and prostate cancer cells. *Journal of Medicinal Chemistry* **2008,** *51* (24), 8109-8114.
11. van Hattum, H.; Branderhorst, H. M.; Moret, E. E.; Nilsson, U. J.; Leffler, H.; Pieters, R. J., Tuning the preference of thiodigalactoside- and lactosamine-based ligands to galectin-3 over galectin-1. *Journal of medicinal chemistry* **2013,** *56* (3), 1350-1354.

12. Sorme, P.; Arnoux, P.; Kahl-Knutsson, B.; Leffler, H.; Rini, J. M.; Nilsson, U. J., Structural and thermodynamic studies on cation-Pi interactions in lectin-ligand complexes: high-affinity galectin-3 inhibitors through fine-tuning of an arginine-arene interaction. *Journal of the American Chemical Society* **2005,** *127* (6), 1737-1743.
13. Halgren, T. A., Identifying and characterizing binding sites and assessing druggability. *Journal of chemical information and modeling* **2009,** *49* (2), 377-389.
14. Maybridge database. www.maybridge.com (accessed March 2019).
15. Wittebole, X.; Castanares-Zapatero, D.; Laterre, P.-F., Toll-like receptor 4 modulation as a strategy to treat sepsis. *Mediators Inflamm.* **2010,** *2010*.
16. *Schrödinger Release 2015-1: Maestro, version 10.1, Schrödinger, LLC, New York, NY*, 2015.
17. *Schrödinger Release 2015-1: Epik, version 3.1, Schrödinger, LLC, New York, NY*, 2015.
18. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S., A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** *47* (2), 279-94.
19. Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** *49* (21), 6177-96.
20. Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B., ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research* **2012,** *40* (D1), D1100-D1107.
21. 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.
22. *Schrödinger Release 2015-1: MacroModel, version 10.7, Schrödinger, LLC, New York, NY*, 2015.
23. 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.
24. 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.

CHAPTER 4:

Design of pseudo-dimannoside based glycomimetics with high affinity for DC-SIGN

4.1 Introduction

This chapter of the thesis was performed in collaboration with Prof. Anna Bernardi, at the Università degli Studi di Milano.

DC-SIGN is a transmembrane C-type lectin expressed at the surface of dendritic cells. It plays a key role in the recognition of several pathogens and in the development of various infections, including Dengue and the HIV virus.[1] Over the past decade, several glycomimetic ligands were designed to act as antagonists of DC-SIGN mediated viral infections using as template the pseudo-dimannoside (ps-diMan) scaffold (Figure 1), composed of a mannose ring connected to a conformationally locked cyclohexane diol.[2]

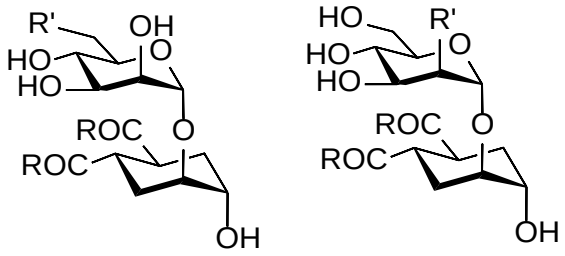


Figure 1: Left) Scaffold present in the literature for the design of DC-SIGN inhibitors. Right) Our proposal of modification in carbon 2, instead of carbon 6.

Here, we have identified by means of fragment-based virtual screening an alternative pocket adjacent to the mannose binding site where positively charged fragments are nicely hosted. This pocket can harbor fragments that can be linked to the carbon 2 of the mannose residue in order to design ligands derived from dimannoside, with improved affinity towards DC-SIGN (Figure 1). Thus, instead of modifications in carbon 6 of the mannose, here we have explored the possibility to perform modifications on carbon 2 (Figure 1). Designed pseudo-dimannoside compounds have been docked in several DC-SIGN protein structures, and some of them are proposed to be synthesized and tested as novel DC-SIGN inhibitors.

4.2 Results

Virtual fragment screening of Maybridge database[3] (more than 1000 thousand drug-like fragments) was performed using Glide on the 2XR5 DC-SIGN X-ray structure[2b], keeping in place the ligand from the X-ray crystallographic structure. More concretely, we focused on the region besides the mannose binding site, comprising residues Asn344, Glu358, Ser360, Asn365, Asp367, Phe313, Leu371 and Lys373. We chose that region because, after a deep observation and study of the protein, we realized that oxygen of carbon 2 of the mannose residue could act as a linker, given the pocket that is present between it and Phe 313 (Figure 2). Actually, this pocket also binds several natural oligomannosides.[4]

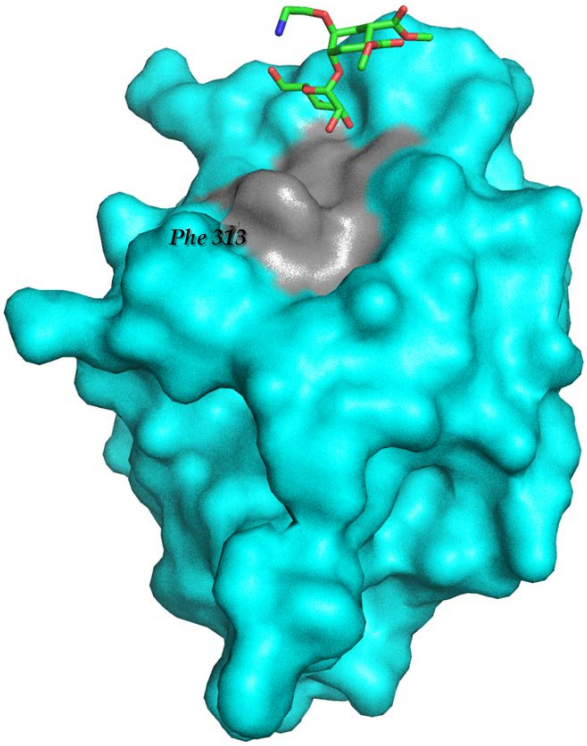


Figure 2: Representation of the X-ray structure of DC-SIGN/ps-diMan complex (PDB ID 2XR5). Alternative pocket (Grey) adjacent to the mannose binding site of DC-SIGN is shown in blue. Phe313 is highlighted and ps-diMan is depicted in green sticks.

Around 50 fragments resulted from this VS protocol, with predicted binding energies from -6.5 to -3.8 kcal/mol. Most of the fragments presented ammonium

groups in their structure, given the polar and negatively charged nature of the DC-SIGN binding pocket. Actually, the ammonium group is placed in a way that interacts with Glu358 through ionic interactions, Ser360 through H-bonds and with Phe313 through a cation-π interaction. Aromatic fragments are always present in the VS results, some containing nitrogen or sulfur in their structure. These fragments are close to residues like Phe313 and Asn344, stablishing π-π interactions. Although some fragments were predicted to bind in different parts of the protein, they were discarded for synthetic and accessibility reasons (Figure 3).

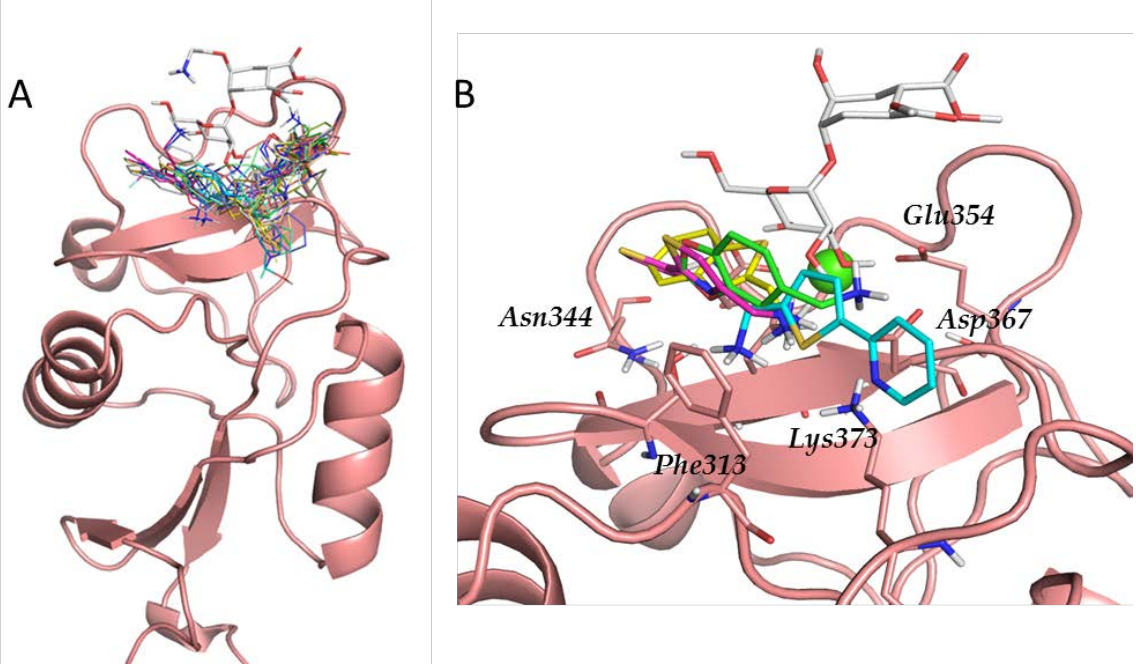


Figure 3: A) Selected fragments from the virtual screening. B) Selected superimposed fragments from the VS into DC-SIGN: fragments 23 (cyano), 250 (green) and 261 (magenta). Bound pseudo-dimannoside (from PDB ID 2XR5) is represented in grey sticks.

Based on the VS results, the suggestions in terms of synthetic accessibility and the availability of commercial fragments, we design 88 ligands with the scaffold of the pseudo-dimannoside molecule. The fragments were joined to the carbon 2 of the mannose residue through three types of linkers: alpha aminoacid linker (48 fragments), beta-aminoacid linker (22 fragments) and triazole linker (18 fragments) (Figure 4, annex figures 1-3).

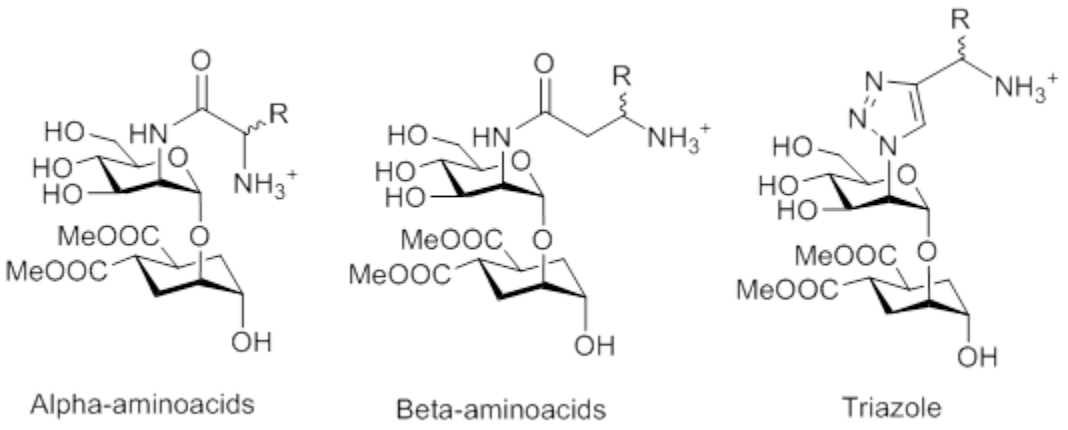


Figure 4: Designed ligands to be docked in DC-SIGN based on fragment-based VS.

In order to validate our docking protocol, the reference molecules for each type of ligand (i.e. the pseudo-dimannoside plus the linker with the ammonium group) were docked in order to discriminate if a possible increase in the binding affinity is due to the presence of fragment or the linker. In all three cases, the reference ligands interact with DC-SIGN in a very similar way than the reported crystal structures.[2b, 5] The mannose moiety is the major responsible of the binding, interacting with the calcium atom present in the binding site of DC-SIGN through oxygens of carbons 3 and 4, while the cyclohexane moiety is more exposed to the solvent. In interacts, though, with Val351. The results showed no appreciable difference in the binding energy between the reference structures (-2.6 kcal/mol for the α-aminoacid reference; -2.7 kcal/mol for the β-aminoacid reference; and -3.0 kcal/mol for the triazole reference) and the crystal ligand of 2XR5 crystal structure were docked as a validation of the docking protocol and to compare the predicted binding energies (Figure 5).

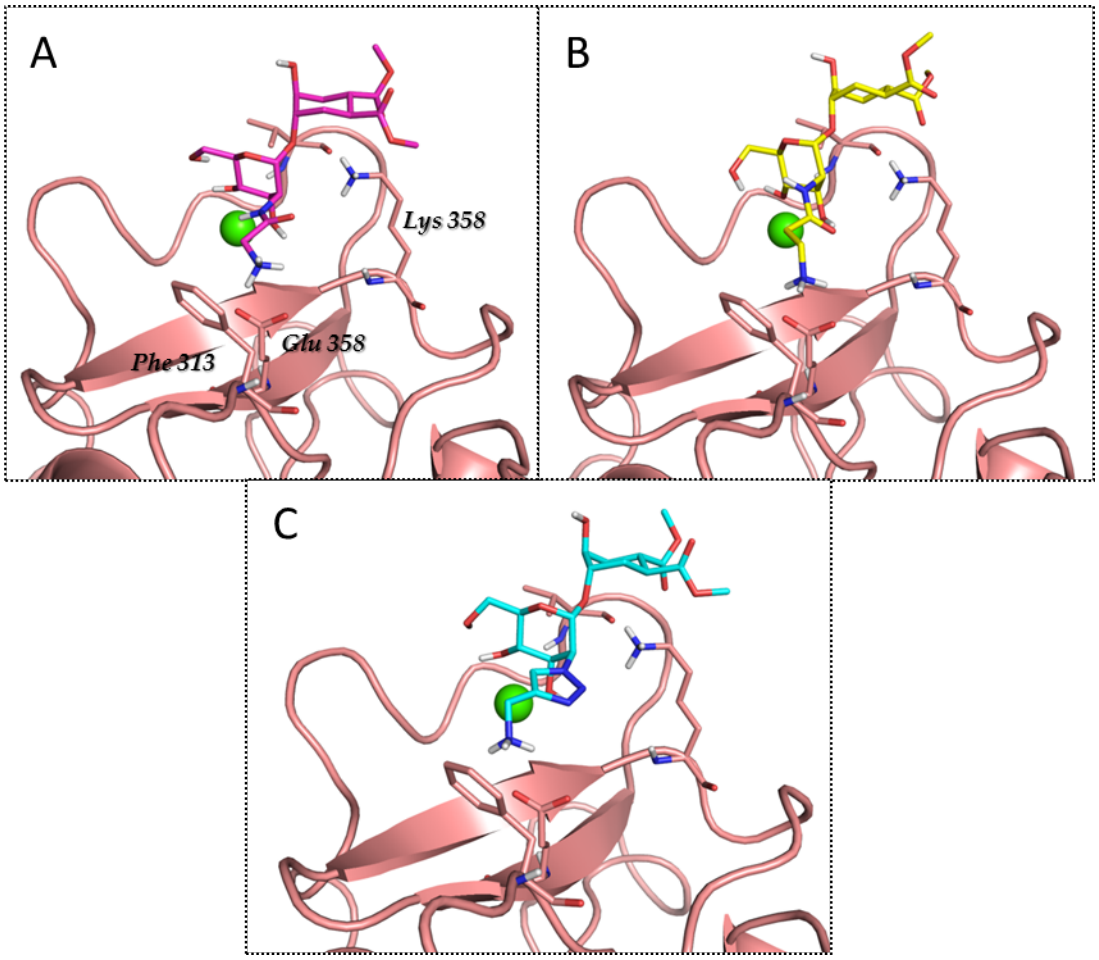


Figure 5: Docking results of the reference compounds of the α-aminoacid (A), the β-aminoacid (B) and the triazole (C) based compounds. Docked ligands and residues Phe313, Val351, Lys368 and Glu358 are highlighted in sticks. Calcium atom is shown as a green sphere.

We also performed the docking of compound Man069. Man069 is a compound designed by Bernardi and her coworkers based on our fragment-based design that showed a high binding affinity towards DC-SIGN (ITC results, data not shown). It is a triazole based compound with two hydroxymethyl-phenyl moieties attached to carbons 4 and 5 of the cyclohexane. The predicted binding pose was very similar than the crystal structure that they have also reported. Main differences can be observed in both hydroxymethyl-phenyl moieties, since the sugar part of the ligand is interacting with the protein like the reference compound. The differences

in the hydroxymethyl-phenyl moiety could be caused by possible disruptions of the structure in the crystal due to the interaction of this moiety with other DC-SIGN (Figure 6)

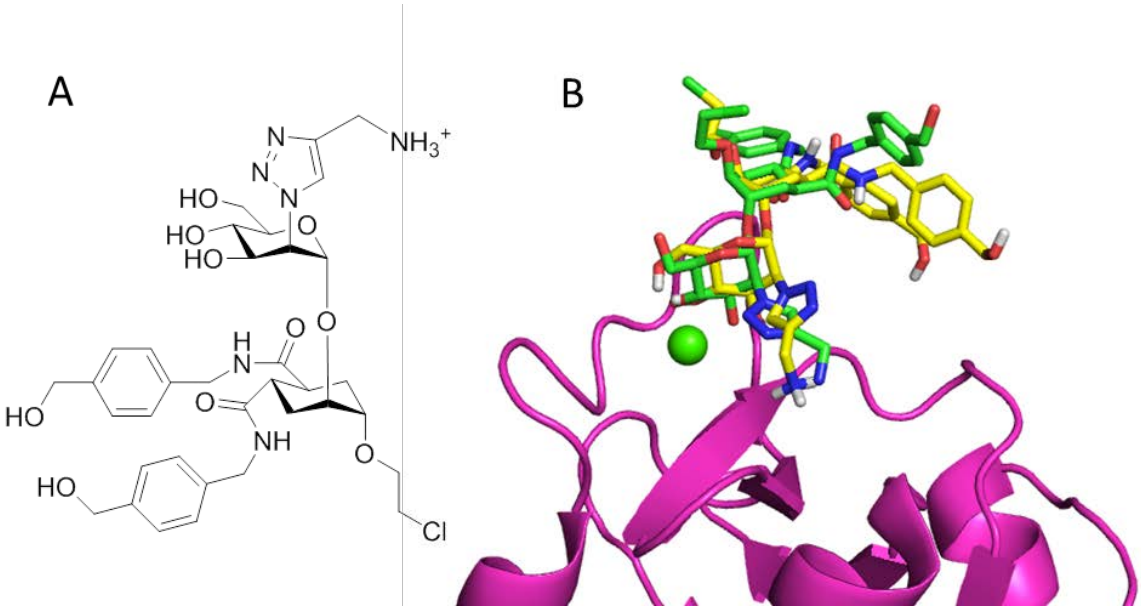


Figure 6: A) Structure of Man069 ligand. B) Superimposition of docked pose of Man069 (green) in DC-SIGN and the crystallographic pose from PDB ID 6GHV (yellow, PDB entry on hold until publication). Calcium atom is shown as a green sphere.

All these compounds were docked on DC-SIGN. Since there were two crystal structures of DC-SIGN protein on the PDB (2XR5 and 2IT5)[2b, 5], we decided to dock all ligands in both proteins. Although there are no major differences between these two crystal structures, subtle changes in the some aminoacids orientation can be found in the carbohydrate binding region (Figure 7). For example, Lys368, which is exposed to the solvent in the case of 2XR5, is oriented towards the ligand in 2IT5 crystal structure. This different orientation could lead to changes in the pose and in the predicted binding energy of the docking (Figure 7, top). Besides that, we also performed a molecular dynamics simulation of the free state of 2XR5 protein for two main reasons: to explore different conformations of the binding site and to identify the major conformation of Val351. Val351 is very close to the binding site and seems to interact with the 2XR5 crystal ligand but, in the crystal structure,

there are two possible conformation of this residue. After 20ns, we observed an aperture of the ammonium pocket, mainly due to the displacement of Phe313 (Figure 7, bottom). So, the docking was performed using these three proteins.

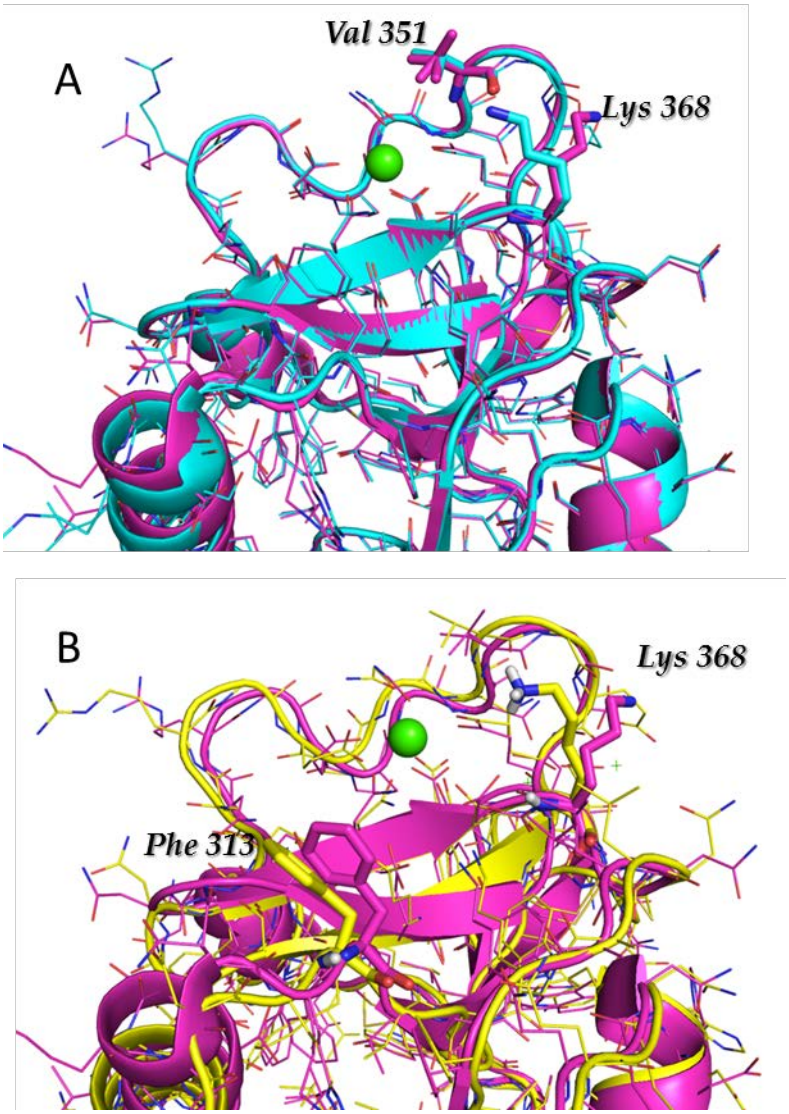


Figure 7: Superimposition of 2XR5 (pink) and 2IT5 (cyan) crystal structures (up) and 2XR5 with the average structure of 2XR5 MD simulation (2XR5_dyn, yellow, down). Calcium atom is shown as a green sphere.

Although in all three protein structures the desired conformation of the reference ligand (i.e. a pose with oxygens 3 and 4 of the mannose residue interacting with the calcium atom in the proper orientation) was achieved, best results in terms of energy were obtained on the average structure of the 2XR5 MD simulation (Annex

tables 1-3). The aperture of Phe313 seems to give more space for the binding of the ligands.

From all docked results, we selected as a good predicted poses the ones which the pseudo-dimannoside is placed and interacts with the protein in the same way than in the 2XR5 crystal structure.

From the docking results we suggested that the best candidates in terms of predicted binding energy would be the triazole based ligands, since its predicted binding energy is lower. Common features between all docked poses can be observed. Oxygens of carbons 3 and 4 of mannose residue are interacting with the calcium atom present in the binding site while the cyclohexane is clearly more exposed to the solvent, although an ionic interaction through its carboxymethyl moieties with Lys368 can be observed. The mannose moiety has another anchorage site, since its oxygen of carbon 4 stablishes a dual H-bond with the side chains of Glu347 and Asn349. On the other hand, the triazole moiety has the perfect size not only to fit into the ammonium pocket, but to expose the ammonium group to Phe313 and Glu358. Figure 8 show the example of compound 29R, one of the best predicted binders. The pseudo-dimannoside moiety is binding to 2XR5 in a very similar way than reference ligand, as well as the triazole and the ammonium. The pyridine fragment, due to the R conformation of the chiral carbon of the linker, reaches another part of the protein and interacts through a π-π stacking with Asn344. Best predicted ligands, apart from 29R, where ligands 26R_F and 28R_S, which a binding pose very similar to ligand 29R (Figure 8).

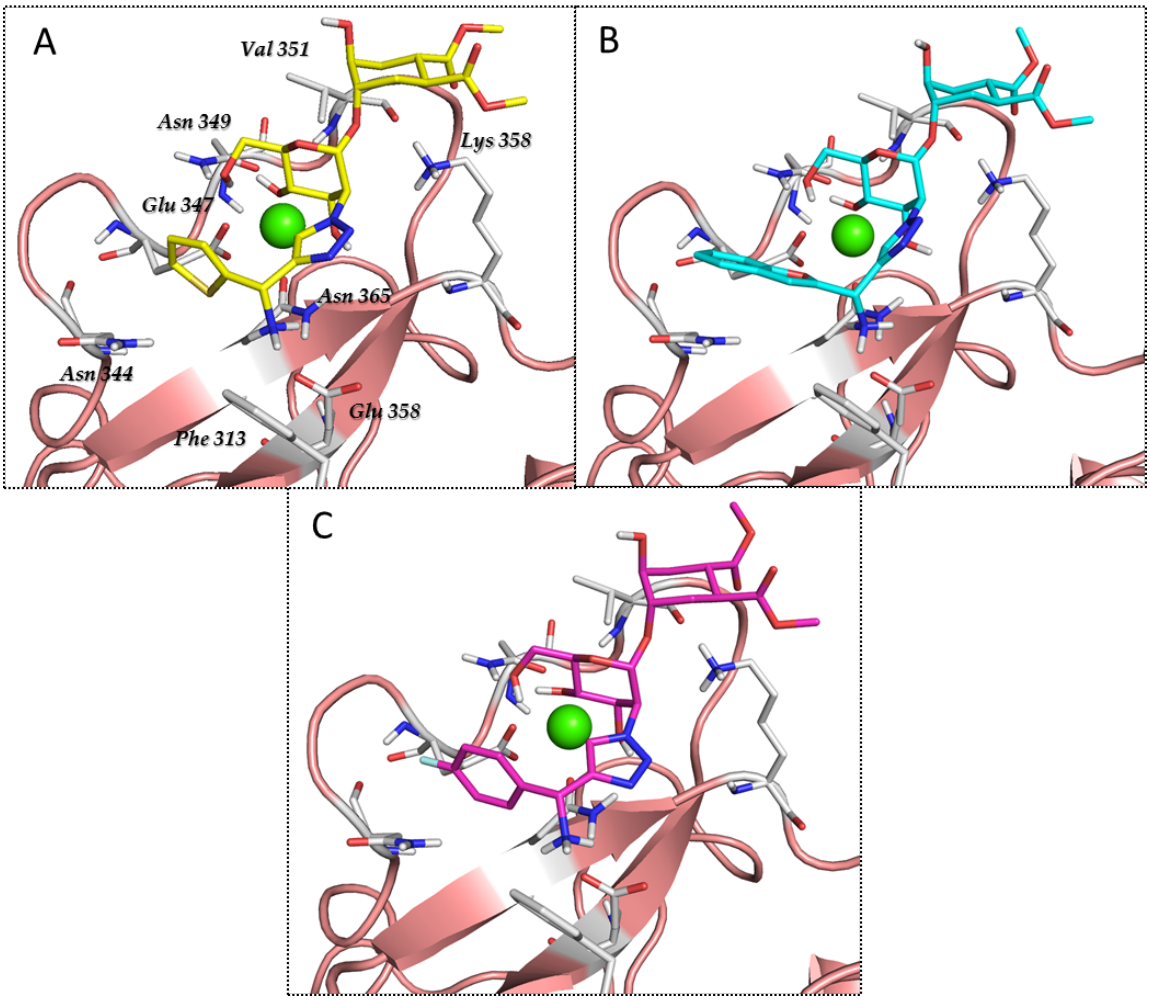


Figure 8: Examples of docked poses of compounds 28R_S (A), 29R (B) and 26R_F (C). Residues that participate in the interaction are highlighted in sticks. Calcium atom is shown as a green sphere.

MD simulations of the DC-SIGN complex with the reference ligand were performed. The mannose residue is kept in the staring bound orientation, while the cyclohexane moiety, more exposed to the solvent, changes its orientation during time. The ammonium moiety was kept in the same pocket, interacting with the initial residues Phe313 and Glu358 during the whole simulation, confirming to be a proper scaffold to hold the charged groups to bind in that pocket.

We also performed molecular dynamics simulation of both pseudo-dimannoside and man069 ligand on 2XR5 crystal structure. After 20ns, no major changes in both

ligand and protein can be observed. Mannose residues in both ligands remain stable and oxygens of carbon 3 and 4 interact during the whole simulation with the calcium atom present in the binding site. In the case of pseudo-dimannoside, the cyclohexane moiety rotates and changes its orientation compared to the mannose residue, but his results was expected because that part of the ligand is exposed completely to the solvent. This behavior is clearly observed in man069, where both 4-hydroxy-methyl-phenyl moieties are completely exposed to the solvent and the change in their orientation during time is relatively high.

4.3 Conclusions

In this chapter, we have studied the DC-SIGN protein in diverse levels. First of all, we have performed a fragment-based virtual screening of the Maybridge database in the 2XR5 crystal structure of DC-SIGN in order to identify additional pockets to design novel DC-SIGN inhibitors. With the results of the VS, we have designed 88 ligands based on a pseudo-dimannoside core with three different linkers: α-aminoacid, β-aminoacid and triazol linker. Then, we have docked these designed ligands into three structures of the DC-SIGN protein: the two conformations from the X-ray crystal structures (PDB IDs 2XR5 and 2IT5) and the average structure resulting from the MD simulation 2XR5 DC-SIGN conformation, representing the average from the conformational ensemble. These three conformations of the protein have revealed subtle changes between them that have been crucial in the understanding of how the designed ligands bind to DC-SIGN. We have revealed that Val351 side chain rotation has an important role in the recognition of DC-SIGN binders and that the motion of Phe313 enlarges the adjacent pocket. This enlargement is the reason why the designed ligands bind better to the average structure of the MD simulation. From the docking studies, we have selected some

compounds to be synthesized with novel scaffolds as glycomimetics with high affinity and DC-SIGN inhibitor activity.

4.4 Materials and Methods

Protein Preparation. 2XR5 and 2IT5 DC–SIGN crystal structures were extracted from the PDB. The proteins were prepared using the Protein Preparation Wizard tool included in Maestro. Water molecules, co-factors of crystallization and ligands were removed, missing atoms and cap termini were added, and hydrogens were added with Epik at physiological pH. Calcium atoms were left, since at least one of them is crucial for the binding of the ligands. The final structure of the proteins was minimized with OPLS3 force field as implemented in Maestro with implicit solvent (water). The final minimized structures were used for docking purposes.

Virtual screening. Virtual screening was performed using GLIDE VS Workflow.[6] 1000 drug-like fragment Maybridge database was screened on DC-SIGN 2XR5 crystal structure. The center of the grid box was placed at the middle point between Phe313, Glu358 and Asn365. Inner box size was 8x8x8 Å and the outer box size was 19x19x19 Å, including the whole adjacent pocket of the calcium binding site, the so-called "ammonium binding pocket". No excluded volumes were used in the grid preparation. Since the Maybridge database was previously optimized by our group, no ligand preparation was performed prior to the VS protocol. Scaling factor and partial charge cutoff were set as default (0.8 and 0.15, respectively) and Epik state penalties for docking were used. 20 % of the poses were kept in the first stage (High Throughput Virtual Screening, HTVS), 50 % in the second step (Standard precision, SP) and 50 % in the last step (Extra Precision, XP),

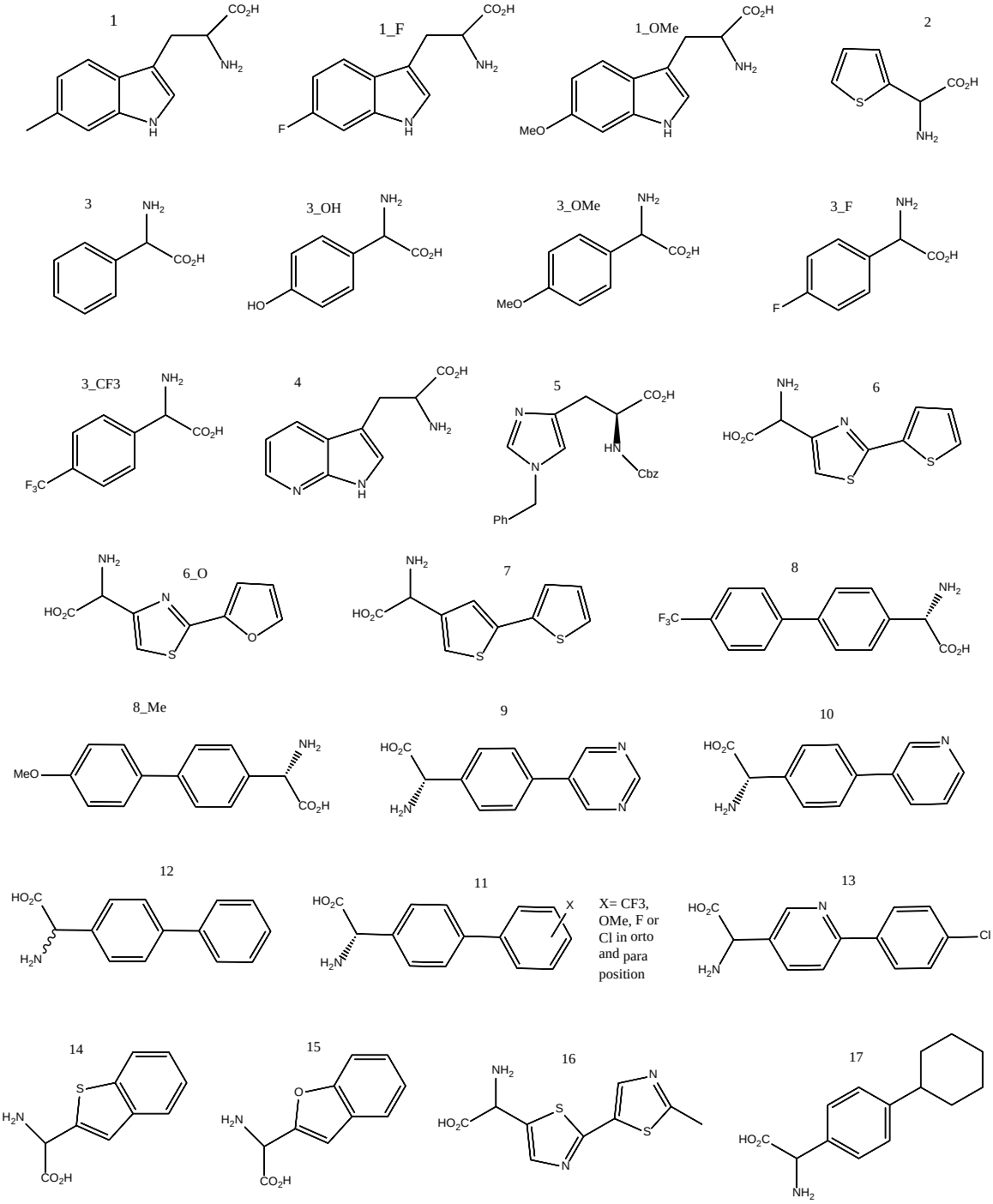
Docking. Docking of constructed ligands was performed using Autodock4 software.[7] Docking settings were set as default. The box size chosen was big enough to ensure that all ligands could be docked. A three-dimensional grid was defined (54 x 50 x 56 Å) centered on the centroid of residues Glu347, Glu354 and Asn365, which are at the center of the binding site . The grid spacing was 0.375 Å, and a distance-dependent dielectric constant was used. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock4 were also used. After docking, the 200 solutions were clustered in groups with root-mean-square deviation less than 2.0 Å. The clusters were ranked by the lowest energy representative of each cluster.

Parameters Derivation. Parameters for molecular dynamics simulation of the new units were set up with the standard Antechamber procedure. 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 Ambertools14 and AMBER14 with Antechamber assigning the general AMBER force field (GAFF) atom types.

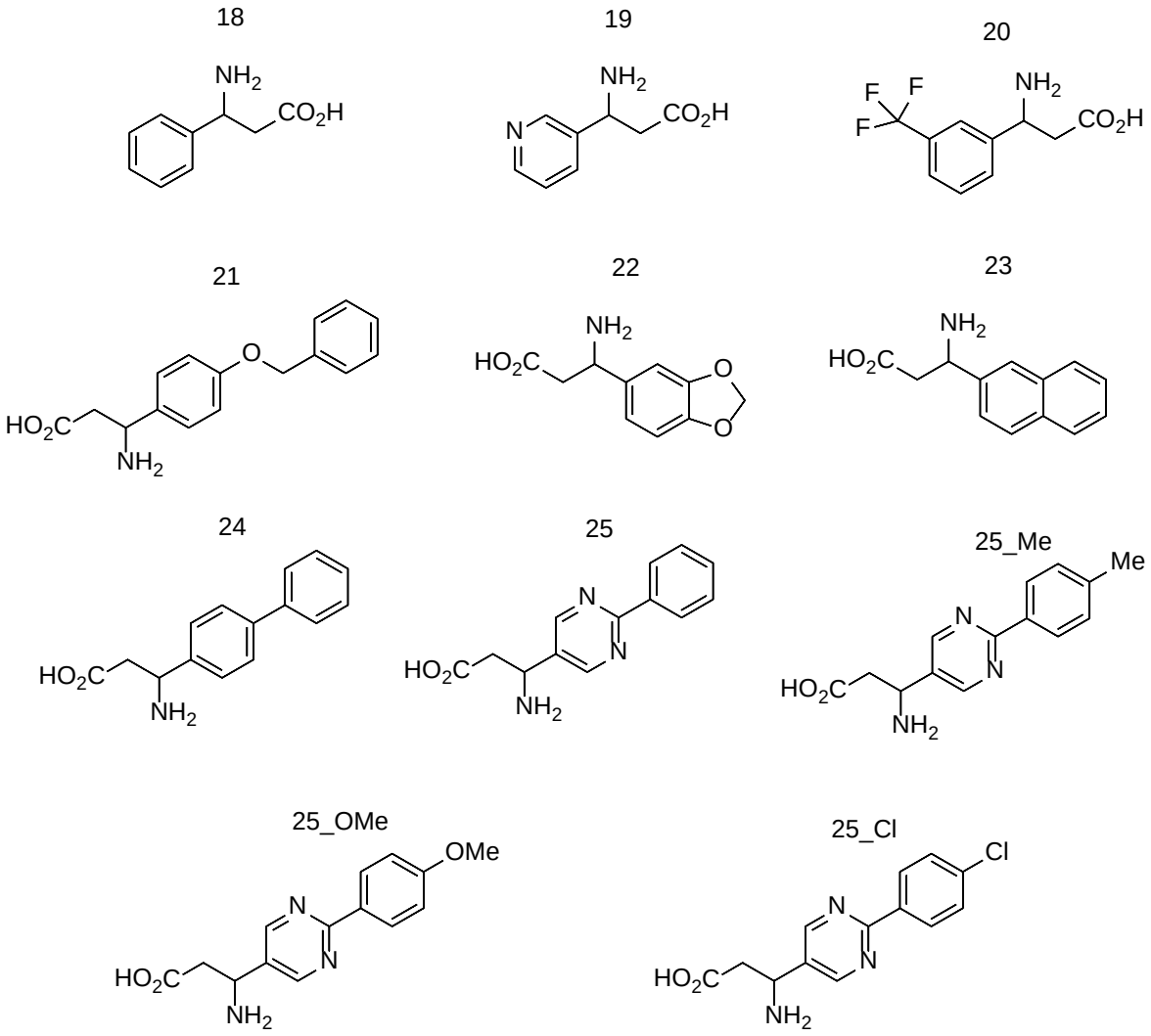
MD simulation. GLYCAM06, gaff, and ff14SB were used as force fields.[8] All MD simulations were carried out by using the sander module in AMBER14. Counterions (Na+ and Cl-) were added to neutralize the system. Each system was then solvated by using TIP3P waters in a cubic box with at least 10 Å of distance around the complex. The shake algorithm was applied to all hydrogen containing bonds, and 1 fs integration step was used. Periodic boundary conditions were applied, as well as the smooth particle mesh Ewald method to represent the electrostatic interactions, with a grid space of 1 Å. Each system was gently annealed from 100 to 300 K over a period of 25 ps. The system were then maintained at temperature of 300 K during 50 ps with a solute restraint and progressive energy minimizations, gradually releasing the restraints of the solute

followed by a 20 ps heating phase from 100 to 300 K, where restraints were removed. Production simulation for each system lasted 100 ns. Coordinate trajectories were recorded each 2 ps throughout production runs, yielding an ensemble of 5000 structures for each complex, which were finally analyzed.

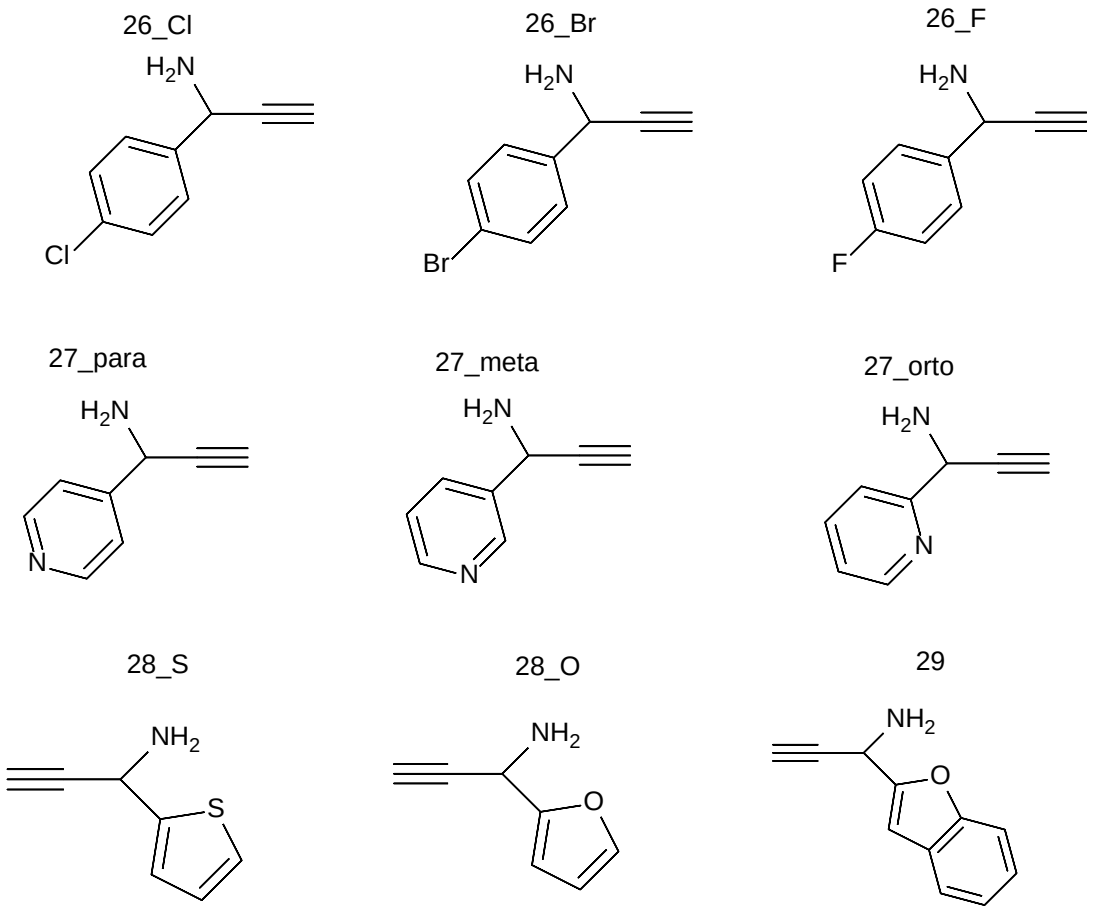
4.5 Annex



Annex Figure 1: α-aminoacids based fragments.



Annex Figure 2: β-aminoacids based fragments.



Annex Figure 3: Triazole based fragments.

Ligand	PBE	Ligand	PBE	Ligand	PBE	Ligand	PBE
1R	-0.5	3R_OMe	+0.1	6S	-0.25	12R	-0.5
1R_F	+0.6	3S	-0.4	6S_O	-1.75	12S	--
1R_0Me	+0.4	3S_CF3	-0.3	7R	-1.5	13R	-0.7
1S	-0.3	3S_F	-0.34	7S	-0.5	13S	--
1S_F	-0.25	3S_OH	--	8R	-0.6	14R	-1.2
1S_OMe	+0.25	3S_Ome	+0.5	8R_OMe	-0.7	14S	-0.3
2R	-1.1	4R	-0.6	8S	+0.9	15R	-1.25
2S	--	4S	-0.1	8S_Ome	+1.4	15S	+0.6
3R	-0.1	5R	+0.6	9R	-0.8	16R	-1.25
3R_CF3	+0.1	5S	-0.4	9S	+0.7	16S	-0.8
3R_F	-1.25	6R	-0.6	10R	+0.2	17R	-1.0
3R_OH	--	6R_0	-1.0	10S	+0.1	17S	-1.0

Ligand	PBE	Ligand	PBE	Ligand	PBE	Ligand	PBE
1R	-3.4	3R_OMe	-3.4	6S	-3.5	12R	-3.8
1R_F	-2.75	3S	-2.9	6S_O	-3.5	12S	-4.1
1R_0Me	-2.8	3S_CF3	-2.5	7R	-4.25	13R	-3.0
1S	-3.3	3S_F	-2.3	7S	-3.8	13S	-2.3
1S_F	-3.3	3S_OH	-2.0	8R	-0.9	14R	-4.0
1S_Ome	-3.5	3S_Ome	-2.25	8R_OMe	--	14S	-2.8
2R	-2.5	4R	-2.75	8S	-1.3	15R	-3.8
2S	-3.0	4S	-2.8	8S_Ome	-2.4	15S	-3.4
3R	-3.3	5R	-2.25	9R	-3.3	16R	-3.6
3R_CF3	-2.75	5S	-3.2	9S	-2.5	16S	-3.5
3R_F	-3.1	6R	-3.5	10R	-3.6	17R	-3.9
3R_OH	-3.1	6R_0	-4.0	10S	-3.5	17S	-4.3

Continuation

Ligand	PBE	Ligand	PBE	Ligand	PBE	Ligand	PBE
1R	-0.9	3R_OMe	-0.5	6S	-1.75	12R	-0.5
1R_F	-0.25	3S	-1.25	6S_O	-2.5	12S	-1.0
1R_0Me	+0.5	3S_CF3	-1.2	7R	-1.5	13R	-0.25
1S	-0.9	3S_F	-1.9	7S	-2.1	13S	-0.2
1S_F	-0.3	3S_OH	-0.6	8R	-0.4	14R	-1.25
1S_OMe	+0.3	3S_Ome	-0.9	8R_OMe	--	14S	-0.8
2R	-1.4	4R	-0.6	8S	-0.4	15R	-1.4
2S	-0.25	4S	+0.4	8S_Ome	+0.3	15S	-1.5
3R	-0.8	5R	-0.25	9R	-0.7	16R	-1.1
3R_CF3	-0.4	5S	-0.5	9S	-1.1	16S	-0.8
3R_F	-0.6	6R	-0.75	10R	+0.1	17R	-1.1
3R_OH	+1.0	6R_0	-0.75	10S	-0.75	17S	-0.8

Annex Table 1: Predicted binding energies obtained with Autodock4 of the α-aminoacids based compounds on 2XR5 crystal structure (Up), 2IT5 crystal structure (Middle) and 2XR5_dyn average structure (Down). PBE =predicted binding energy in kcal/Mol

Ligand	PBE	Ligand	PBE	Ligand	PBE
18R	0,0	22R	--	25R	--
18S	-0,8	22S	--	25R_Cl	+0,6
19R	+0,2	23R	-1,4	25R_Me	+0,2
19S	-1,2	23S	-0,1	25R_OMe	--
20R	+0,6	24R	-0,2	25S	--
20S	+0,5	24S	--	25S_Cl	--
21R	-0,5			25S_Me	--
21S	--			25S_OMe	--

Ligand	PBE	Ligand	PBE	Ligand	PBE
18R	-3,3	22R	-3,7	25R	-3,0
18S	-3,5	22S	-3,0	25R_Cl	-2,7
19R	-2,9	23R	-4,0	25R_Me	-2,5
19S	-2,5	23S	-4,0	25R_OMe	-3,6
20R	-2,6	24R	-3,4	25S	-3,0
20S	-3,0	24S	-3,7	25S_Cl	-3,3
21R	-3,5			25S_Me	-3,5
21S	-3,5			25S_OMe	-3,1

Continuation

Ligand	PBE	Ligand	PBE	Ligand	PBE
18R	+0,1	22R	-1,5	25R	+0,0
18S	-1,3	22S	-1,3	25R_Cl	--
19R	+0,3	23R	-1,0	25R_Me	-0,5
19S	-1,3	23S	-1,5	25R_OMe	+0,7
20R	-0,5	24R	-0,8	25S	--
20S	-0,8	24S	+0,0	25S_Cl	--
21R	--			25S_Me	--
21S	+0,4			25S_OMe	--

Annex Table 2: Predicted binding energies obtained with Autodock4 of the β-aminoacids based compounds on 2XR5 crystal structure (Up), 2IT5 crystal structure (Middle) and 2XR5_dyn average structure (Down). PBE =predicted binding energy in kcal/Mol

Ligand	PBE	Ligand	PBE	Ligand	PBE
26R_Br	-1.25	27R_Nmeta	-0.1	28R_O	--
26R_Cl	-0.4	27R_Norto	-1.0	28R_S	-0.8
26R_F	-0.8	27R_Npara	-1.2	28S_O	-0.8
26S_Br	-1.6	27S_Nmeta	-1.0	28S_S	-1.9
26S_Cl	-1.5	27S_Norto	-1.4	29R	-0.7
26S_F	-0.1	27S_Npara	-1.5	29S	-2.0

Ligand	PBE	Ligand	PBE	Ligand	PBE
26R_Br	-3.5	27R_Nmeta	-3.7	28R_O	-3.9
26R_Cl	-3.7	27R_Norto	-3.2	28R_S	-4.3
26R_F	-4.3	27R_Npara	-3.8	28S_O	-3.3
26S_Br	-3.4	27S_Nmeta	-3.2	28S_S	-3.5
26S_Cl	-3.4	27S_Norto	-2.8	29R	-4.3
26S_F	-3.3	27S_Npara	-3.1	29S	-3.7

Continuation

Ligand	PBE	Ligand	PBE	Ligand	PBE
26R_Br	-1.0	27R_Nmeta	-1.0	28R_O	-0.9
26R_Cl	-1.0	27R_Norto	-0.5	28R_S	-1.0
26R_F	-1.2	27R_Npara	-0.9	28S_O	-1.0
26S_Br	-2.1	27S_Nmeta	+0.2	28S_S	-2.1
26S_Cl	-2.4	27S_Norto	-1.6	29R	-1.7
26S_F	-2.5	27S_Npara	-2.0	29S	-2.0

Annex Table 3: Predicted binding energies obtained with Autodock4 of the triazole based compounds on 2XR5 crystal structure (Up), 2IT5 crystal structure (Middle) and 2XR5_dyn average structure (Down). PBE =predicted binding energy in kcal/Mol.

Bibliography

1. (a) Jin, W.; Li, C.; Du, T.; Hu, K.; Huang, X.; Hu, Q., DC-SIGN plays a stronger role than DCIR in mediating HIV-1 capture and transfer. *Virology* **2014,** *458-459*, 83-92; (b) Tacken, P. J.; de Vries, I. J.; Gijzen, K.; Joosten, B.; Wu, D.; Rother, R. P.; Faas, S. J.; Punt, C. J.; Torensma, R.; Adema, G. J.; Figdor, C. G., Effective induction of naive and recall T-cell responses by targeting antigen to human dendritic cells via a humanized anti-DC-SIGN antibody. *Blood* **2005,** *106* (4), 1278-85.

2. (a) Ordanini, S.; Varga, N.; Porkolab, V.; Thepaut, M.; Belvisi, L.; Bertaglia, A.; Palmioli, A.; Berzi, A.; Trabattoni, D.; Clerici, M.; Fieschi, F.; Bernardi, A., Designing nanomolar antagonists of DC-SIGN-mediated HIV infection: ligand presentation using molecular rods. *Chemical communications (Cambridge, England)* **2015,** *51* (18), 3816-9; (b) Thepaut, M.; Guzzi, C.; Sutkeviciute, I.; Sattin, S.; Ribeiro-Viana, R.; Varga, N.; Chabrol, E.; Rojo, J.; Bernardi, A.; Angulo, J.; Nieto, P. M.; Fieschi, F., Structure of a glycomimetic ligand in the carbohydrate recognition domain of C-type lectin DC-SIGN. Structural requirements for selectivity and ligand design. *J. Am. Chem. Soc.* **2013,** *135* (7), 2518-29.

3. Maybridge database. www.maybridge.com (accessed March 2019).

4. Feinberg, H.; Mitchell, D. A.; Drickamer, K.; Weis, W. I., Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **2001,** *294* (5549), 2163-6.

5. Feinberg, H.; Castelli, R.; Drickamer, K.; Seeberger, P. H.; Weis, W. I., Multiple modes of binding enhance the affinity of DC-SIGN for high mannose N-linked glycans found on viral glycoproteins. *J. Biol. Chem.* **2007,** *282* (6), 4202-9.

6. (a) 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-49; (b) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004,** *47* (7), 1750-9.

7. 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.

8. (a) 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; (b) 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; (c) 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.

CHAPTER 5:

Studies on the molecular recognition process of TLR4 modulators

5.1 Agonist synthetic peptides RS01 and RS09

In this part of the work we have studied the binding of two reported agonist peptides of TLR4. Shanmugam *et al.* reported a series of agonist peptides designed by means of phage display combinatorial peptide technology.[1] Phage display is used as a technique to study protein-protein interactions, mainly. It was first described by Smith in 1985.[2] In this technique, a gene encoding a protein of interest is inserted into a phage coat protein gene. That causes the phage to display the protein on its surface, while containing the gene for the protein, resulting in a connection between genotype and phenotype. Then, those phages can be screened against other proteins in order to detect interaction between them.

The authors used a commercial library containing phages that express a variety of 7 amino acid length peptides. The library was screened against an anti-LPS antibody, in order to discriminate which of the peptides could mimic the LPS structure. Among the 12 initially reported peptides that interact with the antibody, RS01, RS03 and RS04 showed a strong LPS activity at 10μg/mL, comparable to the activity of LPS. On the contrary, RS02, RS05-07, RS08 and RS10 did not trigger TLR4 activation. Finally, RS09, RS11 and RS12 showed a partial agonist activity. We chose two of these peptides (RS01 and RS09) to investigate their binding pose to TLR4/MD-2 and try to shed light into their agonist mechanism.

5.1.1 Modelling of RS01 and RS09 Peptides

Agonist peptides RS01 (Gln-Glu-Ile-Asn-Ser-Ser-Tyr) and RS09 (Ala-Pro-Pro-His-Ala-Leu-Ser) were built using PyMol software. Both peptides were minimized using Maestro, and the resulting structures were used to perform docking calculation in TLR4/MD-2. Since these peptides mimicked LPS, according to the

phage display experiments, the grid of the docking was placed occupying the full hydrophobic pocket of MD-2, as well as the rim of the pocket. According to GlideScore, best predicted binding energies for RS01 and RS09 were very similar, -11.1 kcal/mol and -10.9 kcal/mol, respectively. However, differences in term of the predicted binding site could be observed, as RS01 was predicted to bind at the rim of MD-2 pocket while the prediction for RS09 was to be inserted into the pocket.

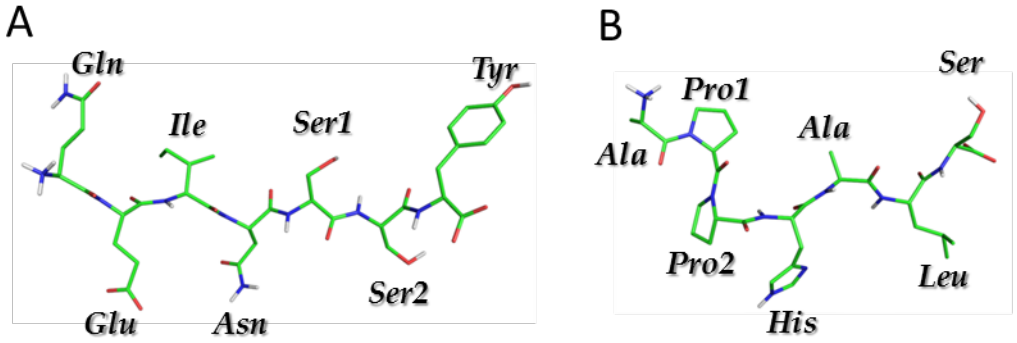


Figure 1: Structures of A) RS01 peptide (Gln-Glu-Ile-Asn-Ser-Ser-Tyr). B) RS09 peptide (Ala-Pro-Pro-His-Ala-Leu-Ser). Non polar hydrogens are not shown.

This different predicted binding mode could be due to the predominant abundance of aliphatic residues in RS09 peptide (two Alanines, two prolines and a Leucine) in comparison to RS01 (only one Isoleucine). Thus, RS09 was docked with the N-terminal region inside the MD-2 pocket, with both prolines interacting with Ile63, Leu61, Ile117 and Tyr65. RS09-His is as well in the MD-2 pocket and interacts with Phe119. Val, Leu and Ser of RS09 peptide are located at the rim of the pocket. Leucine interacts with Ile117 and hydroxyl group of RS09-Ser interacts with the CO group of the Val93 backbone. On the contrary, RS01 was docked completely at the rim of MD-2, with none of its residues inserted in the MD-2 pocket. The NH group of RS01-Gln establishes an H-bond with the CO group of Lys122 backbone, while the side chain of RS01-Asn interacts with the backbone of Ser120 through a H-bond. Close to the C-terminal region of RS01, the side chain of the RS01-Ser1

interacts with side chain of Glu92 through a hydrogen bond and another hydrogen bond with the RS01-Ser2 is formed with the Val93 backbone CO. Note that in the case of RS01, more polar contacts between the peptide and MD-2 were predicted than in the case of RS09. Another important feature of the RS01 predicted binding pose is the fact that RS01 interacts with residues not only from MD-2 but also from TLR4, like Arg239, which stablishes a hydrogen bond with the RS01-Ser2, and Asn314, which interacts with the backbone of the RS01-Tyr.

Since only from the docking we could not discriminate why RS01 is a better agonist than RS09, we decided to perform a MD simulation of both complexes. After 40ns of MD simulation, both complexes remained stable, but only RS01 peptide was able to maintain the agonist conformation of the Ile124-Phe126 loop (with the Phe126 side chain towards the inner region of MD-2, Figure 2). RS01 peptide was kept at the rim of MD-2 during the whole simulation, with a minor displacement through the rim of MD-2. MD-2 hydrophobic pocket seems to collapse slightly, leading to a displacement of the C-terminal region of the peptide, allowing the RS01-Phe to interact with TLR4 residues like Asn314 and Lys337. Moreover, Arg90 and Tyr102 of MD-2 came also in contact with the peptide, through a ionic interaction with the side chain of RS01-Glu and a H-bond with the side chain of RS01-Ser1, in the case of Arg90, and a H-bond with the RS01-Ser2 backbone. Lys120 also stablishes a H-bond with the side chain of RS01-Asn and with the CO backbone of RS01-Ile. Finally, RS01-Ser2 side chain stablishes a H-bond with the side chain of Glu92 of MD-2 and the backbone CO of RS01-Ser2 interacts with the side chain of Ser 118 of MD-2 (Figure 3).

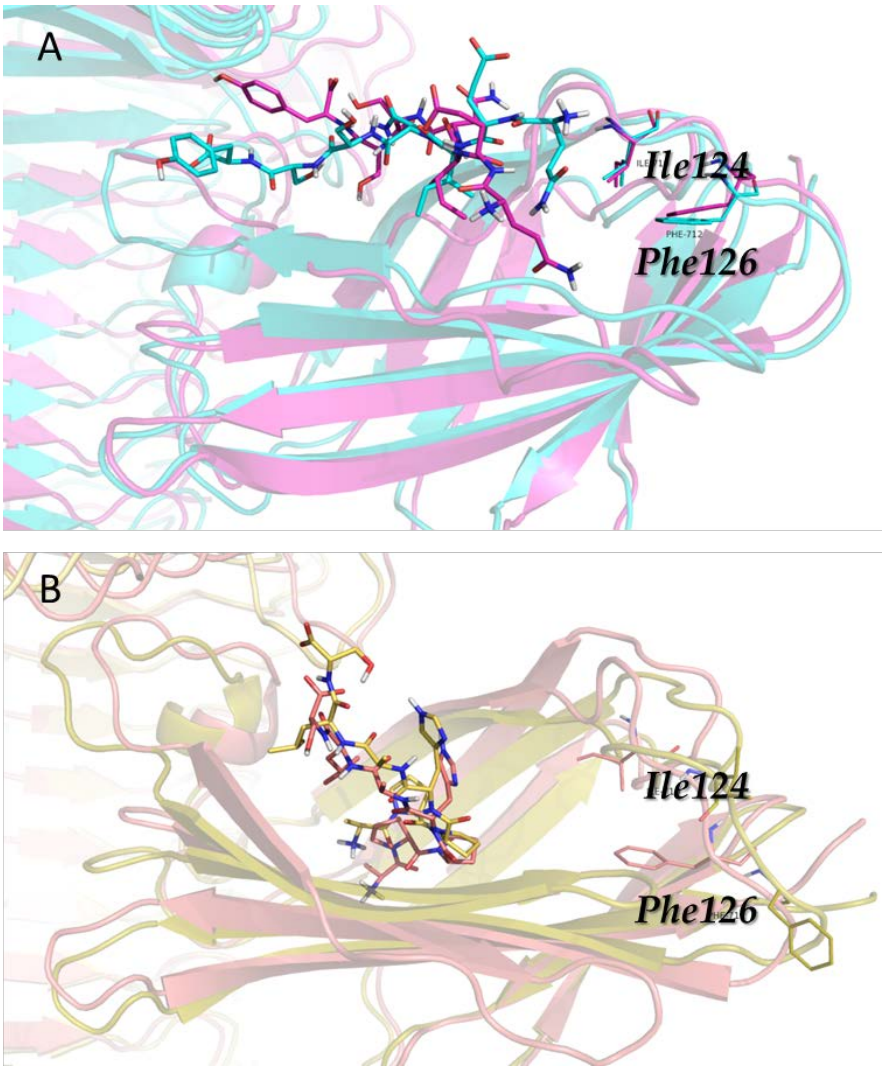


Figure 2: Docking and MD simulation RS01 and RS09. A) Docked pose (cyan) and 100ns MD simulation structure (pink) of RS01 peptide. B) Docked pose (salmon) and 100ns MD simulation structure (yellow) of RS09 peptide. Ile124 and Phe126 are highlighted in sticks.

In the case of RS09, the complex with TLR4/MD-2 remained stable as well. The overall structure of MD-2 did not suffer major changes with the exception of the Ile124-Phe126 loop. These residues changed their position to the so-called antagonist conformation.[3] Phe126, initially buried inside the MD-2 pocket, exposes its side chain to the solvent at the very beginning of the dynamics. Ile124, partially exposed to the solvent at the beginning, enters to the MD-2 pocket and occupies the initial place of Phe126. This is one of most important feature that determines

the agonist of antagonist behavior of a TLR4 ligand. As said before, the peptide does not suffer major changes during the dynamics, as it kept buried deep inside the MD-2 pocket (Figure 2).

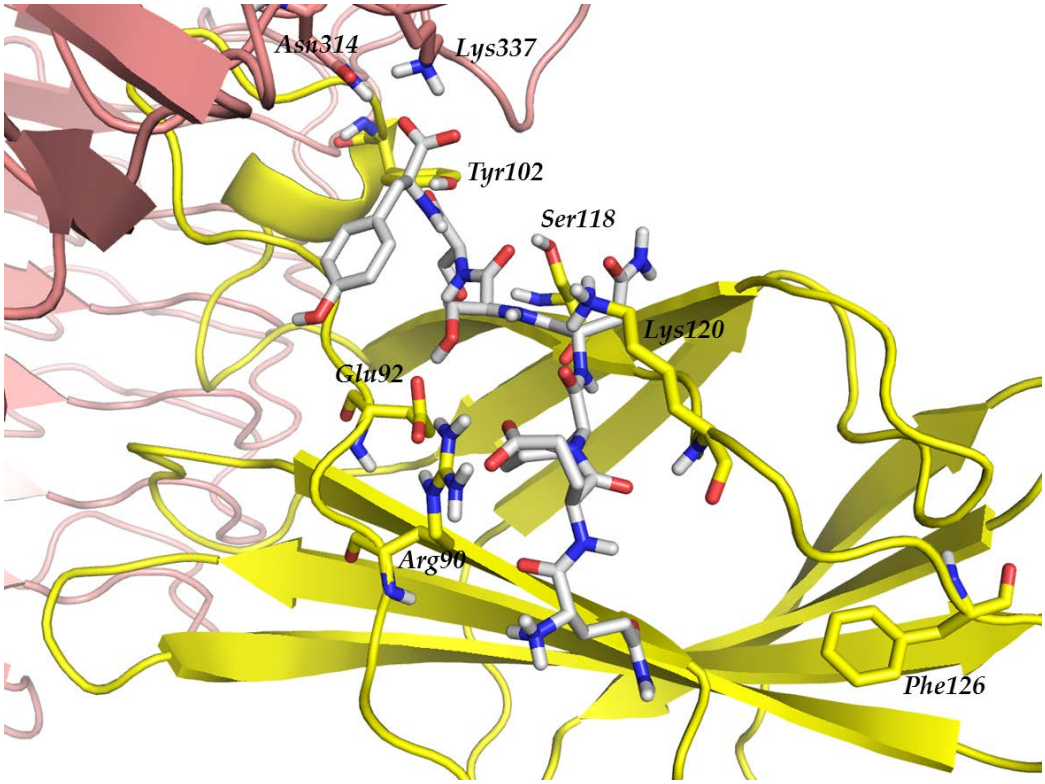


Figure 3: MD simulation of RS01 peptide (grey): average geometry from the last nanosecond. TLR4 is represented in salmon. MD-2 is represented in yellow.

To complement these findings, we also performed a MM-ISMSA protocol to determine the strength of the binding of both peptides to TLR4/MD-2 complex. The average binding energy predicted to RS01 during the dynamics was -67.7 kcal/mol, while the energy to RS09 was -53.44 kcal/mol, indicating a stronger interaction of RS01 peptide than RS09. Overall, these findings are in agreement with the experimental data suggesting that RS01 a better agonist that RS09.

Therefore, we have here proposed a binding mode for these reported peptide agonists in agreement with their biological activities.

5.2 Antagonist Peptide MDMP

In another work published by Duan *et al.* in 2010, a synthetic MD-2 mimetic peptide with antagonist activity on TLR4 is reported.[4] They designed that peptide based on the LPS and TLR4 binding motifs present in the structure of MD-2, called MDMP. LPS binding motif is formed by amino acids between Phe119 and Lys132[3, 5] and TLR4 binding motif is composed by residues from Cys95 to Cys105.[6] It is important to highlight the importance of the disulfide bond between these two residues, due to its formation promotes the adoption of a tertiary structure that is crucial for the binding to TLR4.[7] Thus, the final sequence of the designed peptide was CHGHDDDYSFCFSFEGILFPKGHYR, with a disulfide bond between both cysteines in the structure. In the paper, they conclude that this peptide binds efficiently to TLR4 and block the binding of MD-2 to TLR4, resulting in an effective and promising TLR4 antagonist.[4]

In this part of the work, we try to give some answers related to the binding mode of this peptide to TLR4, to understand the mechanism of action and to propose modifications that could lead to an increase in the antagonist activity.

5.2.1 Modelling of MDMP

Since we wanted to study the ability of MDMP to bind to TLR4, we started by working with the TLR4 binding domain, *i.e.* the region comprised from Cys95 to Cys105 of the MD-2 sequence. In order to get the 3D structure of this region, we extracted it from the MD-2 of 3FXI crystal structure. From this starting geometry, we performed a conformational analysis to get the lowest energy conformations of the peptide.

The resulting structure is shown in Figure 4.

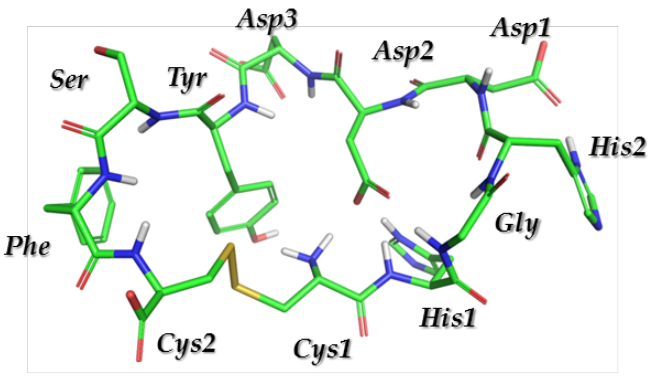


Figure 4: 3D structure of the TLR4 binding motif in MDMP peptide. Non polar hydrogens are not shown.

As it can be seen, the peptide adopts an almost circular (closed) structure, due to, mainly, the disulfide bridge present in it. We also performed a 100ns MD simulation of the peptide, to explore more possible conformations that could affect to the docking results. As it can be seen in Figure 5, the peptide remains stable after 75 ns of MD simulation, with no major changes in its secondary structure after that time. We took the average structure of the last nanosecond as starting geometry to perform a docking in TLR4. The docking was performed with AutoDock4 and the center of the box was placed on the residues that interact with MD-2 in the 3FXI crystal structure, *i.e.* Arg234, Asn265 and Arg264. The backbone torsional angles were kept fixed, in order to preserve the secondary structure during the docking. It is important to mention that this backbone conformation was obtained from previous conformational analysis.

The cluster with higher number of poses after the docking contains the poses that are more similar to the TLR4 binding motif of MD-2. The Tyr and Asp1 MDMP residues interact with Arg264, and MDMP-Ser side chain stablishes a H-bond with Glu266 and Asn265 (Figure 6, A). With these results in hand, we performed a 100ns MD simulation of the complex, in order to see if the pose remains stable. Figure 6 shows the results of the simulation. RMSD of both MDMP and TLR4 did

not show relatively important changes during time. In fact, the peptide adopts a conformation which is more similar to the MD-2 pose on 3FXI at the end of the simulation than the docked pose.

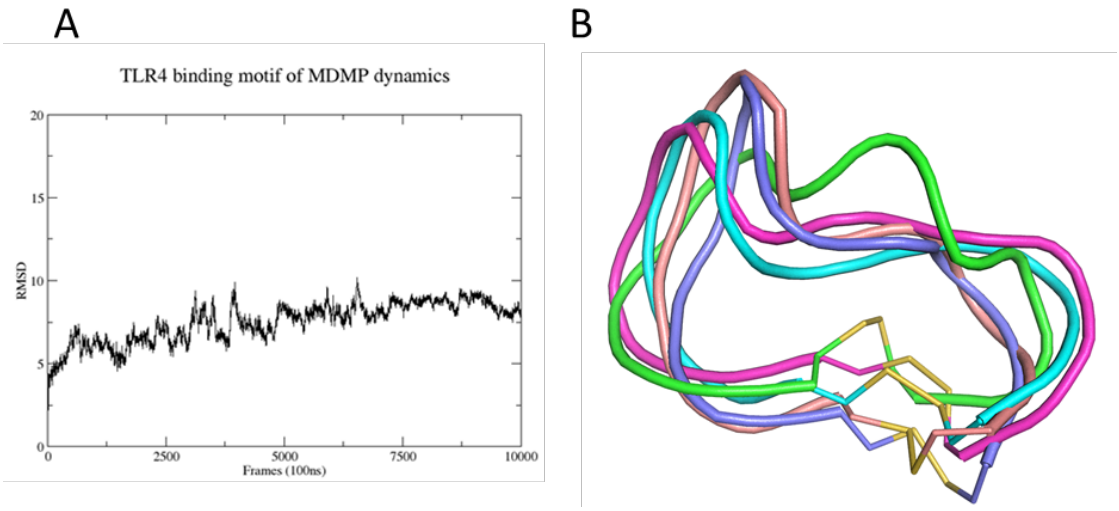


Figure 5: A) RMSD graph of 100ns of MD simulation of the TLR4 binding motif of MDMP. B) Ribbon representation of the average secondary structures during the simulation. Green = Initial structure; Salmon = 25ns; Blue = 50ns; Pink = 75ns; Yellow = 100ns. The cysteine bridge is represented in sticks.

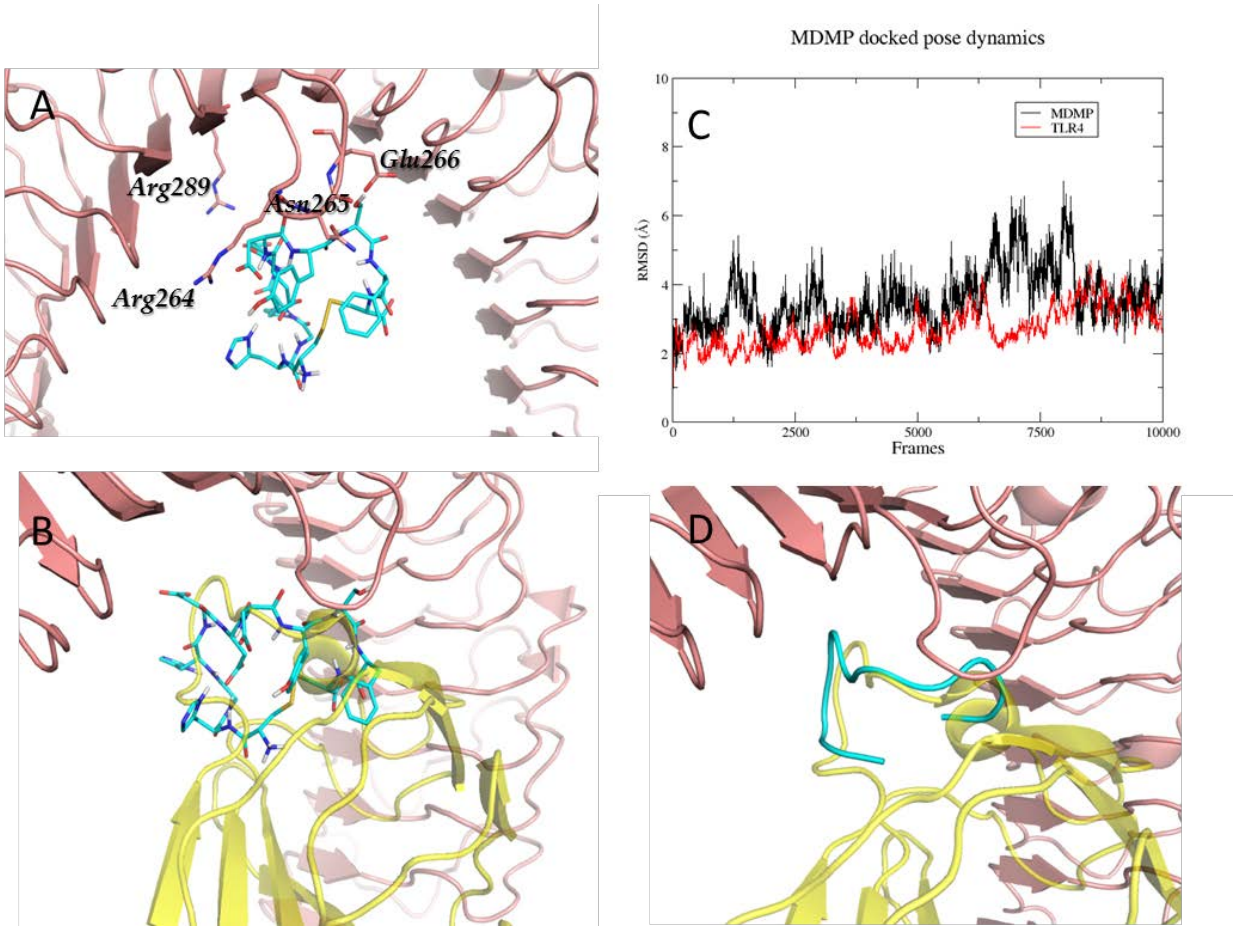


Figure 6: Docking and MD simulation of TLR4/MDMP complex. A) Docked pose of MDMP (blue sticks) in TLR4 (salmon). Interacting residues in TLR4 are highlighted in sticks. A) Superimposition of docked TLR4/MDMP complex with the TLR4/MD-2 from PDB 3FXI (yellow). C) RMSD graph of MDMP (black line) and TLR4 (red line) during the MD simulation of the TLR4/MDMP complex. D) Superimposition of the average structure TLR4/MDMP of the last nanosecond of the simulation (blue) with the TLR4/MD-2 from PDB 3FXI (yellow).

5.3 *Bacteroides vulgatus* LPS

This section of the chapter was done in collaboration with Prof. Molinaro and Prof. Silipo at University of Naples Federico II.

In a previous study, our collaborators characterized a member of the phylum Bacteroidetes, *Bacteroides vulgatus* mpk, a symbiotic commensal of the mouse intestine, providing high genome plasticity and strong immune-modulating

properties leading to prevention of colitis-induction in several mouse models for experimental colitis.[8] In fact, not only the presence, abundance or proportion of certain live microbial species contributes to such microbiota-mediated effects. Indeed, also the structure and, consequently, the toxicity of Gram-negative bacteria lipopolysaccharides determine the outcome of inflammation in a mouse model for experimental colitis.[9]

The group of Prof. Molinaro has characterized the structure of *Bacteroides vulgatus* LPS (LPS_{BV}). It revealed that LPS_{BV} is actually a mixture of a tetracylated and pentacylated LPS and has a unique core OS composition. Briefly, the core region was built up of a hexa-saccharide bearing a phosphorylated Kdo unit, in turn, substituted at position O-5 by a rhamnose (Rhap) residue. This latter was found to be a branched monosaccharide bearing a β-galactofuranose unit (Galf) at its position O-3 and, at position O-4, a β-galactopyranose (Galp) unit. This latter was, in turn, di-substituted by α-fucose (Fucp) and, finally, by a β-glucose (Glcp) which was established to be the residue substituted by the first rhamnose of the O-chain. The O-chain structure was built up of repeating units of β-mannose (Manp) and α-rhamnose (Figure 7). The lipid A part has only one phosphate on the galactose unit and the FA chains have between 15 and 17 carbons, in contrast with *E. coli* LPS, which have only chains with 14 carbons.

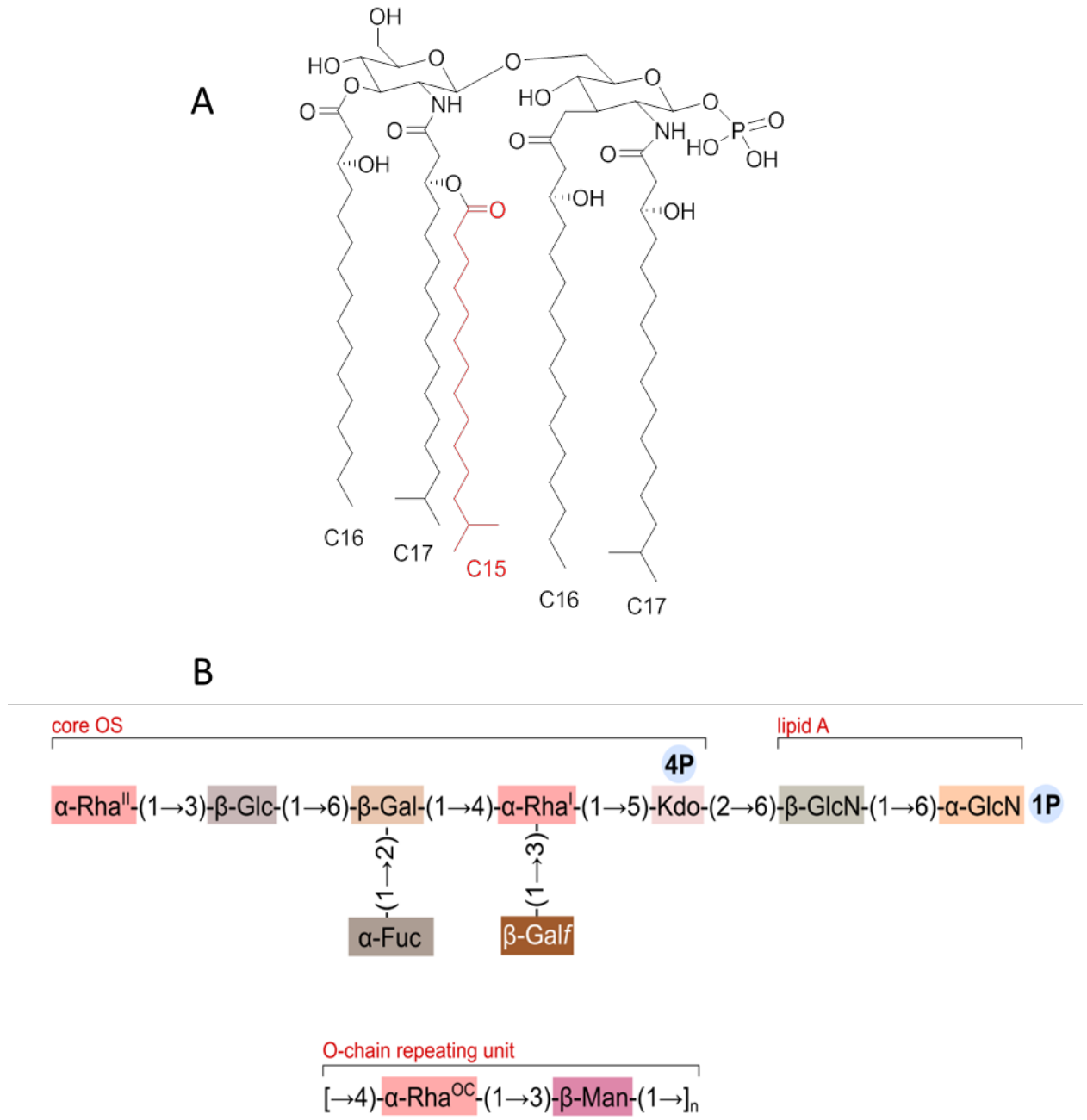


Figure 7: A) Elucidated 2D structure of the pentacylated LPS$_{BV}$. FA chain that is not present in tetracylated is marked in red. B) Schematic representation of the Core OS of LPS$_{BV}$.

Our workers proved that the administration of live *Bacteroides vulgatus* in colonic inflammation mouse models reduced drastically the inflammation and the administration of the mixture of LPS$_{BV}$ alone mimics the effect of the

administration of live bacteria. Moreover, LPS_{BV} protective role seems to be associated with a weak activation of dendritic cells, acting as an immunoprotector. Given that the active principle of live cells is the LPS_{BV}, it is presumable that TLR4 could have a role in the activity of LPS_{BV}.

Here we provide a model of the binding mode of both tetra and pentacylated LPS_{BV} on TLR4/MD-2, trying to provide a 3D perspective to the activity.

5.3.1 Modelling of LPS_{BV}

We performed computational studies using both docking and molecular dynamics (MD) simulation approaches. Given that the structural analysis of LPS_{Bv} revealed a mixture of species that are actually present in the sample employed to experimentally evaluate LPS_{Bv} lipid A activity, we decided to use the two most representative and abundant structures of the mixture for our computational study, namely the *mono*-phosphorylated tetra- and penta-acylated LPS_{Bv} compounds (referred to here on as LPS_{Bv}-tetra and LPS_{Bv}-penta).

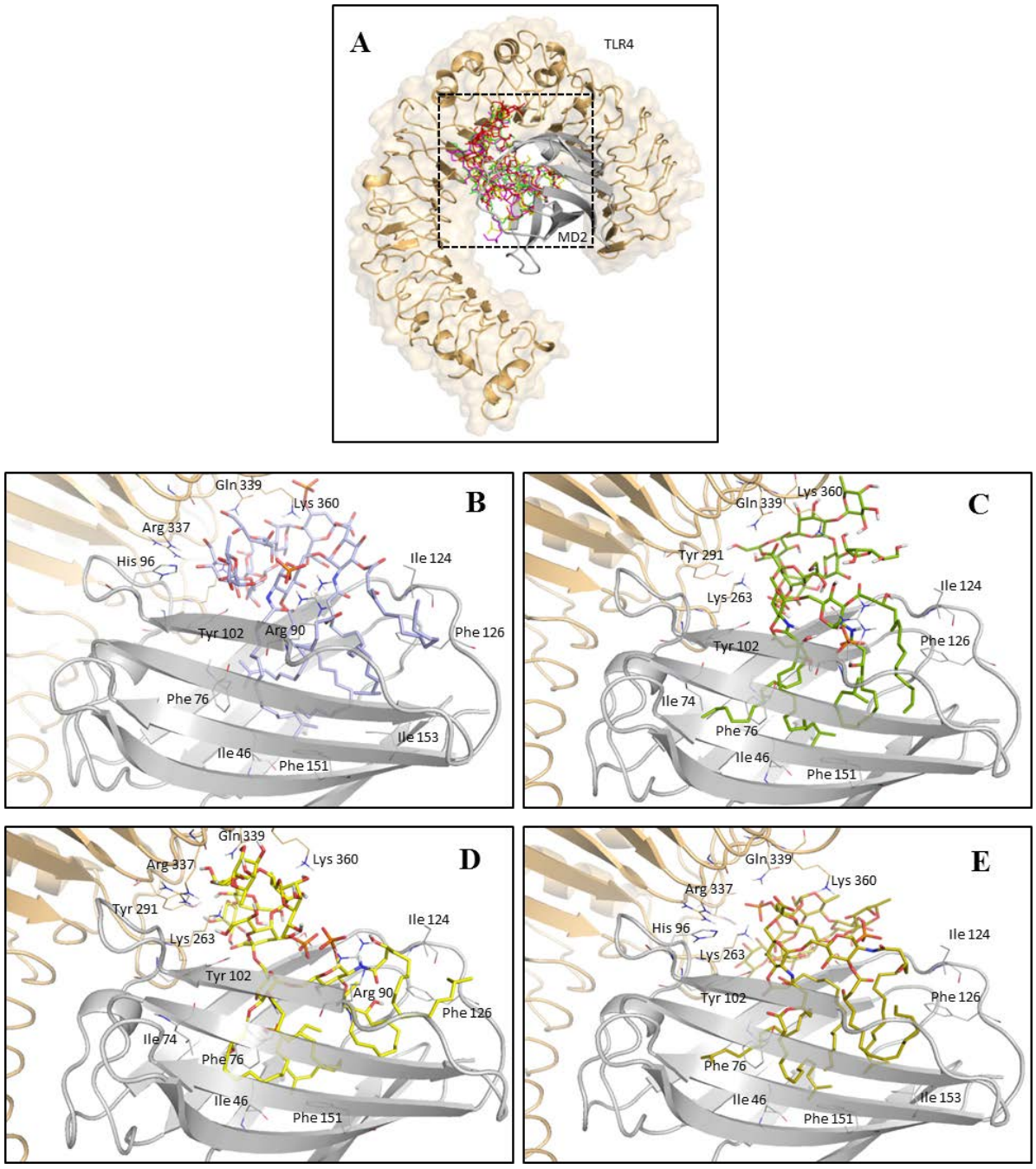


Figure 8: Docking calculations of LPS$_{Bv}$–tetra and LPS$_{Bv}$–penta into TLR4/MD-2. Residues in contact with LPS$_{Bv}$ are displayed in wireframes (A, B, C, and D). A) Superimposition of selected docked poses of LPS$_{Bv}$-tetra and LPS$_{Bv}$-penta in mTLR4/MD-2. B, C, D, E) Detail of representative binding poses of both LPS$_{Bv}$-penta and LPS$_{Bv}$-tetra. (B) and (C) correspond to poses 02 and 09 of LPS$_{Bv}$-penta, respectively. Pose 02 (B) corresponds to a type-A orientation and presents the five FA chains inserted into the MD-2 pocket. Pose 09 (C) corresponds to a type-B orientation and one of its FA chains is located at the MD-2 channel. (D) and (E) correspond to poses 15 and 10 of LPS$_{Bv}$-tetra, respectively, both corresponding to a type-A orientation (no docked type-B orientations were predicted in this case). Pose 10 (D) presents one FA chain at the MD-2 channel.

We performed docking calculations to predict putative binding modes for both ligands inside the murine TLR4/MD-2 system (*m*TLR4/MD-2), to be subsequently submitted to MD simulations to study the stability and dynamics of the predicted bound complexes. Reported X-ray crystallographic structures for similar compounds show that the fatty acid (FA) chains are inserted into the MD-2 pocket, while the oligosaccharide portion establishes polar interactions with complementary residues from the TLR4 and also from the partner TLR4* when a heterodimer is formed (PDB ID 3FXI 3VQ2). In the case of *E. coli* hexa-acylated LPS, five of the FA chains are inserted into the MD-2 pocket and the sixth one is placed in a groove or channel of MD-2 delimited by residues Arg90 and Phe126, thus allowing the conformation of the Phe126 side chain to be kept towards the inner part of MD-2. This disposition has been established as the agonist bound conformation for LPS_{Ec} (PDB ID 3FXI), while the antagonist conformation presents the Phe126 exposed to the solvent and the Ile124 occupies the Phe126 initial position (PDB ID 2E59). Our previous molecular modeling studies for different LPSs or synthetic glycolipids also predicted binding modes similar to the X-ray crystallographic binding poses.

As for the docking of LPS_{Bv}-tetra and LPS_{Bv}-penta, different docked poses with favorable predicted binding energy were obtained: some of them with the FA chains inserted into the MD-2 pocket (most of the poses), and other poses with one FA chain placed in the channel delimited by residues Arg90 and Phe126, and the other FA chains inside the MD-2 pocket, similarly to LPS_{Ec} in human (*h*-) and murine (*m*-) TLR4/MD-2/TLR4*/MD-2* heterodimers (PDB ID 3FXI and 3VQ2). The oligosaccharide moiety was predicted to establish polar interactions with the TLR4 with high similarity to the binding poses found for LPS_{Ec} in the X-ray crystal structures of h- and *m*TLR4/MD-2/LPS_{Ec}/TLR4*/MD-2*/LPS_{Ec}* (Figure 8).

The phosphate group at the Glc_{RID} was found to be oriented towards the side chain of Tyr102 of MD-2 for LPS_{Bv}-tetra and most of the docked poses of LPS_{Bv}-penta. This orientation is similar to that found for *E. coli* LPS in the PDB IDs 3FXI and 3VQ2, and for tetra-acylated lipid IV_A, a biosynthetic precursor of *E. coli* lipid A, in *m*TLR4/MD-2/TLR4*/MD-2* (PDB ID 3VQ1), accounting for their agonist activity, and it will be referred here on as orientation "type-A". For LPS_{Bv}-penta, additionally, a 180°-rotated orientation ("type-B") was also found in few docking solutions, with the phosphate group at Glc_{RID} oriented towards the guanidinium group of Arg90 of MD-2. This type-B orientation is also found for lipid IV_A when bound to *h*TLR4/MD-2 (PDB ID 2E59) in the antagonist bound pose. Lipid IV_A is, indeed, widely known for acting as an antagonist on *h*TLR4/MD-2 as X-ray crystallographic studies of the compound in complex with *h*TLR4/MD-2 showed that all its four acyl chains are placed inside the MD-2 binding cavity in a manner that does not permit dimerization and subsequent activation of the downstream signaling. The phosphate group linked to Kdo, in both compounds, establishes polar interactions with the side chains of residues Lys263, Arg337 or Lys360, depending on the pose, which are key residues in the recognition of LPSs by the TLR4/MD-2 complex, according to the literature data.[24] In all the poses of both LPS_{Bv}-tetra and LPS_{Bv}-penta, also other interactions from the different sugar moieties (depending on the binding pose) with the side chains of particular residues are identified: Gln339 (TLR4), Ser120 (MD-2) and Tyr102 (MD-2). These residues are also present in the interactions network involving the oligosaccharide of LPS_{Ec} in the reported X-ray crystal structures (PDB IDs 3FXI and 3VQ2).

From the docking results, we selected 13 different TLR4/MD-2/LPS_{Bv} complexes (from 8 and 5 selected docked poses of the LPS_{Bv}-penta and LPS_{Bv}-tetra compounds, respectively) to be submitted to 20ns MD simulations. Different

aspects were monitored along the simulation (Annex figures 1-3): i) stability of the LPS_{Bv} ligands and the TLR4/MD-2 system, ii) stability and values of the ligand binding energies, iii) stability of the FA chains kept inside the MD-2 pocket (or one of them in MD-2 channel), iv) stability of the Phe126 and Ile124 side chains towards the inside and outside of MD-2, respectively (agonist conformation).

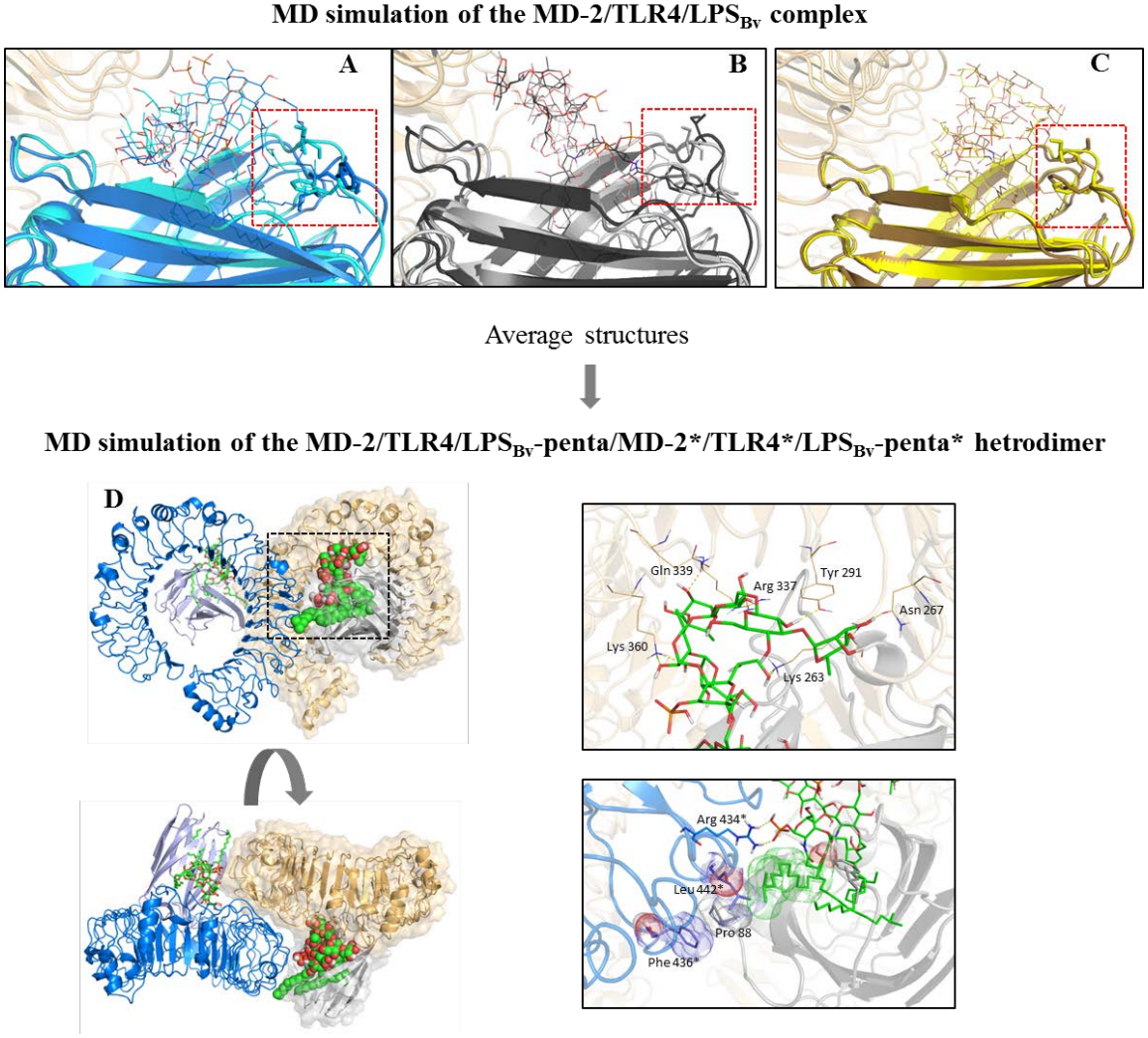


Figure 9: MD simulations of different complexes of TLR4/MD-2with LPS_{Bv}–penta and LPS_{Bv}–tetra. (a, b and c) Superimposed geometries for TLR4/MD-2/LPS_{Bv} complexes along the MD simulation (initial geometry in cyan, brown, and grey; after 20ns of MD simulation in deep blue, yellow, and black). (a) TLR4/MD-2/LPS_{Bv}-penta complex from docked pose 09. (b) TLR4/MD-2/LPS_{Bv}-penta complex from docked pose 19. (c) TLR4/MD-2/LPS_{Bv}-tetra complex from docked pose 10. Red squares highlight Ile124 and Phe126 shifting during the MD simulation. (d) 3D Structure of the TLR4/MD-2/LPS_{Bv}-

tetra/TLR4/MD-2*/LPS$_{Bv}$-tetra* heterodimer complex from MD simulations (from pose 10, average after 20 ns simulation time). Front (top) and lateral (bottom) views. (f) Detail of the polar interactions of the core OS with TLR4. (g) Detail of the interactions taking place at the dimerization interface between MD-2 and the partner TLR4*. Representative interactions are: the FA chain at the MD-2 channel stablishes a hydrophobic interaction with Leu442* of the partner TLR4*; the phosphate group of GlcN$_{RID}$ interacts with the guanidinium group of Arg434* side chain; the Pro88 of MD-2 stablishes a π-π interaction with the Phe436 side chain from TLR4*.An additional docked pose (pose 15) was also selected, and two different models of the TLR4/MD-2/LPS$_{Bv}$-tetra/TLR4*/MD-2*/LPS$_{Bv}$-tetra* heterodimer were built.*

In the case of LPS$_{Bv}$-penta, after the 20ns MD simulation time, five of the eight selected poses kept the agonist conformation of Phe126, while the other three suffered a partial rotation of the Phe126 side chain towards the outer part of MD-2, even allowing Ile124 side chain to occupy the initial position of Phe126 side chain in few cases, but without observing complete rotation towards the antagonist conformation. From the RMSD of LPS$_{Bv}$ (Annex Figure 1), it can be observed that the major changes took place in the core OS, due to the accommodation of the polar interactions, while lipid A moiety remained stable in all cases. Among the eight TLR4/MD-2/LPS$_{Bv}$-penta complexes, and taking into account the stability along the MD simulation, we selected two of them with type-A orientation (from docked poses 02 and 19) and two with type-B orientation (from docked poses 04 and 09) as the best complexes to build four models of the full TLR4/MD-2/LPS$_{Bv}$-penta/TLR4*/MD-2*/LPS$_{Bv}$-penta* heterodimer (Figure 9).

In the case of compound LPS$_{Bv}$-tetra, five docked poses were selected as the best ones, being all of them with type-A orientation. Only one of these five selected poses (from docked pose 10) kept the agonist conformation of Phe126 after the MD simulation (Annex Figure 2).

We finally constructed six TLR4/MD-2/LPS$_{Bv}$/TLR4*/MD-2*/LPS$_{Bv}$* full heterodimer models which were submitted to 20ns MD simulations: four from the LPS$_{Bv}$-penta

docked poses (two type-A oriented and two type-B oriented), and two from the LPS_{Bv}-tetra (type-A oriented) (Figure 9). A similar analysis to that performed to the previous MD simulations of the monomers was performed. Except for the model from pose 09 of LPS_{Bv}-penta, which suffered a geometry change due to the approaching of N-terminal and C-terminal domains, coming in close contact, no significant changes were observed during the simulations (with very stable RMSD values), and the Phe126 side chain remained in the partially rotated agonist conformation, in a similar position to that found in the previous models, while Ile 124 side chain is stable towards the inner part of MD-2 (Annex figure 3).

Interestingly, the model from pose 10 of LPS_{Bv}-tetra maintained the agonist conformation of both Phe126 and Ile124 side chains.

5.4 Cardiolipin

This part of the work was performed in collaboration with Prof. Fukase, at the University of Osaka, Japan.

Cardiolipin (CL) is an important component of the inner mitochondrial membrane, constituting the 20% of the total lipid composition of this organelle. The name cardiolipin is derived from the fact that it was first isolated from animal hearts, more concretely, from a beef heart.[10] CL contains two phosphatidylglyceride backbone molecules and therefore contains also four fatty acids (Figure 10).[11] Different fatty acids bound to all four positions give rise to a highly diversified CL pool in most mammalian tissues. CL is involved in many essential functions linked to mitochondrial membranes, including mitochondrial morphology, mitochondrial metabolism, and respiration. Accordingly, defects in the biosynthesis and remodeling of CL have been linked with severe disorders such as Sengers disease and Barth syndrome.[10, 12]

In the recent years, some works have linked CL to TLR4.[13] There is strong evidence in the literature that CL can bind to TLR4, but it is not clear still the exact role of this fatty acid in the activation of TLR4, since the bibliography in this field has a variety of results. Prof. Fukase and his coworkers observed a possible synergistic effect of murine CL with LPS (data not published). When they incubated HekBlue Cells with LPS and low concentrations of CL (50 and 500ng/mL), they observed a higher secretion of IL-6 compared to the cells that were only treated with LPS and CL separately. Moreover, at higher concentrations of CL, the synergistic effect disappeared. With these results in hand, they suggested that, at lower concentrations of CL, the TLR4 dimer is formed by one monomer containing LPS bound to the MD-2 pocket and the other monomer containing CL. On the other hand, at high concentration of CL compared to LPS, both MD-2 pockets are occupied and, thus, the activation of TLR4 is lower. Based on these results, we wanted to study the binding mode of CL in TLR4/MD-2, which is not known, and to propose a mechanism for the observed synergistic effect.

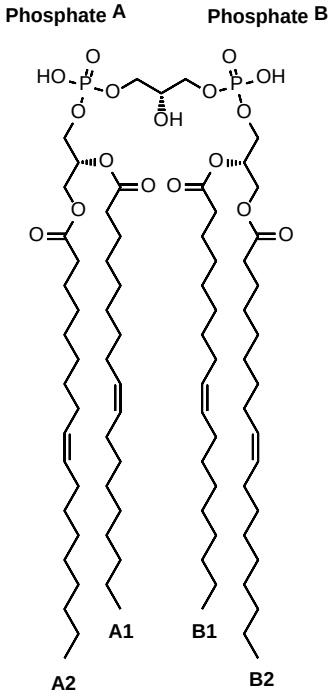


Figure 10: Structure of murine CL. Label of the fatty acid chain is indicated.

5.4.1 Modelling of Cardiolipin

The molecule of murine CL was constructed using Maestro and, after a round of minimization, was docked using AutodockVina on the agonist conformation of TLR4/MD-2 (PDB ID 3FXI). Among all the docked poses (20 in total), we only took into account the ones with all four FA chains inserted into the MD-2 hydrophobic pocket. Only six poses presented the FA chains inserted in this way. All these poses presented common features. The phosphate backbone was predicted to bind at the rim of MD-2, interacting with the polar residues present in that region. For example, in the best docked pose (pose 1, Figure 11), phosphate group of FA chain A, is interacting with Arg 90 of MD-2, while the other is exposed to the solvent. Arg 264 of TLR4 stablishes a hydrogen bond with the oxygen present on the ester bond of chains B. The hydroxyl group of the backbone is oriented and interacting with Arg90 and Glu92 side chains of MD-2. On the other hand, the FA chains are buried inside MD-2 pocket, partially mimicking the interactions of LPS with the hydrophobic residues of MD-2. In fact, thanks to the larger FA chains in CL compared to *E. coli* LPS (18 against 14), CL is able compensate it lower number of FA chains (4 against 6) and mimic the orientation of the FA chains of LPS (Figure 9).

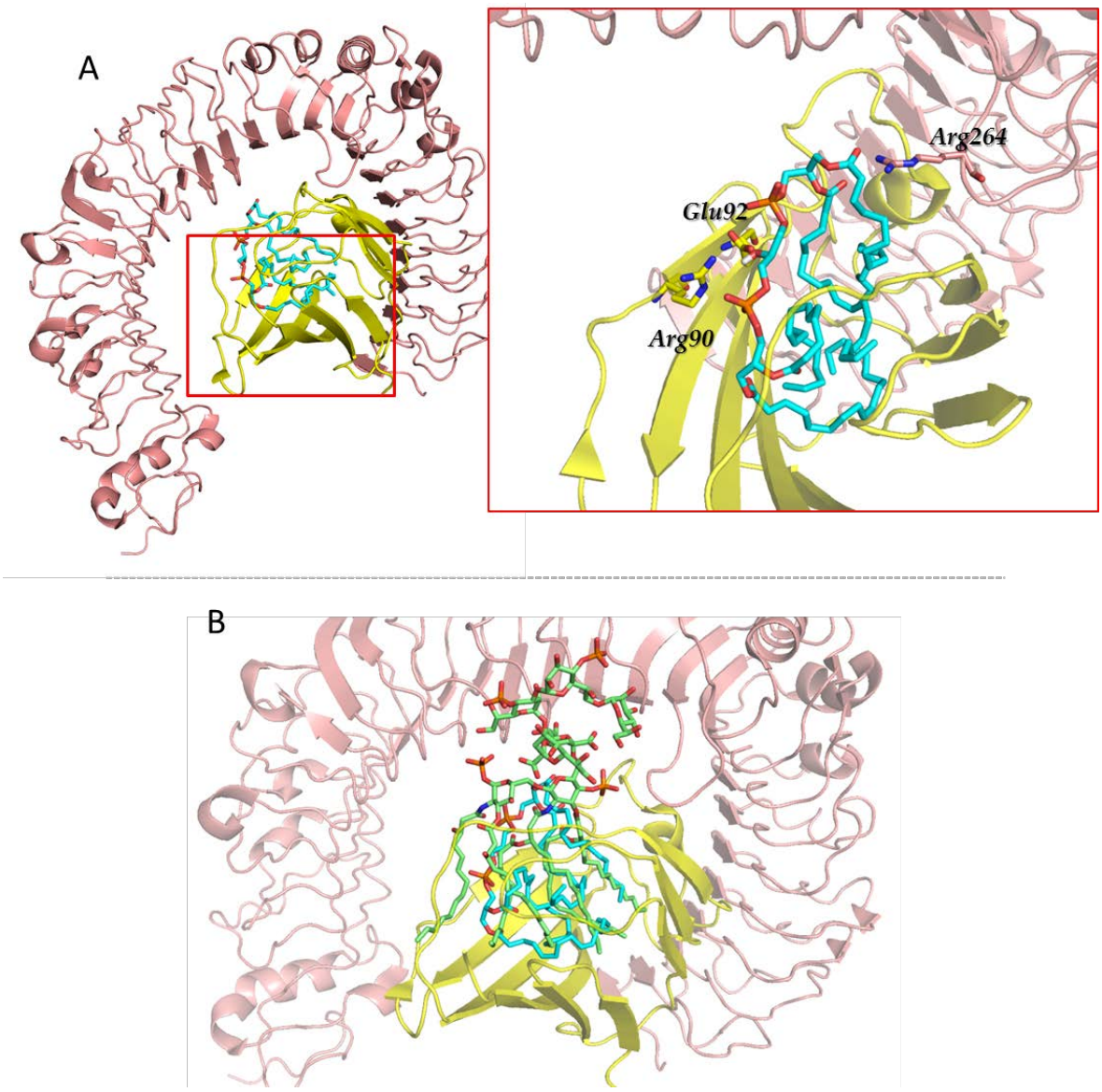


Figure 11: Docking of CL. A) Docked pose 1 of CL (blue) on TLR4/MD-2(salmon/yellow). Polar interacting residues are highlighted in sticks. B) Superimposition of E. Coli LPS (green) with CL docked pose 01 (blue).

Actually, the observed differences between the binding poses are mainly due to the position of the phosphate backbone, since the FA chains are placed in a very similar way in all cases. Thus, the backbone, which is always predicted to be oriented and interacting with the polar residues of the rim of MD-2, seem to have more freedom of movement in the binding pose of CL to TLR4.

We examined with detail the best six docked poses, and we chose pose 1 to perform MD simulation, since apart of being the pose with the best predicted binding energy, presented a binding mode similar to that for LPS. Thus, pose 1 of Cl in complex with TLR4/MD-2 was submitted to 100ns MD simulation, to determine if CL could retain the agonist conformation of MD-2. In fact, after 100ns MD simulation, the complex remained stable and the agonist conformation of Phe126 of MD-2 was retained. The FA chains did not suffer major changes during the simulation, since they were buried inside MD-2 pocket and with very stable interactions.

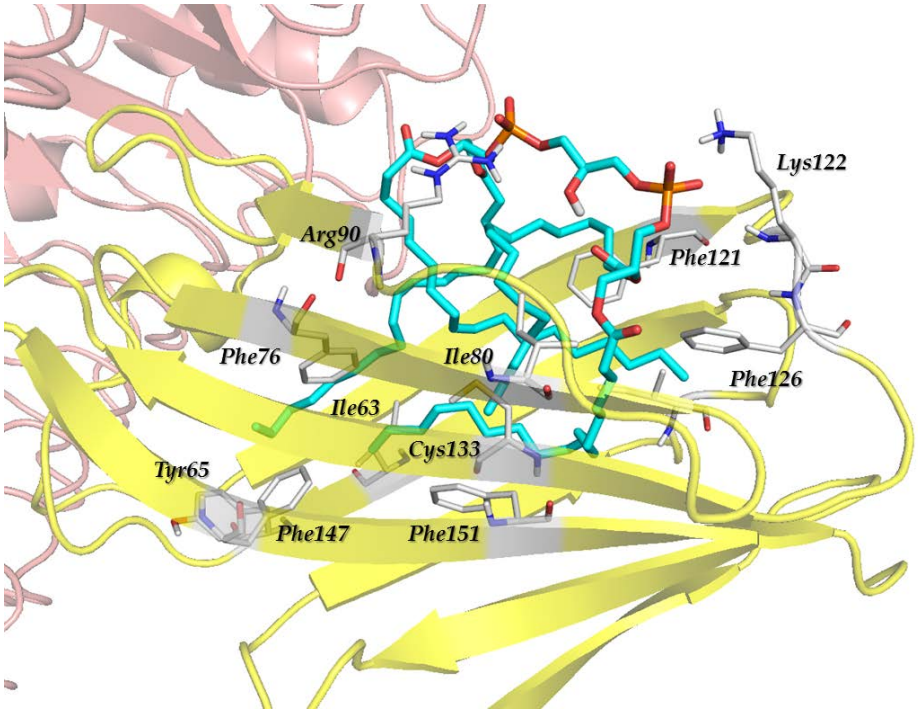


Figure 12: Hydrophobic interactions at the end of 100ns MS simulation of CL docked pose 1 on TLR4/MD-2.

For example, chain B1 interacts with Tyr65, Ile63, Phe76 and Phe147, while chain B2 interacts with Phe121, Phe126 Phe76 and Ile80. On the other hand, chain A1 interacts with Phe121, Ile80 Cys133 and Phe141 and chain A2 with Ile80, Phe737, Ile632 and Phe 126 (Figure 12). Regarding the phosphate backbone, it changes

subtly its orientation (Figure 13). Phosphate A, which was interacting initially with Arg 90, at the end of the simulation is placed on the other side of MD-2, interacting with the side chain of Lys122. Phosphate B, in contact with Lys 337 of TLR4 at the beginning of the simulation, relocates and keeps in contact with Arg676. Arg239 of TLR4 also has an important role in the binding of CL to TLR4, since its interaction with the carboxy group of the backbone is kept during the whole simulation. Thus, in agreement with our results, cardiolipin can bind to the TLR4/MD-2 complex and is able to maintain the agonist conformation of the protein.

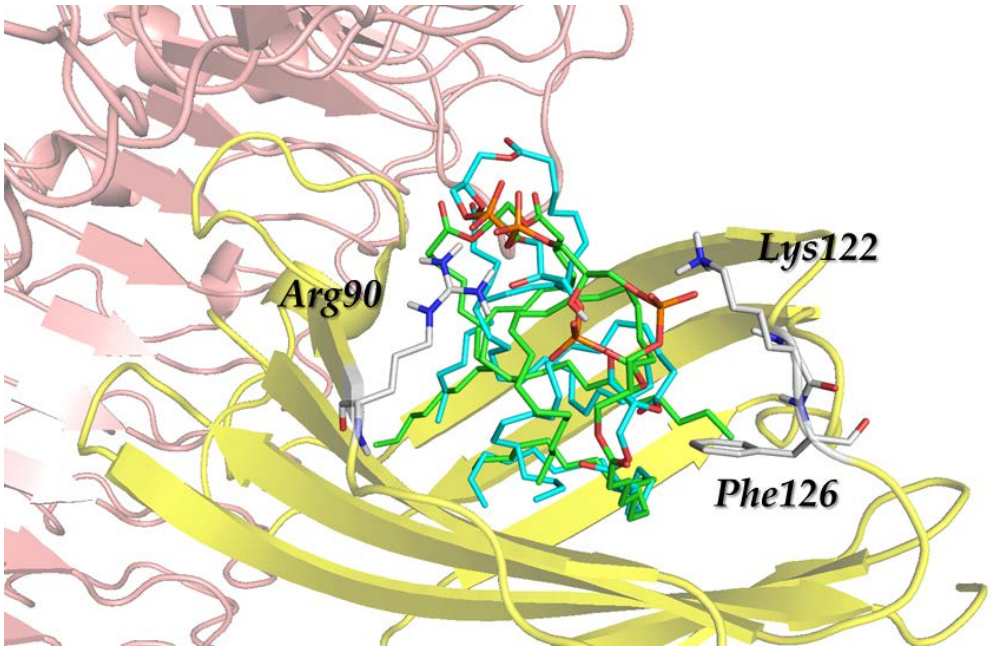


Figure 13: MD simulation of CL docked pose 01. Initial pose is represented in blue lines. Binding pose after green lines.

In order to explore the mechanism of the observed synergic effect, we constructed a mixed heterodimer of LPS and CL *i.e.* a TLR4 dimer complex with a LPS molecule inserted in MD-2 of one monomer and CL inserted in MD-2 of the other monomer. The docked Pose 01 was superimposed to one of the monomers of TLR4/MD-2 complex in 3FXI crystal structure. Then, the template monomer was removed, in order to finally obtain the mixed dimer. The mixed dimer was

submitted to 100ns MD simulation. In both monomers, the agonist conformation of Phe126 side chain was maintained. The overall structure of TLR4/MD-2 dimer was maintained and no major changes in LPS pose can be observed. In the case of CL, the phosphate backbone suffers a minor relocation, but the interactions between MD-2 and the partner TLR4, like Arg90 with Gln436 and Lys125 with Glu422, are not affected. Thus, we propose that CL and LPS can form a stable TLR4/MD2 heterodimer

5.5 Conclusion

In this chapter, we have studied several reported (and yet to be reported) agonist and antagonists of the TLR4/MD-2 system were very distinct chemical nature.

In the case of agonist peptides RS01 and RS09, we proposed a binding mode to the MD-2 protein. RS01 was binding at the rim of the pocket, while RS09 was more buried inside, due to the presence of hydrophobic residues in its structure. Furthermore, after an MD simulation, only RS01 was able to maintain the agonist conformation of MD-2, while Phe126 of the TLR4/MD-2/RS09 complex changed its orientation towards the antagonist conformation. These results were in agreement with the experimental data stating that RS01 was a more potent agonist than RS09.

The binding mode of the antagonist peptide has been also elucidated. This peptide was a mimic of TLR4 binding motif of MD-2. After docking and MD simulations, it was shown that the predicted binding site in TLR4 is the same as for MD-2, thus mimicking the same MD-2 binding motif. In fact, its TLR4 bound conformation was kept in a similar way to that for MD-2. Modifications of the structure of this peptide could lead to the design of a more potent antagonist.

On the other hand, *Bacteroides vulgatus* LPS was an interesting case of study, because it is a mixture of two actually different LPS: tetracylated and pentacylated. Our molecular modelling studies indicate that both LPS_{Bv}-tetra and LPS_{Bv}-penta can bind to the TLR4/MD-2 system in a similar pose to that for LPS_{Ec} (type-A orientation), with some variations on the FA chains and core OS binding, promoting slight differences in the resulting TLR4/MD-2 conformations. The predominance of a conformation close to the agonist conformation of the Phe126-Ile124 MD-2 switch, from our molecular dynamics results, indicates that this LPS_{Bv} mixture could trigger changes at a molecular level that are in the frontier between an agonist and antagonist conformation and this could explain the weak agonist behavior of LPS_{Bv}.

Finally, we have demonstrated that CL binds to TLR4/MD-2. The molecular modelling reveals that the FA chains are inserted in the MD-2 hydrophobic pocket while the phosphate polar head is located at the rim of MD-2, interacting with residues from both TLR4 and MD-2. The FA chains buried in the pocket are stable during the simulation time and the phosphate groups have more freedom of movement, but without affecting the overall binding pose. Moreover, we have shown that a heterodimer of TLR4/MD-2 with a molecule of CL in one monomer and a molecule of *E. coli* LPS in the other monomer can be formed, and that fact could explain the synergistic effect observed *in vitro.*

5.6 Materials and methods

Building of the ligands and the macromolecule. Compounds LPS_{Bv}-tetra and LPS_{Bv}-penta, peptides RS01 and RS09 and cardiolipin, were constructed using Maestro interface. The 3D coordinates of *E. coli* LPS from PDB ID 3FXI were used as template. The resulting structures were minimized using OPLS3 force field,

water as a solvent and steepest descent as the method of minimization, with a maximum of 10000 iterations. The 3D structure of the murine *m*TLR4/MD-2 complex was obtained from the X-ray crystal structure with PDB ID 3VQ2. The 3D structure of human TLR4/MD-2 was obtained from the 3FXI crystal structure. Water molecules, LPS and glycans attached to the structure were removed. Missing hydrogens were added and protonation state of ionizable groups was computed using Protein Preparation Wizard of Maestro. The complex was minimized with 10000 steps of steepest descent method and optimized with OPLS3 force field.

Parameters derivation. Parameterization of new units was performed using Gaussian09 by optimizing the geometries with 6-31G basis set at the Hartree-Fock level of theory. Charges were derived by applying the RESP methodology implemented in Antechamber, assigning the general Amber14 force field (GAFF) atom types.

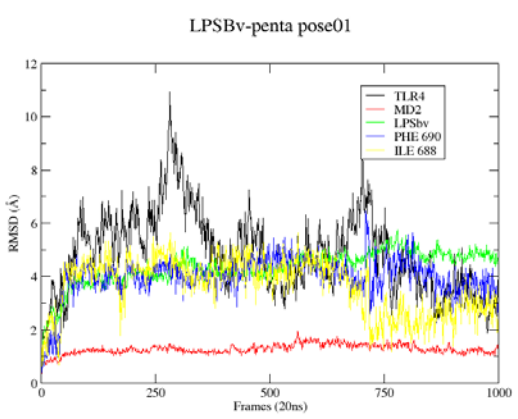
Docking calculations. Docking of compounds LPS_{Bv}-tetra and LPS_{Bv}-penta and CL inside the *m*TLR4/MD-2 system was performed using AutoDock Vina.[24] Agonist peptides RS01 and RS09 were docked with Autodock4. Grid box was centered at the centroid defined among the α-C atoms from Arg 90, Glu 92 and Phe 121 of MD-2, occupying the full MD-2 pocket and part of the TLR4/MD-2 dimerization interface, and 1 Å of grid spacing was set (default value in AutoDock Vina). The spacing in the case of Autodock4 was set to 0.375Å. Box size was set to 34.5 Å in the *x*-axis, 26.25 Å in the *y*-axis and 35.25 Å in the *z*-axis. In the case of AutoDock Vina, exhaustiveness was set as default (200) and 20 predicted binding poses, the maximum allowed in AutoDock Vina, were extracted. The original Lennard-Jonnes and hydrogen-bonding potentials provided by AutoDock4 were also used. After docking, the 200 solutions were clustered in groups with root-mean-square

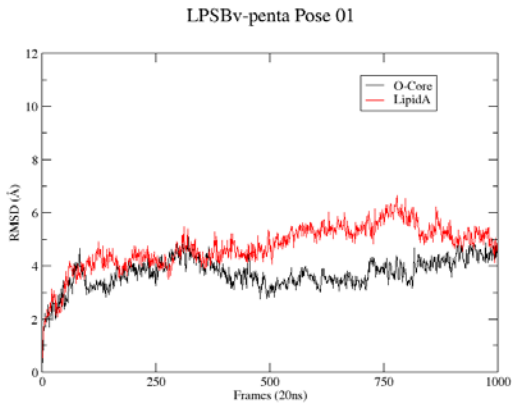
deviation less than 2.0 Ǻ. The clusters were ranked by the lowest energy representative of each cluster.

Molecular dynamics (MD) simulations. All the MD simulations were carried out using Amber14. Several steps of equilibration were performed before running the MD simulation. The first one consisted of 1000 steps of steepest descent minimization followed by 7000 steps of conjugate gradient minimization; a 100 $kcal \cdot mol^{-1} \cdot A^{-2}$ harmonic potential constraint was applied to the proteins and the ligand. In the four subsequent steps, the harmonic potential was progressively lowered respectively to 10, 5, and 2.5 $kcal \cdot mol^{-1} \cdot A^{-2}$) for 600 steps of conjugate gradient minimization each time, and then the whole system was minimized uniformly. In the following step, the system was heated from 0 to 100 K using the Langevin thermostat in the canonical ensemble (NVT) while applying a 20 $kcal \cdot mol^{-1} \cdot A^{-2}$ harmonic potential restraint on the proteins and the ligand. The next step heated up the system from 100 to 300 K in the isothermal–isobaric ensemble (NPT) under the same restraint condition as in the previous step. In the last step, the same parameters were used to simulate the system for 100 ps but no harmonic restraint was applied. At this point, the system was ready for the production run, which was performed using the Langevin thermostat under NPT ensemble, at a 2 fs time step.

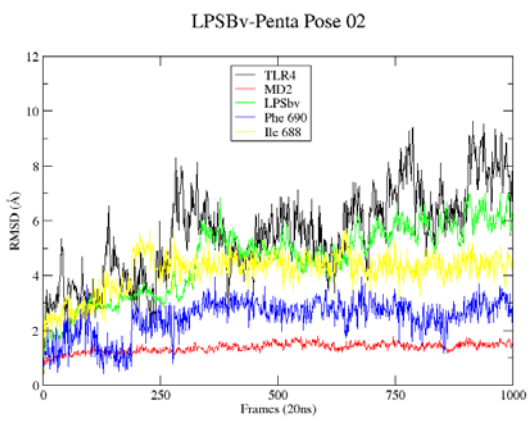
5.7 Annex

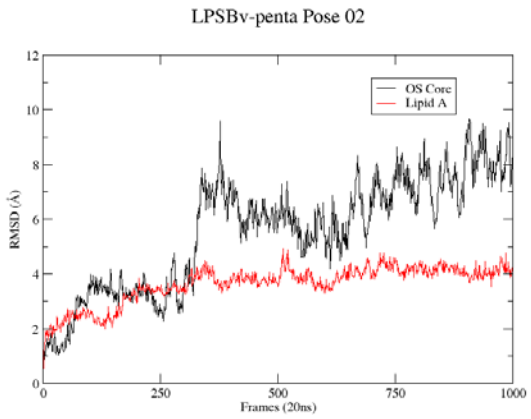
a


LPSBv-penta pose01
TLR4
MD2
LPSbv
PHE 690
ILE 688
RMSD (Å)
Frames (20ns)

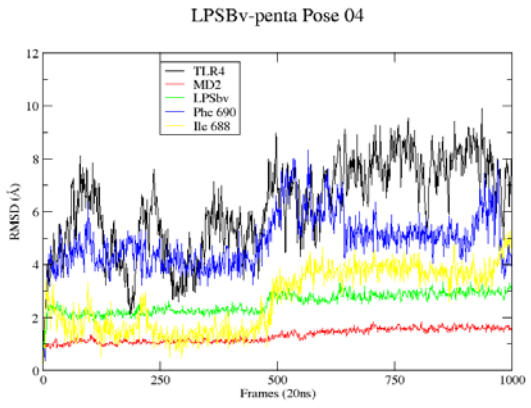

LPSBv-penta Pose 01
O-Core
LipidA
RMSD (Å)
Frames (20ns)

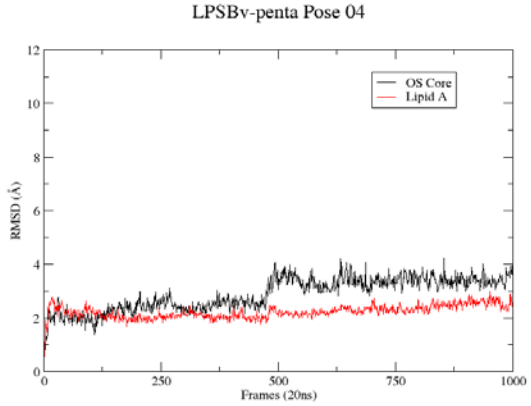
b


LPSBv-Penta Pose 02
TLR4
MD2
LPSbv
Phe 690
Ile 688
RMSD (Å)
Frames (20ns)

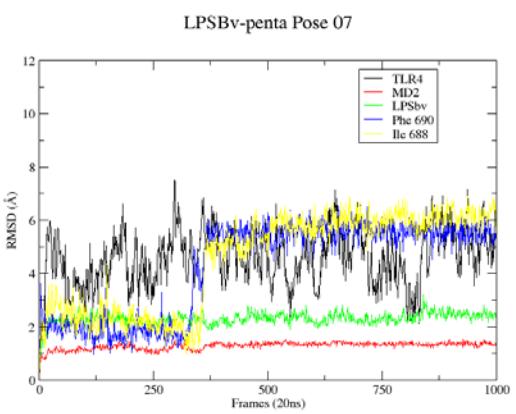

LPSBv-penta Pose 02
OS Core
Lipid A
RMSD (Å)
Frames (20ns)

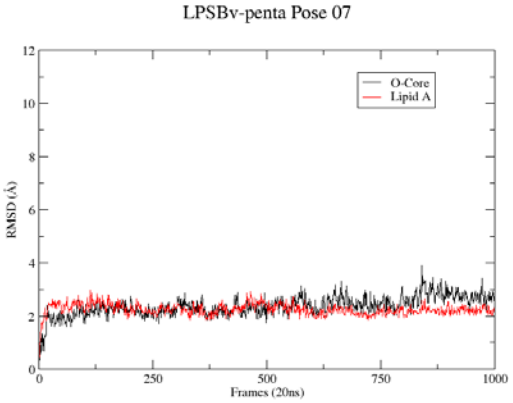
c


LPSBv-penta Pose 04
TLR4
MD2
LPSbv
Phe 690
Ile 688
RMSD (Å)
Frames (20ns)

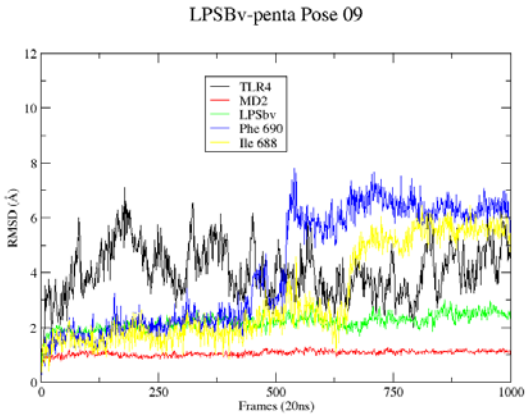

LPSBv-penta Pose 04
OS Core
Lipid A
RMSD (Å)
Frames (20ns)

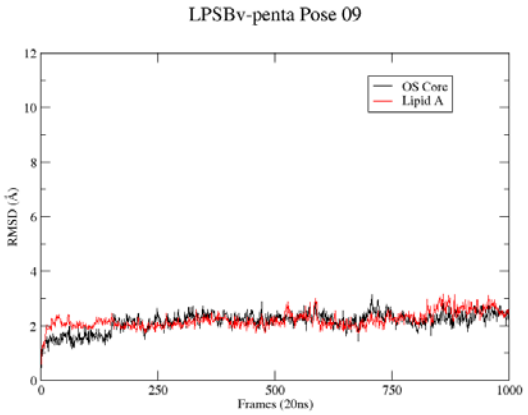
d


LPSBv-penta Pose 07
TLR4
MD2
LPSbv
Phe 690
Ile 688
RMSD (Å)
Frames (20ns)

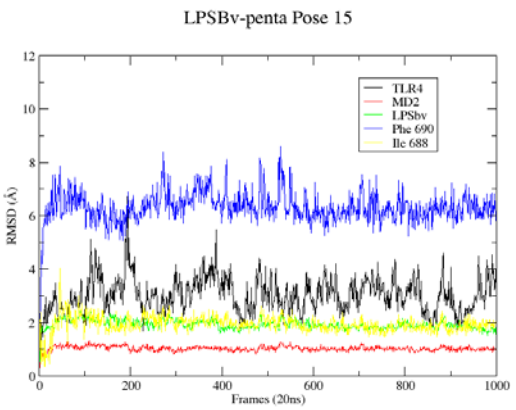

LPSBv-penta Pose 07
O-Core
Lipid A
RMSD (Å)
Frames (20ns)

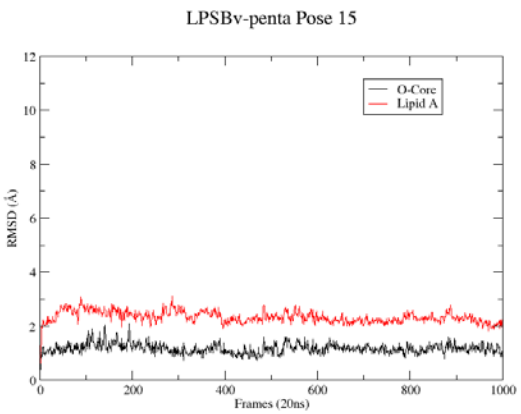
e

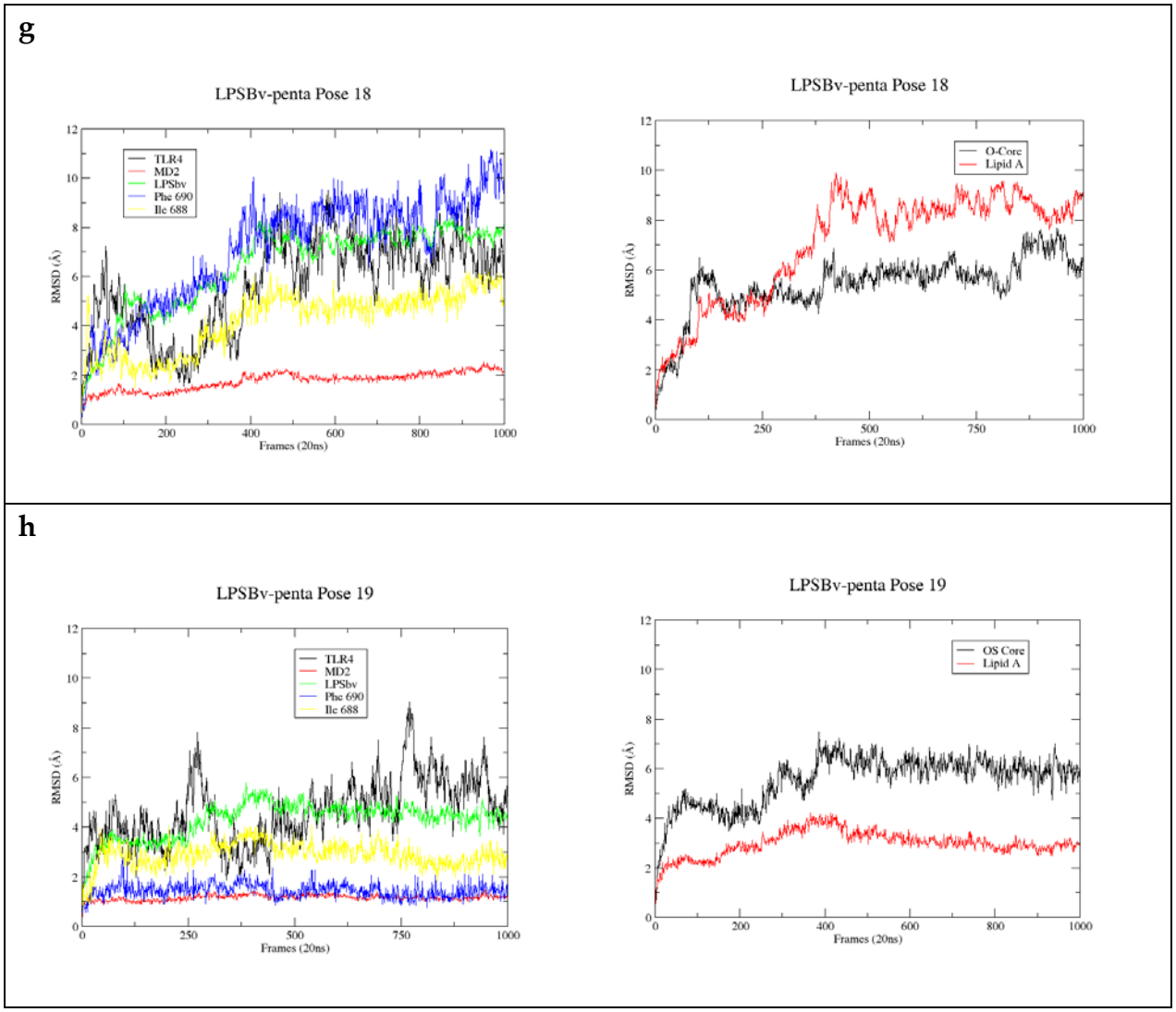

LPSBv-penta Pose 09
TLR4
MD2
LPSbv
Phe 690
Ile 688
RMSD (Å)
Frames (20ns)


LPSBv-penta Pose 09
OS Core
Lipid A
RMSD (Å)
Frames (20ns)

f

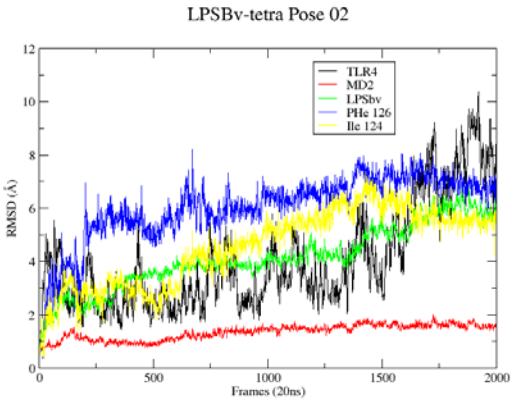

LPSBv-penta Pose 15
TLR4
MD2
LPSbv
Phe 690
Ile 688
RMSD (Å)
Frames (20ns)

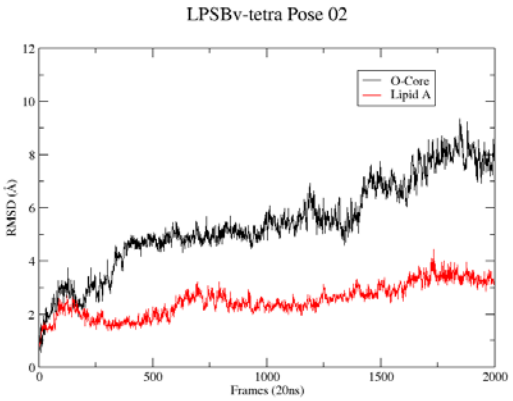

LPSBv-penta Pose 15
O-Core
Lipid A
RMSD (Å)
Frames (20ns)



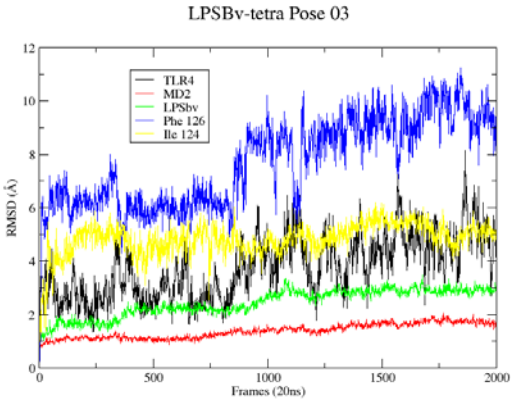
Annex Figure 1: RMSD values along the 20ns MD simulations of eight TLR4/MD-2/LPS$_{Bv}$-penta complexes. RMSD (Å) values (y-axis) of TLR4/MD-2/LPS$_{Bv}$-penta complexes from docking predicted binding poses 01 (a), 02 (b), 04 (c), 07 (d), 09 (e), 15 (f), 18 (g) and 19 (h) along 20ns of MD simulation (x-axis). First column represents the RMSD of TLR4 (black line), MD-2 (red line) LPS$_{Bv}$ (green line), Phe 126 (blue line) and Ile 124 (yellow line). Second column represent the RMSD of LPS$_{Bv}$ separated in core OS (black line) and lipid A (red line). All RMSD were calculated with ptraj. Each frame has been collected every 20 ps of the simulation. Hydrogens were not included in the calculation in all cases and in the case of TLR4 and MD-2 values, only α-carbons were considered in the calculation.

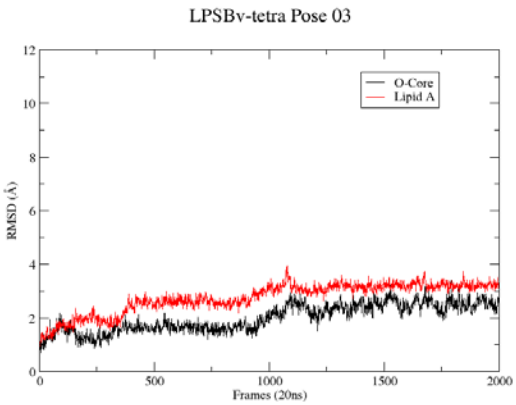
a


LPSBv-tetra Pose 02
TLR4
MD2
LPSbv
PHe 126
Ile 124
RMSD (Å)
Frames (20ns)

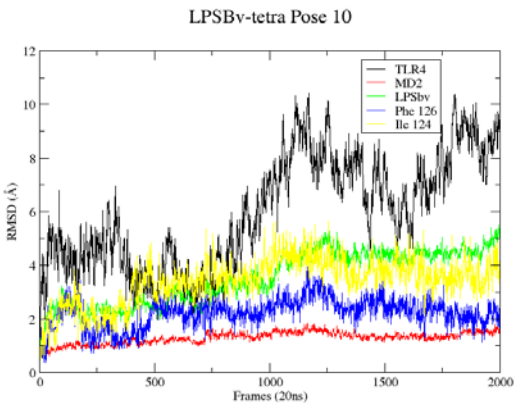

LPSBv-tetra Pose 02
O-Core
Lipid A
RMSD (Å)
Frames (20ns)

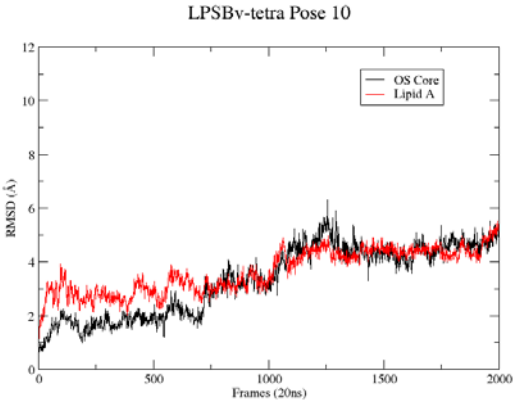
b

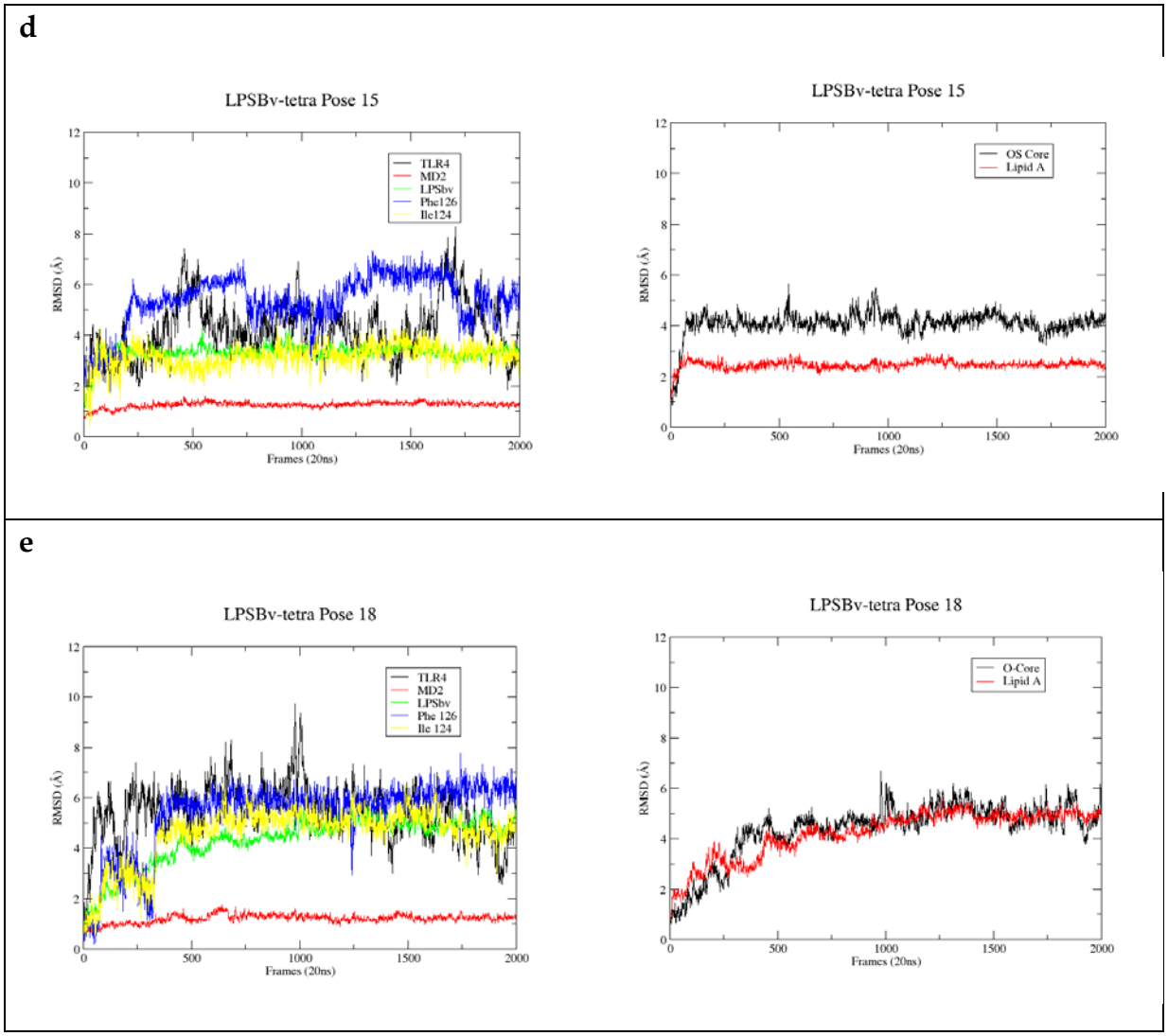

LPSBv-tetra Pose 03
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)


LPSBv-tetra Pose 03
O-Core
Lipid A
RMSD (Å)
Frames (20ns)

c

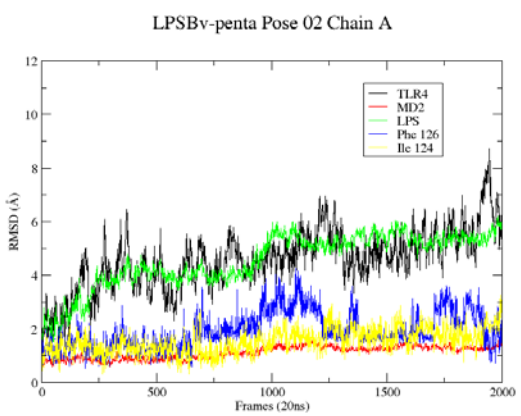

LPSBv-tetra Pose 10
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

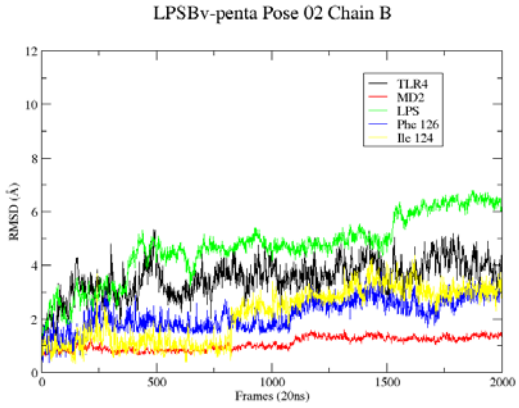

LPSBv-tetra Pose 10
OS Core
Lipid A
RMSD (Å)
Frames (20ns)



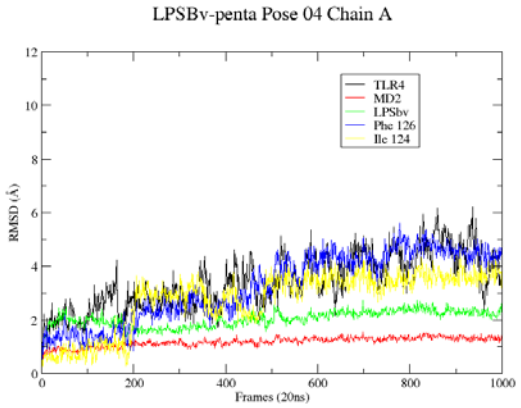
Annex Figure 2: RMSD values along the 20ns MD simulations of five TLR4/MD-2/LPS$_{Bv}$-tetra complexes. RMSD values of TLR4/MD-2/LPS$_{Bv}$-penta complexes from docking predicted binding poses 02 (a), 03 (b), 10 (c), 15 (d) and 18 (e). First column represents the RMSD of TLR4 (black line), MD-2 (red line) LPS$_{Bv}$ (green line), Phe 126 (blue line) and Ile 124 (yellow line). Second column represents the RMSD of LPS$_{Bv}$ separated in core OS (black line) and lipid A (red line). All RMSD were calculated with ptraj. Each frame has been collected every 10 ps of the simulation. Hydrogens were not included in the calculation in all cases and in the case of TLR4 and MD-2 values, only α-carbons were considered in the calculation.

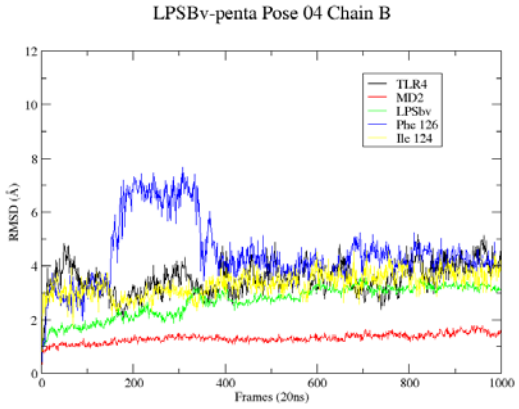
a


LPSBv-penta Pose 02 Chain A
TLR4
MD2
LPS
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

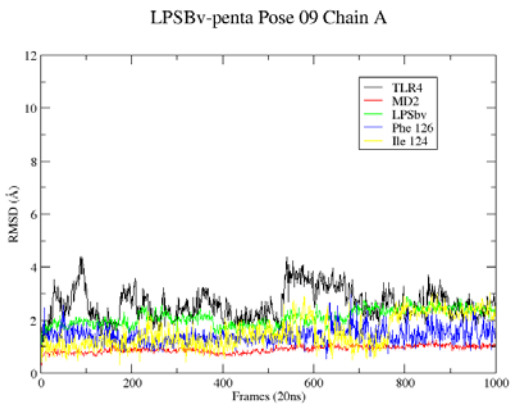

LPSBv-penta Pose 02 Chain B
TLR4
MD2
LPS
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

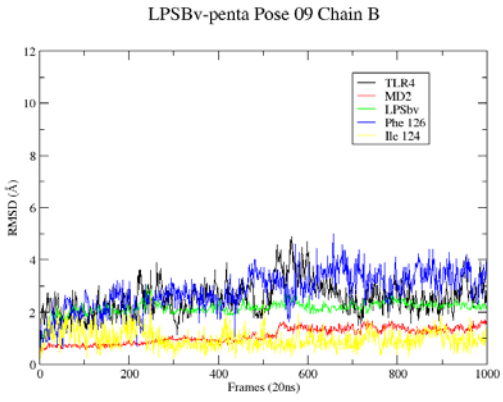
b


LPSBv-penta Pose 04 Chain A
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

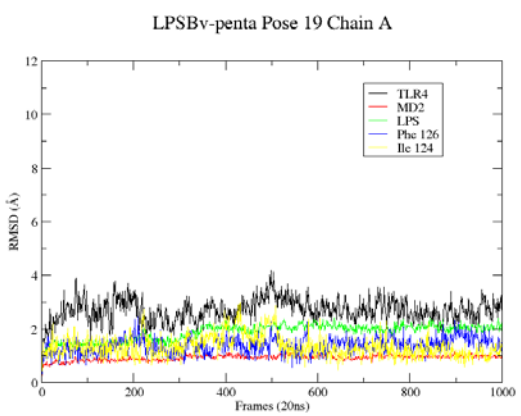

LPSBv-penta Pose 04 Chain B
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

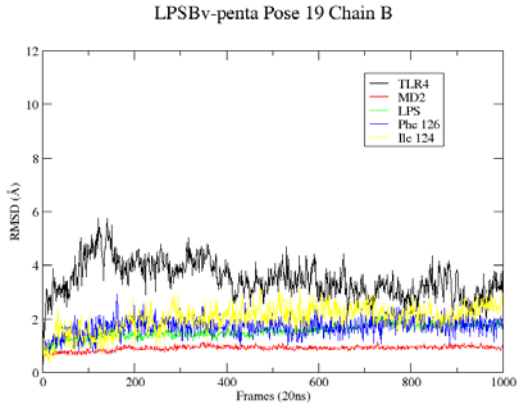
c


LPSBv-penta Pose 09 Chain A
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

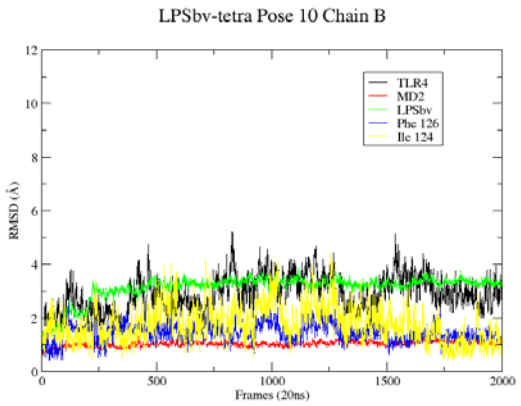

LPSBv-penta Pose 09 Chain B
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

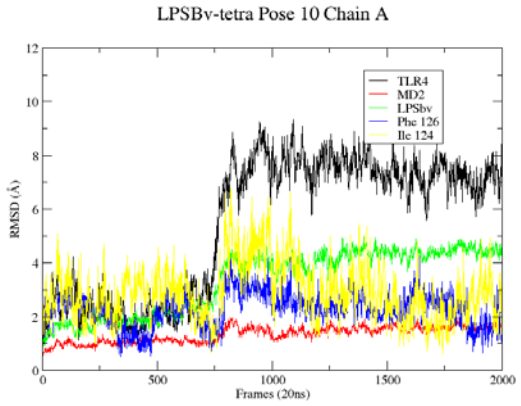
d


LPSBv-penta Pose 19 Chain A
TLR4
MD2
LPS
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

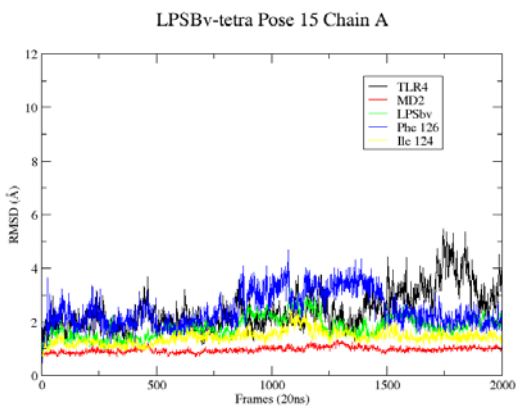

LPSBv-penta Pose 19 Chain B
TLR4
MD2
LPS
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

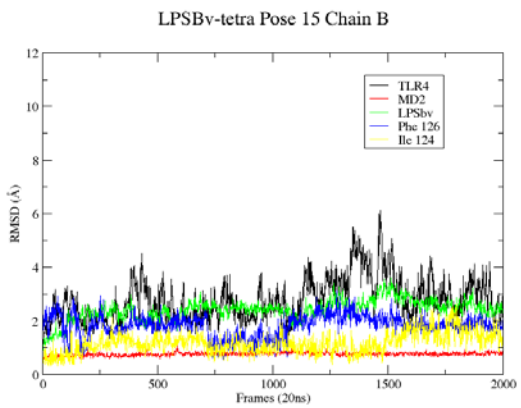
e


LPSbv-tetra Pose 10 Chain B
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)


LPSBv-tetra Pose 10 Chain A
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

f


LPSBv-tetra Pose 15 Chain A
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)


LPSBv-tetra Pose 15 Chain B
TLR4
MD2
LPSbv
Phe 126
Ile 124
RMSD (Å)
Frames (20ns)

Annex Figure 3: RMSD values along the 20ns MD simulations of the six TLR4/MD-2/LPS$_{Bv}$/TLR4/MD-2*/LPS$_{Bv}$* full heterodimer models. RMSD values of both LPS$_{Bv}$-penta and LPS$_{Bv}$-tetra full heterodimer models after 20ns of MD simulation. The models were constructed from an average structure of the monomers MD simulation and using 3VQ2 crystal structure as template. (a) LPS$_{Bv}$-penta pose 02; (b) LPS$_{Bv}$-penta pose 04; (c) LPS$_{Bv}$-penta pose 09; (d) LPS$_{Bv}$-penta pose 19; (e) LPS$_{Bv}$-tetra pose 10; (f) LPS$_{Bv}$-tetra pose 15. First column represents the RMSD values of the first monomer (TLR4/MD-2/LPS$_{Bv}$, residue number 1-735) and the second column represents the RMSD values of the second monomer (TLR4*/MD-2*/LPS$_{Bv}$*, residue number 736-1470). Black line, TLR4; Red line, MD-2; Green line, LPSBv; Blue line, Phe 126; Yellow line, Ile 124. All RMSD were calculated with ptraj. Each frame has been collected every 10 ps of the simulation. Hydrogens were not included in the calculation in all cases and in the case of TLR4 and MD-2 values, only α-carbons were considered.*

Bibliography

1. Shanmugam, A.; Rajoria, S.; George, A. L.; Mittelman, A.; Suriano, R.; Tiwari, R. K., Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PloS one* **2012,** *7* (2), e30839.
2. Smith, G. P., Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **1985,** *228* (4705), 1315-7.
3. Ohto, U.; Fukase, K.; Miyake, K.; Satow, Y., Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. *Science* **2007,** *316*, 1632-4.
4. Duan, G. J.; Zhu, J.; Wan, J. Y.; Li, X.; Ge, X. D.; Liu, L. M.; Liu, Y. S., A synthetic MD-2 mimetic peptide attenuates lipopolysaccharide-induced inflammatory responses in vivo and in vitro. *Int Immunopharmacol* **2010,** *10* (9), 1091-100.
5. Visintin, A.; Latz, E.; Monks, B. G.; Espevik, T.; Golenbock, D. T., Lysines 128 and 132 enable lipopolysaccharide binding to MD-2, leading to Toll-like receptor-4 aggregation and signal transduction. *J. Biol. Chem.* **2003,** *278* (48), 48313-20.
6. Re, F.; Strominger, J. L., Separate functional domains of human MD-2 mediate Toll-like receptor 4-binding and lipopolysaccharide responsiveness. *J. Immunol.* **2003,** *171* (10), 5272-6.
7. Mullen, G. E.; Kennedy, M. N.; Visintin, A.; Mazzoni, A.; Leifer, C. A.; Davies, D. R.; Segal, D. M., The role of disulfide bonds in the assembly and function of MD-2. *Proc Natl Acad Sci U S A* **2003,** *100* (7), 3919-24.
8. (a) Steimle, A.; Autenrieth, I. B.; Frick, J. S., Structure and function: Lipid A modifications in commensals and pathogens. *Int. J. Med. Microbiol.* **2016,** *306* (5), 290-301; (b) Steimle, A.; Gronbach, K.; Beifuss, B.; Schafer, A.; Harmening, R.; Bender, A.; Maerz, J. K.; Lange, A.; Michaelis, L.; Maurer, A.; Menz, S.; McCoy, K.; Autenrieth, I. B.; Kalbacher, H.; Frick, J. S., Symbiotic gut commensal bacteria act as host cathepsin S activity regulators. *Journal of autoimmunity* **2016,** *75*, 82-95.
9. Gronbach, K.; Flade, I.; Holst, O.; Lindner, B.; Ruscheweyh, H. J.; Wittmann, A.; Menz, S.; Schwiertz, A.; Adam, P.; Stecher, B.; Josenhans, C.; Suerbaum, S.; Gruber, A. D.; Kulik, A.; Huson, D.; Autenrieth, I. B.; Frick, J. S., Endotoxicity of lipopolysaccharide as a determinant of T-cell-mediated colitis induction in mice. *Gastroenterology* **2014,** *146* (3), 765-75.
10. Dudek, J.; Hartmann, M.; Rehling, P., The role of mitochondrial cardiolipin in heart function and its implication in cardiac disease. *Biochimica et biophysica acta. Molecular basis of disease* **2019,** *1865* (4), 810-821.
11. Pangborn, M. C., The composition of cardiolipin. *J. Biol. Chem.* **1947,** *168* (1), 351-61.

12. (a) Ghosh, S.; Iadarola, D. M.; Ball, W. B.; Gohil, V. M., Mitochondrial Dysfunctions in Barth Syndrome. *IUBMB life* **2019**; (b) Musatov, A.; Sedlak, E., Role of cardiolipin in stability of integral membrane proteins. *Biochimie* **2017,** *142*, 102-111.
13. Pizzuto, M.; Lonez, C.; Baroja-Mazo, A.; Martinez-Banaclocha, H.; Tourlomousis, P.; Gangloff, M.; Pelegrin, P.; Ruysschaert, J. M.; Gay, N. J.; Bryant, C. E., Saturation of acyl chains converts cardiolipin from an antagonist to an activator of Toll-like receptor-4. *Cellular and molecular life sciences : CMLS* **2019**.

CHAPTER 6:

Identification of non LPS-like TLR4 modulators by virtual screening tools

6.1. Introduction

Identification of drug-like molecules with potential therapeutic applications for the treatment of TLR-related diseases has attracted considerable interest due to their clinical potential. TLR modulators have the potential to be used with different biomedical applications, especially in the field of infection,[1] inflammation[2] and autoimmune diseases,[3] and also in cancer[4] and in central nervous system (CNS) disorders such as Alzheimer´s disease.[5] Even though several inhibitors of the TLR4/MD-2 complex acting on MD-2 were found in the literature, only a minority shows promising characteristics to become an available drug. Eritoran for example showed promising results in phase I and II clinical trials, but in phase II failed in showing better properties than existing treatments for sepsis.[6] In fact, just few candidates are currently under clinical development due to the difficulty to find molecules with appropriate physic-chemical properties and low toxicity.[7]

Therefore, it is imperative to find new chemical entities, and not necessary with LPS-like structure, as TLR modulators with drug-like properties in order to facilitate their development as drugs. There are some small molecules that exemplify this possibility. For example, some pyrimido[5,4-b]indoles that have shown to stimulate TLR4 and could potentially be used as adjuvants or immune modulators;[8] synthetic analogues of natural product euodenine A have exhibited potent and selective agonist towards TLR4;[9] and synthetic peptides to mimic the TLR4/LPS interaction have also been reported.[10] Also several small non LPS-like molecules with TLR4 antagonist activity have been developed, such as ethyl 4-oxo-4-(oxazolidin-3-yl)-butenoate derivatives (OSL07),[11] benzothiazole-based inhibitors,[12] ethyl phenyl-sulfamoylcyclohexenecarboxylate derivatives (TAK-242 or resatorvid),[13] and β-amino alcohol derivatives.[14]

In the context of drug discovery, virtual screening (VS) techniques have already proved to make hit identification more goal-oriented, allowing the access to a huge number of chemically diverse binders (from public and commercial databases) with a relatively low-cost in terms of time and materials (see Chapter 2, section 3). This computational approach has been subjected to extensive attention and revision over the years, from the early perspective of being an emerging method,[15] until the current time where new challenges are faced.[16] These VS approaches constitute a current strategy in drug design for the identification of novel chemical entities with a given binding ability.[17]

We could say that TLRs are not standard receptors which could be approached following classical strategies in drug design. The complexity of the system and the characteristics of their complexation with the pathogen associated molecular patterns (PAMPs) make them especially difficult to tackle following classical procedures in drug design and discovery. This is why TLRs constitute a special case study in this context. Specifically, on the field of TLR4 research, VS studies have been recently reported leading to novel ligands with drug-like properties, trying to overcome the solubility problems associated with LPS mimetics (see Chapter 1, section 1.4.1). Among these works, Joce *et al*.[18] have developed a novel *in silico* screening methodology including molecular mechanics and implicit solvent methods to incorporate the evaluation of binding free energies and have screened the Enamine database collection.[19] The resulting clusters were filtered by selecting the representative compounds that were submitted to fast molecular docking for the generation of binding poses and subsequent MD simulations to rank the ligand poses according to their predicted binding affinities. Final filtering led to the identification of compounds T5342126 and T6071187 (Figure **5.1**) as small drug-like inhibitors of the TLR4/MD-2 protein-protein interactions. Their biological activity

and selectivity were tested *in vitro*, and their TLR4/MD-2 antagonist activity was confirmed. In other study, Švajger *et al*.[20] performed parallel ligand-based and structure-based VS in order to identify novel TLR4 antagonists targeting the TLR4/MD-2 interface, by using the ZINC drug-like subset (~11.3 million drug-like compounds) from the ZINC database.[21] The identified ligands after ligand-based VS resulted in being either insoluble in water, or inactive, or presented cytotoxicity on HEK293 cells. However, the structure-based VS identified 40 putative TLR4/MD-2 ligands that were assessed *in vitro*. After the first assays, only 14 compounds were sufficiently water-soluble and completely non-cytotoxic at 100 μM. These compounds received further biological evaluation, and finally, three compounds with promising antagonistic activities were discovered: ZINC25778142, ZINC49563556 and ZINC3415865 (Figure 1).

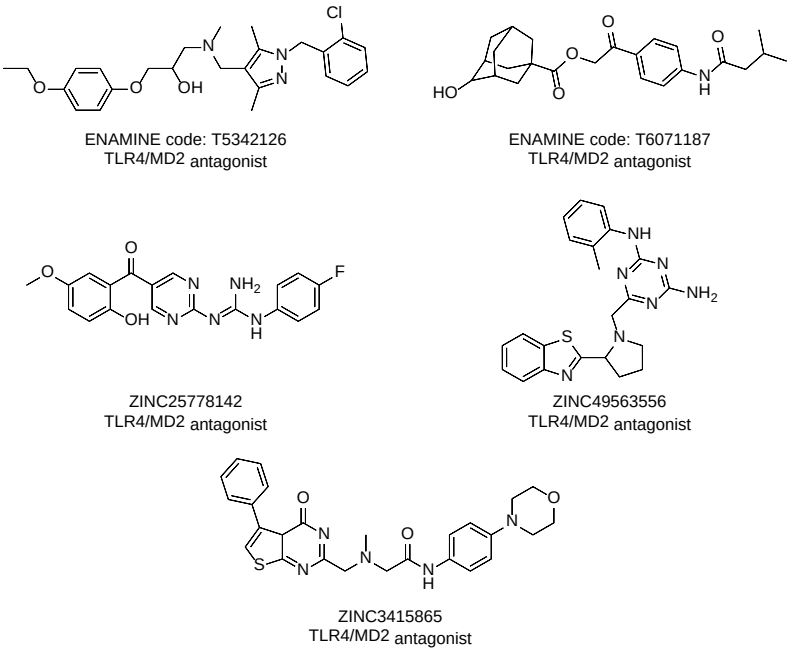


Figure 1: Novel TLR4/MD-2 modulators identified by VS approaches.

On the other hand, despite the enormous effort of spent time and money on research and development, the number of new drugs brought to market drastically decreases each year.[22] Significant investments by pharmaceutical companies for

optimizing the drug discovery pipeline have been undertaken, and new techniques such as structure-based drug design, combinatorial chemistry and high throughput screening (HTS) techniques have emerged. However, the impact of these innovations has not been as important as it was expected both in short and long term.[23] Drug repositioning (also known as drug repurposing, drug redirecting, or drug reprofiling) is a process of discovering new uses outside the scope of the original medical indication for existing drugs. Before 2004, no traces of this process have been found in the literature,[24] but it has gained an increasing attention within the international drug development community over the last few years, and represents a new promising direction.[25]

Different terms are used to describe drug repositioning, but all mean a way to find new indications for existing drugs or potential drug candidates, including those in clinical development where mechanism-of-action is relevant to multiple diseases: drugs that have failed to demonstrate efficacy for a particular indication during Phase II or Phase III trials but with no major safety concerns; drugs that have been discontinued for commercial reasons; marketed drugs for which patents are close to expiry; and drugs candidates from academic institutions and public sector laboratories that have not been fully pursued yet, are also took into account. In this way, drug repositioning represents unique translational opportunities, and is believed to offer great benefits over the *de novo* drug discovery, reducing the development risks and timeline to potentially 3-12 years,[25f] substantially increasing the probability of success to brought drugs into market due to existing knowledge about the drugs, and providing relatively inexpensive solutions as therapies for rare and neglected diseases[26] that frequently offer limited potential revenue to pharmaceutical companies. Successful repurposing examples, discovered by

serendipity, are sildenafil (Viagra®), acetyl salicylic acid (Aspirin®),[27] and thalidomide.[28]

In this work, we aimed to identify novel TLR4 modulators with non LPS-like structure by means of virtual screening following a "computed-aided drug repositioning" approach. We have screened almost 3,000 approved drugs and drug-like molecules, and identified several compounds with TLR4 antagonist activity. This approach has shown to be a promising and efficient tool for discovering new uses from existing drugs and holds the great potential for precision medicine in the age of big data.

6.2 Results and Discussion

-Receptors

There are several available 3D structures of TLR4, as hetero/homo-dimers, and in complex with some ligands (agonists and antagonists) and/or co-receptors.[29] In the case of the agonist conformation of the *h*TLR4/MD-2 monomer complex, 3D coordinates from TLR4/MD-2 heterodimer were obtained from the PDB (PDB-ID: 3FXI).[30] In the case of the antagonist conformation, since the full crystallographic structure of the *h*TLR4/MD-2 complex is not available, a model built by us was used. This model was built using the human MD-2 protein in antagonist conformation (PDB-ID: 2E59)[31] superimposed onto the MD-2 subunit of the agonist full complex (PDB-ID: 3FXI chain C) through PyMOL (see Chapter 3). Also in order to consider different antagonist conformations of TLR4, we used PDB-ID: 2E56 (only in the case of SPECS and Log P 1000 databases).

-Databases

Database processing constitutes a fundamental step in VS approaches. It is crucial to generate the proper chemical library, with the adequate geometries, ionization states, conformations, etc. Furthermore, it is very important to discard any molecule that will not be a good candidate in the further steps of the VS study in relation to the particular system on hand. A good database processing will assure a rigorous and well-conducted VS, as well as it will avoid computational cost and identification of unsuitable drug candidates.

Different commercial, public and *in-house* databases have been used: Log P 1000, SPECS and ZINC as commercial databases, and as *in-house* databases, a diversity collection of compounds from laboratories of Prof. Péter Mátyus (PM) from Semmelweis University (Budapest), Prof. Jose Carlos Menéndez[32] (JCM) from Complutense University of Madrid, Prof. J. R. Pedro (JRP),[33] and Prof. A. Marco (AM)[34] from the University of Valencia.

Commercial databases

ZINC15 (ZINC Is Not Commercial 2015)[21, 35] is a public access database and tool set, developed to enable ready access to compounds for VS, ligand discovery, pharmacophore screens, benchmarking, and force field development. Nowadays ZINC15 database contains over 120 million purchasable compounds. For the purpose of this work, we were only interested in the approved compounds which represented, that time, a total of 2459 structures categorized under the substance subset called **WORLD** that is standing for approved drugs in major jurisdictions, including the FDA. Being a computational drug repositioning study, the compounds present in the ZINC15 database were filtered by clinically approved drugs. Thus 2459 from the WORLD subset over 100 million compounds in total

were kept for the repurposing study. These compounds were submitted to a preparation process and the number of compounds increased from 2459 to 2949.

Log P 1000 dataset[36] a small diverse subset of the ZINC database.21, 37 The subset was obtained by a similarity search based on 128 molecular VolSurf+ descriptors38 covering biologically relevant properties such as shape, surface, volume, molecular weight, polar surface area, hydrogen bonding capacity, lipophilicity and solubility. This was followed by an additional elimination of permanently charged compounds and compounds with a molecular weight lower than 150 Da, resulting in a diverse set of drug-like compounds.

SPECS dataset is a database of commercially available drug-like compounds.[39] Due to the large number of compounds (almost 300000) and computational limitations, a reduction of the final screening set was necessary. The MOE software[40] was used to perform a cluster analysis on which a diverse subset was created. Therefore the fingerprint of each molecule was calculated in form of a bit-packed version of the molecular access system (MACCS) structural keys (BIT_MACCS),[41] encoding 166 unique features. The Tanimoto coefficient was used as a measure of similarity between fingerprints.[42] A similarity of 85% was used for the cluster search. This resulted in a reduced diverse subset of SPECS comprising 23774 compounds that were used for screening.

In-house databases

We selected *in-house* collections with a wide range of chemical structures from different collaborators expert in different types of chemistry. It is important to mention that these chemical libraries are available to perform the biological assay, in the case of these compounds give very good results in the VS studies.

First, we used the diversity collection of heterocyclic compounds, based on their interesting structural characteristics, from PM database with around 1 964 molecules; a second *in-house* dataset with quinoline, quinazoline and acridine structures from JCM database, with 68 compounds; and a third *in-house* collection of 25 and 85 compounds from JRP and AM, respectively, including pyrroles, indoles, naptholes, heterocyclic derivatives from JRP and analogues of natural products colchicine and pironetin from AM library. All of them had the available samples to be tested in case there were successfully screened. It is important to mention that, given that paclitaxel had shown antagonistic activity in human TLR4, while agonistic activity in mouse TLR4, showing the species-specific ligand recognition by MD-2, we were prompted to include tubuline binders in our VS approach. Prof. Marco is a well-recognized synthetic chemist specialized in the synthesis of natural products analogues, being analogues to tubuline binders among them.

The binding of paclitaxel to TLR4 had been demonstrated, however not the induction of the cytokine response. Based on these reported results, we included other tubulin binders and related compounds as putative TLR4 ligands in order to discover novel TLR4 modulators. We chose a family of compounds analogue to natural products colchicine and pironetin (Figure 2). Regarding their antitumoral activity, and their ability to bind to tubulin components and microtubules, paclitaxel is a tubulin-interacting drug that stabilizes microtubules, while colchicine causes disruption of microtubules, and pironetin derivates bind to α-tubulin, inhibiting tubulin assembly. These opposite effects are due to the different tubulin sites with which they interact. We also included compounds derived from stilbene, like resveratrol since they are studied for their antimitotic properties and their antitumor activity, all of them included in the AM database.

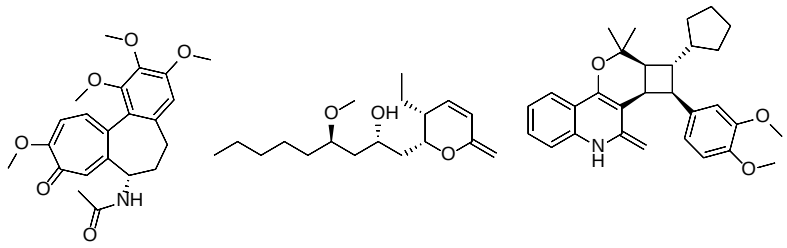


Figure 2: Colchicine, pironetin and euodenine A structures.

Filtering

In order to prepare the databases for the VS, different tautomers were considered according physiological pH leading to the corresponding increase in the total number of screened compounds (see Materials and Methods). Finally, in this study, a database composed by around five hundred thousand compounds was built, including known binders (data from the literature) and decoys.

We have considered the following filters:

1- Lipophilicity of the molecules: a maximum logP of 6 were considered, taking into account that the natural LPS and reported synthetic glycolipids have a logP very high: 29.14 ± 0.83, 14.35±0.73 and 13.53±0.47 for lipid IVa, P01 and ONO-4007 respectively. This limit is a reasonable margin above the value of 5 according to Lipinski´s rules (oral bioavailability).
2- Molecular weight (MW): we considered a wide range between 300 and 700 Da given the MW of glycolipids targeting TLR4, with a reasonable margin above the value of 500 according to Lipinski´s rules.
3- pH: only possible tautomers at physiological pH were considered within a range of 7± 0.5.
4- Prediction of favorable binding from at least two docking programs and in two different conformations of TLR4.

Searching for TLR4/MD-2 Modulators: Virtual Screening

-General overview of the VS protocol

Molecular docking screening was performed against the different databases based on both, the agonist conformation of *h*TLR4/MD-2 complex from PDB-ID: 3FXI, and our modeled antagonist conformation of *h*TLR4/MD-2 complex. Ligand Based (LBVS) and Structure Based (SBVS) VS were carried following the protocols showed in the Figure **5.3**. Log P 1000 and SPECS databases were submitted to LBVS (step III) and SBVS (step I) with the FLAP tool. WORD (subset ZINC database) and *in-house* databases were submitted to SBVS with the combined Glide/AutoDock/VINA approach (step II). The resulting screened compounds (189 in total) were re-docked (step IV) by means of FLAP and Glide programs, finally yielding the selection of 27 compounds that were experimentally tested (step V). Seven compounds were identified as TLR4 antagonists.

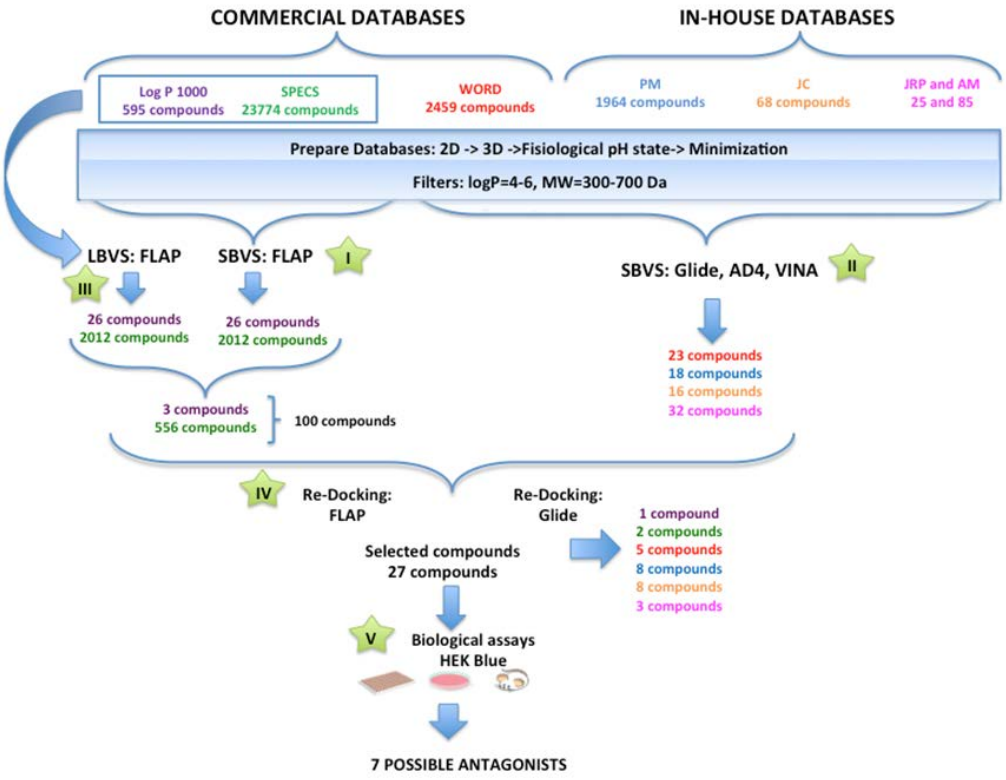


Figure 3: Flow chart of the VS protocol.

-Performance of the VS study

STEP I. Structure-Based Virtual Screening (SBVS) with FLAP

The investigated compounds from Log P 1000, SPECS datasets are drug-like molecules and not glycolipids like the known binders lipid A, lipid IVa or eritoran. The lipid chains in those glycolipids contain a vast number of free bonds which would increase the docking time exponentially. An additional complication would probably arise from the circumstance that most scoring functions are calibrated on drug-like molecules and would likely have difficulties evaluating the interactions correctly, especially in the entropic term, due to the many degrees of freedom and the large hydrophobic surface.[43]

The literature defines three categories of compounds that may inhibit MD-2.[44] The first class consists of inhibitors that compete with LPS for the binding in the hydrophobic pocket but without being able to trigger the final dimerization; Paclitaxel is one example for this class of inhibitors.[45] Molecules of the second category bind covalently to the residue Cys133; compound JTT705 is one example.[46] The final class of inhibitors does not enter the hydrophobic pocket completely but binds in the opening region of the cavity and prevent LPS from entering the pocket; representatives of this category are compounds JSH, curcumin, xanthohumol and isoxanthohumol (Figure 4).[47] The majority of the side chains of the residues form the MD-2 pocket are hydrophobics (Leu, Ile, Phe and Val), but the rim of the cavity, on the other hand, contains almost no hydrophobic residues. The surface contains many positively and negatively charged amino acids which are important for the interaction between TLR4 and MD-2.

FLAP's SBVS method was used to perform target based VS on the TLR4/MD-2 receptor with SPECS and Log P 1000 databases. As a benchmark, the method was

applied initially to the set of known active compounds. The result is shown in Table 1 with the ligands ranked according to their *Glob-Sum* score. This score is a global similarity score calculated by summing the four single contributions: shape (H), hydrogen-bond acceptor (N1), hydrophobic (DRY), and hydrogen-donor acceptor (O) descriptors; *Glob-Sum* is the global sum of all four energy values. Note that the scores of the single contributions are derived from the individual best conformation for this type of score, which might be different, while the *Glob-Sum* score comes from the one conformation for which the sum of scores is maximal.

Antagonist	Glob-Sum	Antagonist	Glob-Sum
Paclitaxel	3.245	6-shogaol	2.498
JSH	2.714	Isoxanthohumol	2.465
Curcumin	2.695	Isoquiritigenine	2.239
1D10G	2.669	Cinnamaldehyde	2.179
CAPE	2.621	C34	2.136
Xanthohumol	2.611	OSL7	1.799
JTT705	2.513	Sulforaphane	1.707

Table 1: Known antagonists of MD-2, ranked descending by Glob-Sum score obtained from structure-based virtual screening (SBVS).

Since the *Glob-Sum* score does not reflect any experimental binding affinity, the results of the known ligands allow having an idea at which value a screened ligand can be considered as a potential hit. The highest score was obtained by paclitaxel (*Glob-Sum*=3.245) which was then used as a cutoff value for the screened unknown ligands. From Log P 1000 and SPECS libraries, 26 and 2012 compounds were obtained, respectively, having a score equal to or higher than 3.245. The highest contribution to the global score is given by the hydrophobic score which can easily be explained by the high hydrophobicity of the target pocket and the screening model that is obtained from it.

To better understand the interactions of the potential inhibitors retrieved by LBVS and SBVS with TLR4/MD-2, a molecular re-docking approach was carried out. In order to narrow down the number of compounds to dock, only molecules were selected which obtained a good score in the LB and the SBVS approaches. From the SBVS, in total, 2038 compounds (26 from Log P 1000 and 2012 from SPECS) obtained a score higher than the cutoff value of 3.245. Since an analogous cutoff value was not available for the LBVS approach, the same number of compounds was chosen here, *i.e.* the top ranked 26 and 2012 compounds from Log P 1000 and SPECS, respectively.

Of the Log P 1000 set 3 common compounds were found in the top ranks of both LBVS and SBVS, while SPECS shared 556 top-ranked compounds. This total number of 559 compounds still seemed large, considering the time-consuming FLAP docking program. For this reason only the top 100 highest scoring ligands were taken for the final re-docking (step IV). This selection procedure was found to be in agreement with examples from the literature.[48]

STEP II. Structure-Based Virtual Screening (SBVS) with Glide, AutoDock and VINA

WORD database from ZINC and *in-house* databases (PM, JCM, JRP and AM) were docked into both agonist and antagonist protein conformations, using three docking programs, Glide, AutoDock and VINA, to avoid the limitation of one scoring function. The receptor grid was set up in order to fully contain the *E. coli* LPS, allowing small molecules to interact with the entire MD-2 pocket, as well as its rim and its entrance (see Materials and Methods). During the docking process, all the ligands were kept to facilitate visual inspections, comparisons and selections between the three docking programs. 50 poses per ligand were generated with AutoDock, 20 poses per ligand with VINA (which is the maximum for the program), and only one pose per ligand was generated with Glide, using HTVS, SP

and XP protocols in order to also facilitate the comparisons, choosing Glide as the main docking software. For the docking program validation analysis, either with, Glide, AutoDock or VINA, the scoring results for all the compounds were consistent and correlated to each other. However, the correlation between AutoDock and VINA is stronger than between Glide and AutoDock or VINA. The docked compounds, as well as all their corresponding predicted binding poses, were visually analyzed to detect any computational errors. The docking scores, defined by the average score of all the poses from one ligand for each docking program, were analyzed. Among all the compounds, according to each scoring function, only the top 25% from each docking program has been kept for the next analysis step.

Among the databases, 23, 18, 16 and 32 compounds were selected respectively, being ranked at the top 25% for at least two docking programs at the same time and for one or both conformations, prioritizing a correlation with Glide, were kept for visual cluster rank analysis.

II.a Molecular Docking Using Glide

The molecules were subjected to a grid-based ligand docking with energetics (Glide, Schrodinger, version 6.9)[49] using the Virtual Screening Workflow protocol (See Materials and Methods). Regarding the docking step parameters, Epik state penalties for docking were used, and the non-polar part of the ligand potential were soften by scaling the van der Waals radii of ligand atoms with small partial charges. The full workflow includes three docking stages, each step differing from the preceding step in the amount of time taken to dock each molecule and the scoring system used to evaluate each pose. The first stage performs HTVS (High Throughput Virtual Screening) docking. The ligands that are retained are then

passed to the next stage, which performs SP (Standard Precision) docking. The survivors of this stage are passed onto the third stage, which performs XP (eXtra Precision) docking, a more powerful and discriminating procedure. The Dock flexibility method was used for HTVS, SP and XP dockings allowing us to penalize non-planar conformation for amide bonds. A post-docking minimization was also performed, as well as constraints for the docking stages. One pose per compound state was generated and 100% of the best compounds that passed the HTVS, SP and XP docking have been kept. For HTVS and SP docking, all states have been retained, but only the best scoring state for the XP docking.

II.b Molecular Docking Using AutoDock and VINA

Docking was also performed independently with both VINA[50] and AutoDock.[51] In AutoDock 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 50 to sample more docked poses.

VINA (Vina Is Not AutoDock) uses an Iterated Local Search global optimizer[52] based on Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which approximates Newton's method and the number of docking poses was set to 20, which is the maximum for the program. TLR4/MD-2 receptors were kept rigid and the ligands were set partially flexible (*i.e.* maximum of 32 dihedral angles) for AutoDock and totally flexible for VINA.

II.c Screened Compounds from Combined Glide/AutoDock/VINA Results

Selected screened compounds from each docking program (steps II.a and II.b) were visually inspected and those binding outside the MD-2 were discarded. Finally, only the top 10% in the case of WORD and PM databases, and 20% from JCM, JRP and AM databases, were kept. Among them, 89 compounds were ranked

at the top for at least two docking programs at the same time, and were predicted to bind into one (at least) of the two TLR4 conformations (agonist/antagonist). These 89 compounds were then submitted to the following analysis step: (step IV) re-docking with AutoDock and Glide.

STEP III. Ligand-Based Virtual Screening (LBVS) with FLAP

Only in the case of the Log P 1000 and SPECS commercial database, we also performed LBVS with the FLAP tool.[53] The FLAP LBVS method uses the common reference framework to align a set of candidate molecules to the template binder, to find the optimal overlap according to the GRID MIFs. The similarity between the fields is quantified by the Tanimoto coefficient. In the output table the user can see the individual scores obtained by the single MIF contributions (*Glob-Prod*), as well as a global score representing the sum (*Glob-Sum*) for each compound.

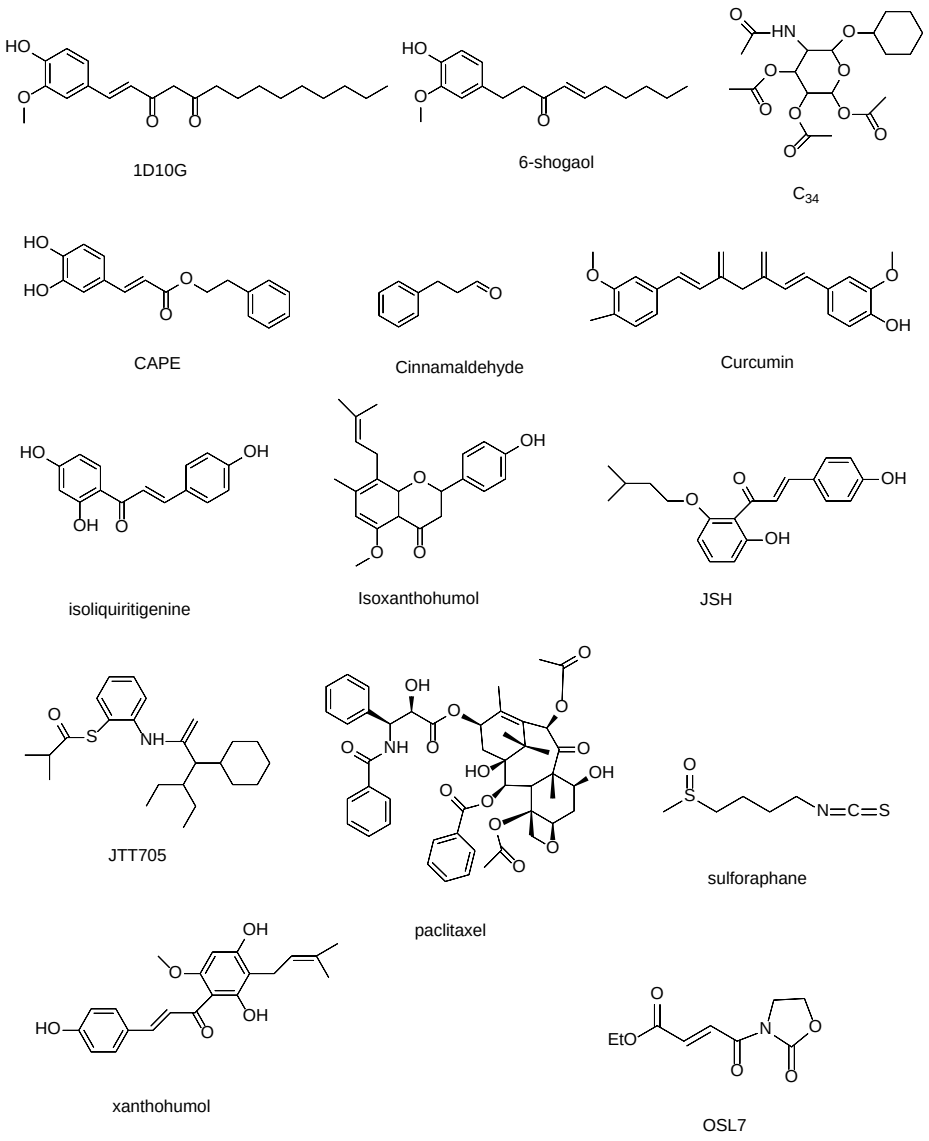


Figure 4: Known antagonists of the MD-2 reported in the literature.

For the LBVS with FLAP, a set of known active antagonists of MD-2 was built based on a literature search (Figure 4). The two datasets Log P 1000 and SPECS were screened on each known active separately and ranked by their obtained *Glob-Sum* scores.

LBVS was here performed individually by using the 14 known ligands as templates for the screening. The best ranked results are shown in Table 2. The 2D representations of the compounds of Log P 1000 and SPECS can be found in Annex Figure 1 and 2, respectively. The similarity between template and test molecule of the single contributions is a value between 0 (no similarity) and 1 (high similarity).

The four single contributions are shape (H), hydrogen-bond acceptor (O), hydrogen-bond donor (N1) and hydrophobic (DRY) potential. The global score value that was used as result for the screening analysis is the *Glob-Sum* which is the global sum of all four energy values.[54] Note that the scores of the single contributions are derived from the individual best conformation for this type of score, which might be different, while the *Glob-Sum* score comes from the one conformation for which the sum of scores is maximal.

Template	Compound	Glob-Sum	H	N1	DRY	O
6-shogaol	152	1.326	**0.663**	0.508	0.224	0.239
	481	1.742	0.598	0.271	0.254	**0.702**
Xanthohumol	568	1.269	**0.699**	0.283	0.340	0.124
	19907	1.912	0.703	0.368	0.508	0.359
Paclitaxel	383	0.847	**0.505**	0.175	0.134	0.337
	20513	1.022	**0.565**	0.171	0.105	0.321
1D10G	368	1.152	**0.579**	0.203	0.181	0.310
	20700	1.857	**0.654**	0.306	0.260	0.734
JSH	492	1.165	**0.598**	0.359	0.229	0.144

	21315	1.421	**0.515**	0.371	0.304	0.329
Isoliquiritigenine	42	1.181	**0.637**	0.364	0.195	0.010
	120	1.706	**0.750**	0.431	0.343	0.294
Isoxanthohumaol	138	1.054	**0.638**	0.234	0.308	0.010
	28	1.493	**0.634**	0.430	0.304	0.305
CAPE	575	1.149	**0.625**	0.242	0.181	0.243
	22298	1.528	0.587	0.230	0.159	**0.700**
Curcumin	548	1.041	**0.631**	0.242	0.204	0.010
	23010	1.562	0.519	0.264	0.173	**0.623**
Sulforaphane	46	1.104	**0.650**	0.361	0.166	0.000
	3203	1.184	**0.684**	0.296	0.000	0.000
Cinnamaldehyde	40	1.489	**0.684**	0.581	0.383	0.000
	23599	1.500	0.580	**0.673**	0.273	0.000
OSL7	35	1.007	**0.648**	0.295	0.128	0.000
	1171	1.285	**0.702**	0.445	0.191	0.000
C34	187	1.142	**0.506**	0.205	0.137	0.512
	10959	1.428	0.560	0.216	0.102	**0.903**
JTT705	439	1.033	**0.539**	0.391	0.188	0.010
	14650	1.127	**0.592**	0.347	0.188	0.000

Table 2: Best ranked compounds of Log P 1000 (blue) and SPECS (white) set for each known ligand (black).

Table 2 shows that for each of the known actives the best scoring SPECS compound scored higher than the best scoring one from the Log P 1000 database. This could be explained by the sole fact that the SPECS set contain a much higher number of compounds than Log P 1000. Consequently the probability is higher to find a good scoring compound.

Regarding the single contributions of the four similarities, the shape similarity (H) seems to have the highest impact on the global score in most of the cases. In four cases (6-shogaol, CAPE, Curcumin, C34) the hydrogen-bond acceptor and in one case the hydrogen-bond donor (N1) similarity made the biggest contribution to the global score.

All five compounds are from the SPECS set. The reason why the influence of hydrophobic (DRY) similarity is comparatively low might be the relatively small size of the compounds. While strong hydrogen-bond similarities can be derived from single donor or acceptor atoms, the hydrophobic potential needs larger apolar surfaces to show a strong impact.

STEP IV. Re-docking with FLAP and Glide

The screened compounds from the above steps (189 compounds in total) were re-docked by means of FLAP and Glide programs, finally leading to the selection of 27 compounds. This selection was based in: i) the agreement in the most probable clusters from both programs; and ii) in the visual analysis of the best clusters from both docking programs with special attention to the ligand/receptor interactions (discussed below).

The finally selected compounds were: **1** compound from Log P 1000 (**ID-5382**), **2** compounds from SPECS (**AG-690/11203225**, and **AF-399/1512855**) (Table 3); **5** compounds from WORD database (compounds **146**, **157**, **177**, **179**, and **208)** (Annex Table 1)**; 8** compounds from **PM** (**PM1097, PM1811, PM1779, PM567, PM1090, PM810, PM1758**, and **PM1200**) (Annex Table 2), **8** from **JCM** (**MS14, MS20, MS21, MS32, MS35, MS40, MS45**, and **MS49**) (Annex Table 3) and **3** compounds from **JRC** and **AM** (**JRP07, JRp07p**, and **JRP10**) (Annex Table 4)**.**

Regarding the three ligands from the Log P 1000 and SPECS datasets, those were initially screened with FLAP, had an S-score equal to or higher than the threshold

of 1.074, obtained by the best-scoring known inhibitor sulforaphane. The compounds, their 2D description and the respective scores are listed in Table 3. The highest scoring compound is **ID-5382** from the Log P 1000 set, with an S-score of 1.231. The two compounds of the SPECS set **AG-690/11203225** and **AF-399/15128553** obtained a score of 1.114 and 1.074 respectively.

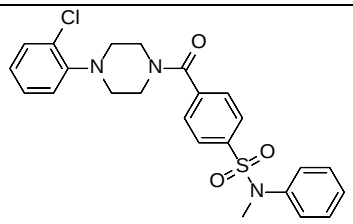
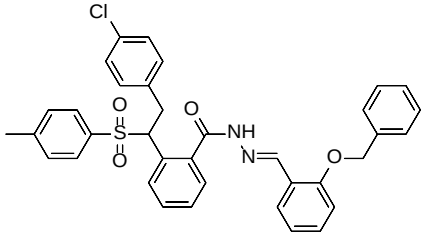
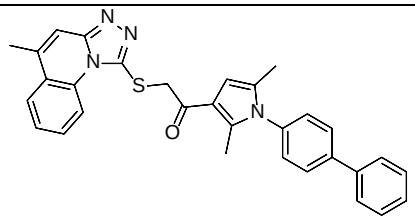
Compound	S-score	Structure
ID-5382	1.231	
AG-690/11203225	1.114	
AF-399/15128553	1.074	

*Table 3: 2D description and the respective scores from **ID-5382**, **AG-690/11203225** and **AF-399/15128553**.*

When studying the docked poses, in the three cases (**ID-5382**, **AG-690/11203225** and **AF-399/15128553**) the docked ligand is located at the entry of the hydrophobic pocket of MD-2 adopting similar poses. The principal interactions are hydrophobic and polar ones. The three compounds show polar interactions with Arg90 and Lys122. The hydrophobic interactions are more wide spread and not with the same set of amino acids for the three compounds. It is observed that compound **ID-5382**

is in close contact with the hydrophobic residues Ile46, Leu61, Leu78, Phe121, Ile124, Val135 and Phe151. Compound **AG-690/11203225** interacts with the residues Ile52, Phe76, Leu78, Ile80, Val82, Glu92, Phe121, Ile124, Val135 and Ile153. Finally, the hydrophobic interactions of **AF-399/15128553** were established with Ile46, Leu61, Ile80, Val82, Leu87, Phe121, Ile124, Tyr131 and Phe151 side chains. (Figure 5)

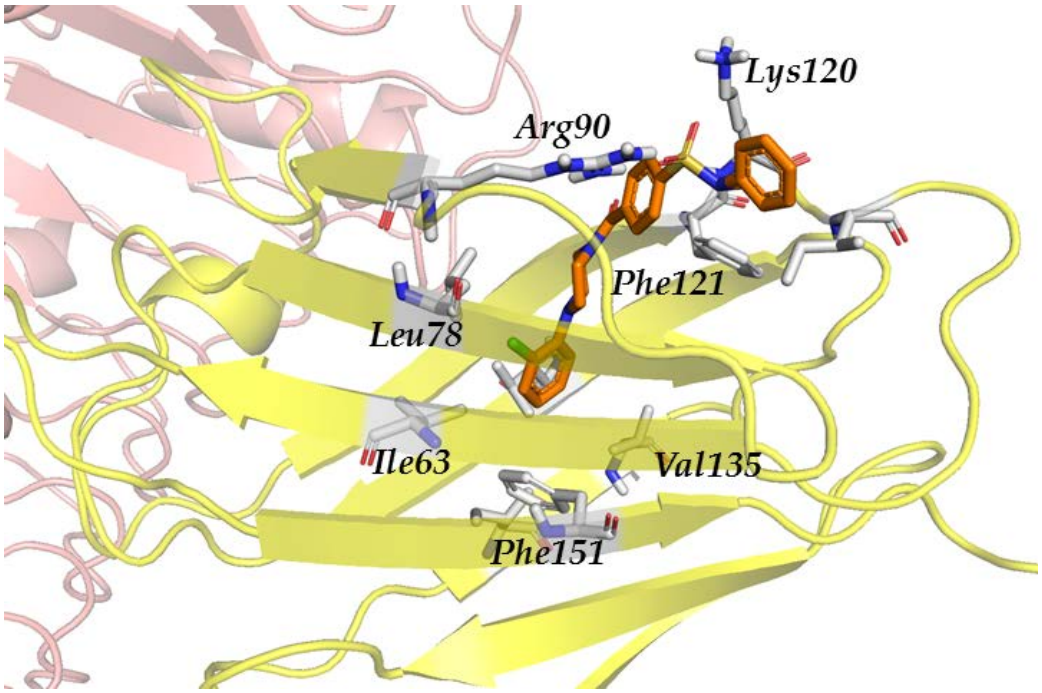


Figure 5: Redocked ID-5382 in TLR4/MD-2.

The two amino acids, Arg90 and Lys122, are able to build salt bridges with the compounds **ID-5382** and **AG-690/11203225** due to their sulfonyl group. This would explain why the polar score is significantly higher for these two ligands than for compound **AF-399/15128553** which possesses no sulfonyl groups. The latter one only forms hydrogen-bonds between Arg90 and Lys122 side chains and the basic nitrogen atoms from the triazol ring. Interactions with Cys133, as those reported in the literature,[55] could not be observed, due to the inability to predict/model covalent bonds with FLAP.

From JRP and AM databases, the selected compounds stablished stacking interactions with Phe76, and CH-π interactions are observed with the side chain of Cys133, Phe151, Phe104 and Leu61. Other interactions observed are hydrophobic ones with the residues Val24, Ile32, Ile44, Val48, Ile52, Leu78, Ile80, Ile94, Ile117, Phe119, Val135 and Ile153.

Regarding the compounds from PM databases, they stablish π-π interactions with Phe104 and Phe151, also CH-π interactions with Phe76 and Phe121. Other interactions observed are hydrophobic with Ile32, Ile52, Leu61, Ile117, Val135, leu149 and Ile153.

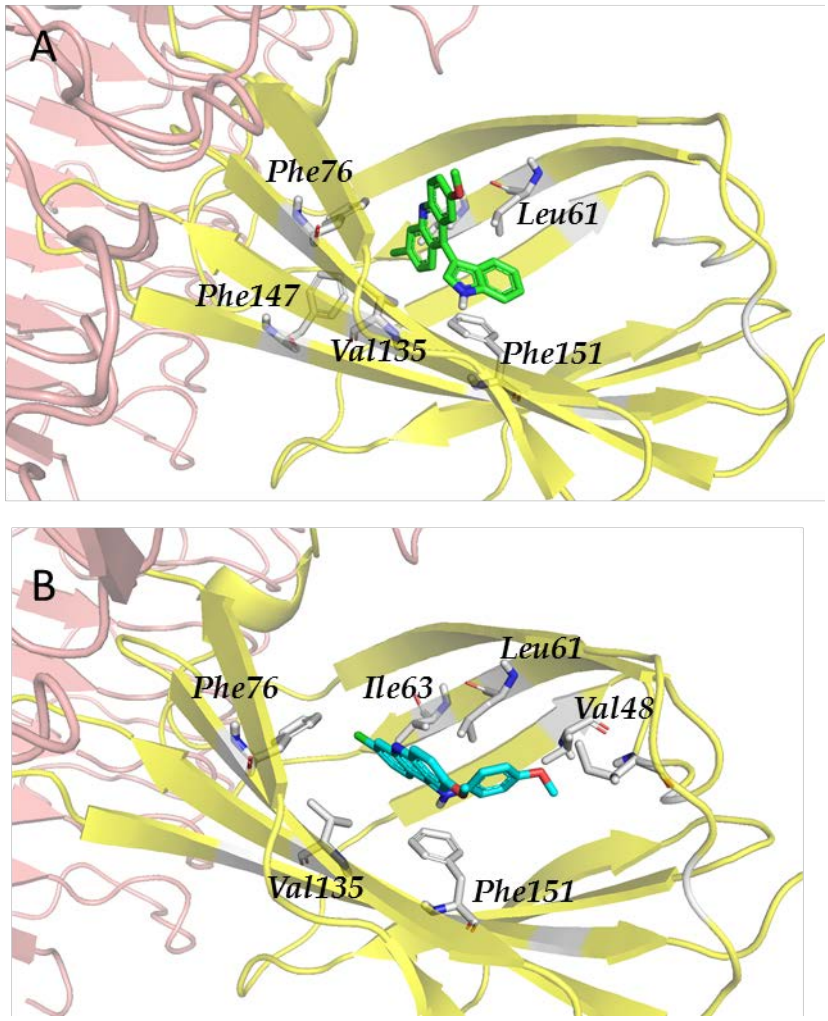


Figure 6: Redocked MS-21 (A) and MS-32 (B) in TLR4/MD2.

Finally, in the case of the compounds from JCM databases, the principal interactions observed are: π-π interactions with Phe76 and Phe151 and also hydrophobic interactions with Ile32, Ile52, Leu61, Leu63, Ile94 and Val135 (Figure 6).

In summary, it is possible to say that for the Log P 1000, SPECS, and the *in-house* databases (JCM, PM, JRP and AM), the docked ligands are located at the entry of the hydrophobic cavity of TLR4/MD-2 in similar poses. The principle interactions are hydrophobic with the inner region of MD-2, and polar ones at the rim of MD-2. All the compounds show polar interactions with Arg90 and Lys122. The hydrophobic interactions, however, are more wide spread and not with the same set of amino acids for all the compounds. Arg90 is assumed to participate in interactions with sulforaphane, JTT705, isoxanthohumol, isoliquiritigenin, CAPE and JSH. Lys122 interacts with OSL07 and cinnamaldehyde.

Some of the identified side chains are also participating in the interaction with known ligands. Ile80 for example interacts with xanthohumol, JSH, OSL07, cinnamaldehyde and 6-shogaol. The side chains Phe121 and Ile124 interact with all known ligands. Hydrophobic interactions with the ligand are basically with aromatic cycles.

Regarding drug repurposing results (WORD subset from ZINC database), the analysis revealed 5 compounds outperforming the remaining ones (Annex Table 1): compounds **146**, **157**, **177**, **179**, and **208**. Surprisingly, compared to the previous analysis done only with the best Glide pose step II), compounds **157** and also compound **212** did not show good results in the last analysis. Indeed, having a more wide number of poses in Glide permitted to see, for these two compounds, that the first pose was not part of the most probable cluster, or any cluster at all, for

both conformations. Moreover, it has been shown that the most probable clusters for these two compounds were ranked in a low energy position, and with a medium total percentage of interaction against the main residues. Compound **208**, previously identified in the first analysis, having a good cluster position, was observed to have medium total percentage of interaction against the main residues. Compounds **157**, **208** and **212** were kept as a query for future structure similarity search.

In ascending order of better predicted binding, compounds **146**, **177** and **179** outperformed all the compounds. Compounds **146** and **177**, already revealed by the first cluster analysis, have shown having in each pose, interactions with almost all the main residues. Moreover, in about 50% of the poses, they were able to make 2 hydrogen bonds at the same time, and in about 70% of the poses, able to make 2 salt bridge interactions simultaneously. Regarding compound **179**, it was predicted having the highest affinity potential with all the main residues. It interacts with all the main residues with high affinity, making in 80% of the poses, up to 3 hydrogen bonding and a salt bridge in 50% of the poses. Compounds **146**, **177** and **179** were also kept as queries for future structure similarity search.

Regarding compound **146**, it is known as Diphenoxylate. It is a meperidine congener used as an antidiarrheal, usually in combination with atropine. At high doses, it acts like morphine. Its unesterified metabolite difenoxin has similar properties and is used similarly. It has little or no analgesic activity. According to DrugBank it is categorized as: analgesics, opioid, antiperistaltic agents, alimentary tract and metabolism, antidiarrheals, intestinal anti-inflammatory/anti-infective agents, and antipropulsives. Because TLR pathways can be related to inflammatory and microbial pathologies, it can be conceivable that Diphenoxylate could have a certain affinity for TLR4. It has also been shown that Diphenoxylate

can regulate NF-kβ,[56] a protein present downstream in the TLR pathway. Moreover, some studies have proven the binding between morphine and TLR4,[57] suggesting also a conceivable effect of Diphenoxylate to TLR4.

Compound **177** is known as Ono-Rs 411 or Pranlukast. It is a cysteinyl leukotriene receptor-1 antagonist. It antagonizes or reduces bronchospasm caused, principally in asthmatics, by an allergic reaction to accidentally or inadvertently encountered allergens. It is classified as: anti-asthmatic agents, respiratory system, drugs for obstructive airway diseases, leukotriene receptor antagonists, cytochrome P-450 CYP2C9 inhibitors, cytochrome P-450 CYP2C9 inducers, and CYP3A4 inhibitors. Besides, some studies have shown that Pranlukast can inhibit NF-kβ activation,[58] a protein present downstream in the TLR activation pathway. It has also been shown that it indirectly induces cytoplasmic membrane depolarization of Gram-negative bacteria, promoting *E. coli* outer membrane detachment,[59] which some are recognized by TLR4.

Compound **179** is known as Vemurafenib, a V600 mutant BRAF enzyme inhibitor for the treatment of late-stage melanoma.[60] Vemurafenib inhibits the active form of the kinase,[61] firmly anchoring itself in the ATP-binding site. By inhibiting only the active form of the kinase, it selectively inhibits the proliferation of cells with unregulated BRAF, normally those that cause cancer. It is classified as: antineoplastic agents, protein kinase inhibitors, antineoplastic and immunomodulating agents, cytochrome P-450 CYP1A2 inhibitors, cytochrome P-450 CYP1A2 inducers, CYP2D6 inducers, CYP2D6 inducers (strong), and CYP3A4 inhibitors. Up to date, it has been shown that TLR4 and its signaling pathway promote the migration of human melanoma cells,[62] but no studies showing an effect from Vemurafenib to TLR4 have been done yet.

All hit structures show a very common scaffold and binding pattern: two hydrophobic moieties separated by a polar linker. The larger hydrophobic part occupies the hydrophobic MD-2 cavity, while the smaller one is placed in the same hydrophobic side region where also one of the lipid A alkyl chains is located in the bound X-ray structure. Key interactions are those stablished with residues Arg90, capable of making salt bridges and hydrogen bonds, Phe121, able to make strong hydrophobic interactions, and situated closely to Phe126, and Tyr131, also able to make hydrogen bonds. These interactions were common for all the compounds, and conferred them a strong predicted binding energy. The polar linker seems to be interacting with two of the positively charged amino acids Arg90 and Lys122 at the entry region of the pocket which have already been described in the literature to interact with known active compounds. The literature reports a covalent interaction between Cys133 of MD-2 and some of the known actives. This observation could be reproduced with some of the known actives. The identified screening hits, however, represent an interesting scaffold for a new class of possible inhibitors for the TLR4/MD-2 complex.

STEP V. Biological testing

After molecular docking and VS would have allowed us to predict the binding and the affinity of putative TLR4 modulators, we subsequently tested them in HEK-Blue cells transfected with *h*TLR4, to check the ability of acting as TLR4 agonists or antagonists, and in J744 macrophage cells to check their ability to decrease TNF-α secretion. These studies were performed in collaboration with Belén de Andrés (CNM-ISCIII) and Manuel Fresno (UAM/CBM-CSIC).

The ability of molecules to interfere with LPS-triggered TLR4 activation in HEK-Blue *h*TLR4 cells model was investigated. This HEK293 cell line is stably transfected with human TLR4, MD-2, and CD14 genes. In addition, HEK-Blue™

cells express a secreted Alkaline Phosphatase (SEAP) produced upon activation of NF-kβ. LPS binding activates TLR4 and NF-kβ leading to SEAP secretion, which is detected by an alkaline phosphatase substrate in cell culture media (Figure 5).

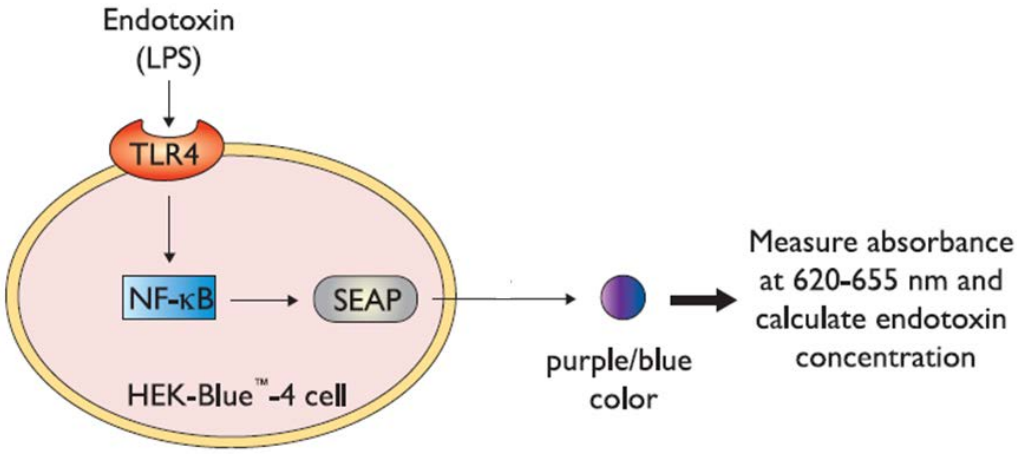


Figure 7: Cell-based colorimetric assay for the detection of biological active endotoxin (figure extracted from https://www.invivogen.com/hek-blue-lps-detection-kit).

In this assay, HEK293 cells transfected with human CD14 and TLR4/MD-2 were treated with increasing concentrations of synthetic molecules and then stimulated with LPS (LPS, 100 ng/ml). TLR4 activation is monitored as SEAP production. The results are normalized to activation by LPS alone and expressed as the mean of percentage ± SD of three independent experiments. The screened 27 compounds were tested and, from them, compounds B (**ID-5382**), F (**MS21**), H (**MS32**), I (**MS35**), X (**PM1090**) and Z (**PM1200**) inhibited TLR4 activation in a dose-dependent way (Figure 6).

As a negative control, compounds were tested in Null cell line (InvivoGen), transfected with the same plasmids as HEK-Blue but without TLR4, MD-2, and CD14 genes, and no effect was observed.

The toxicities of all compounds are being assaying by MTT assay and no inhibitory effects on cell viability have being observed in the concentration range used for biological characterization.

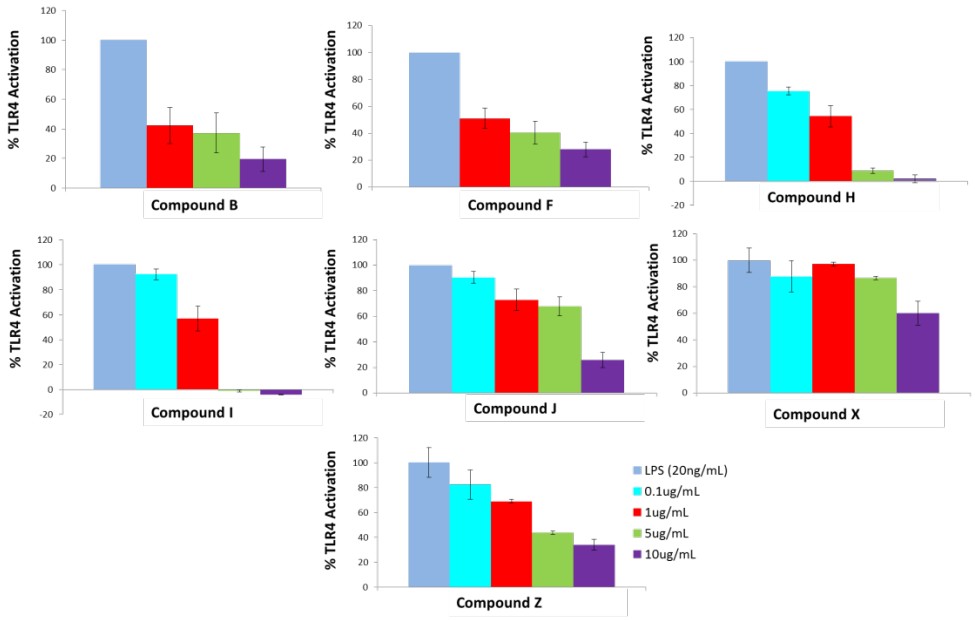
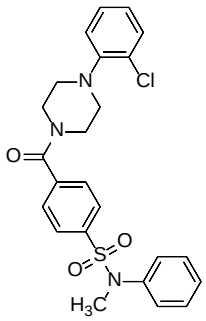


Figure 8: Results are expressed in % of TLR4 activation (positive control: LPS 20ng/mL).

We finally identified the following TLR4/MD-2 antagonist non LPS-like compounds:

Compound	Structure	LogP ChemSketch	LogP Molinspiration
ID-5382 (B)		2.77+/- 0.49	4.13

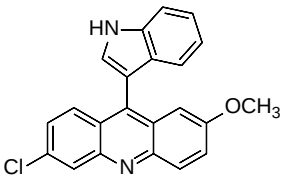
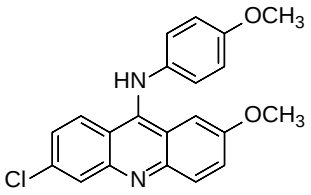
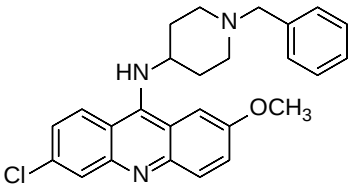
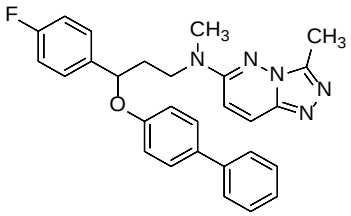
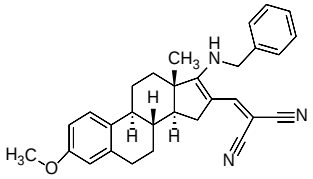
MS21 (F)		5.89+/- 0.40	6.331
MS32 (H)		4.20+/- 0.83	6.434
MS35 (I)		4.68+/-0.84	6.728
PM1090 (X)		5.70+/- 0.89	6.063
PM1200 (Z)		6.29+/- 0.45	6.116

Table 4: Non LPS-like compounds identified with TLR4/MD-2 antagonist activity.

We also tested the activity of these compounds on J744 macrophage cells, in order to detect the amount of TNF-α secreted by the cells in presence of the compounds. Moreover, in order to test the toxicity of selected compounds, MTT assay was performed on the same cell line (Figure 9). Compound H showed an strong TNF-α production inhibition at both concentrations and no cytotoxicity effect was observed. These results make compound H one of the most promising scaffolds. Compound F showed also an inhibitory activity at 1ug/mL, but high levels of cytotoxicity were observed at 5ug/mL. On the other hand, compound X inhibits the

production of TNF-α markedly at 5ug/mL, while no cytotoxic effect was observed at both concentrations. Compounds B, J, I and Z did not show any inhibitory effect and compound I showed a high cytotoxicity even at 1μg/mL.

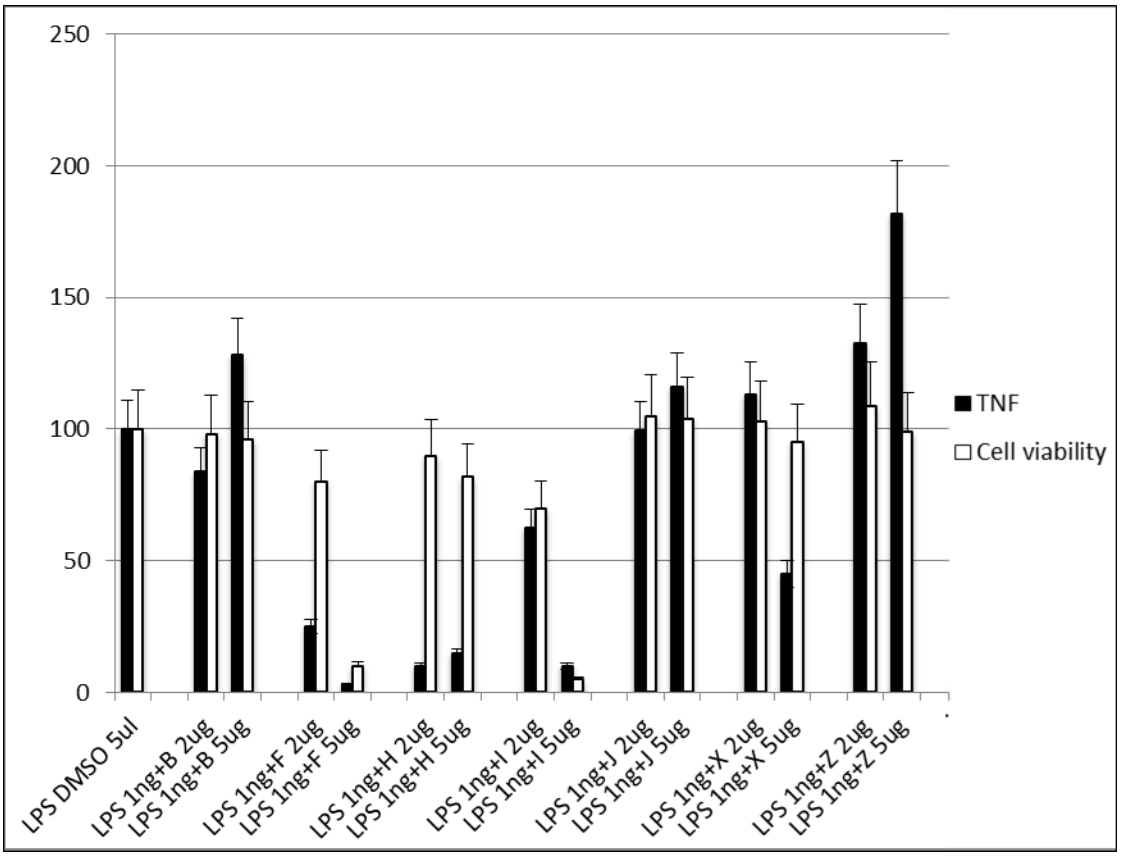


Figure 9: % of expresion of TNF-α secretion (black bars) and cell viability (white bars) in J744 cell line upon the treatment with screened compounds B, F, I, J, X and Z.

6.3 Conclusions

In this work, we have applied VS and computational repositioning strategies for the finding of novel TLR4 modulators. Our computational protocol has made use of different conformations of TLR4/MD-2, and included ligand-based and structure-based VS and deep ligand/receptor analysis. Our protocol has shown to be a robust approach for the identification of seven non LPS-like compounds with

TLR4/MD-2 antagonist activity. Compounds B (**ID-5382**), F (**MS21**), H (**MS32**), I (**MS35**), X (**PM1090**) and Z (**PM1200**) inhibited TLR4 activation in a dose-dependent way as putative of TLR4 modulators. In addition, compounds F and H showed an antagonist activity in J744 cell line, with no signs of cytotoxicity. The computationally identified hits represent interesting non LPS-like scaffolds for a new class of possible inhibitors for the TLR4/MD-2 complex.

6.4 Materials and Methods

-Computational Methods

Library Preparation

Importation. All the databases were saved as a SD File and imported in Maestro software (Schrodinger, version 10.4),[63] which is an all-purpose molecular modeling environment. During the importation process, the chirality and the atom type of each compound has been checked.

Ligand Preparation using LigPrep. LigPrep (Schrodinger, version 3.6)[64] is a program specialized in preparing all-atom 3D structure of drug like molecules, was used for many purposes: to refine the geometry of the ligands imported from the databases; to generate accurate, energy minimized 3D molecular structures; to expand tautomeric, ring conformation, and stereoisomers in order to produce broad chemical and structural diversity from each input structure and to predict protonation states. The 3D structures were minimized using OPLS 2005;[65] to generate ionization states, Epik[66] was used, in order to simulate the physiological pH. In many cases, the compounds contain water molecules or ions, these extra molecules were removed with Desalt option.

The generating tautomer options were also used in order to generate up to 8 tautomers per input structure. Regarding the setting stereoisomer options, the choice of retaining the specified chiralities to keep this information from the input file and fixed these chiralities for the entire calculation has been made. The number of stereoisomers generated was limited up to 32 per ligand. From a 2D structure, it is not immediately obvious which ring conformations give the lowest energy or are preferred for binding to an active site. Therefore it was decided to generate one low energy ring conformation per ligand with LigPrep. The final output was in Maestro format to keep the total information calculated for all the compounds. For the VS, the compounds were selected according to their molecular weight and their lipophilicity, between 300 Da and 700 Da, and between 4 and 6 respectively, using the property calculation tool from the Maestro software.

Protein Preparation

In the case of the agonist conformation of the TLR4/MD-2 monomer, 3D coordinates from TLR4/MD-2 heterodimer were obtained from the PDB (PDB-ID: 3FXI).[30] By contrast, in the case of the antagonist conformation, since the full crystallographic structure of the TLR4/MD-2 complex is not available, a model built by us was used. This model was built using the human MD-2 protein in antagonist conformation (PDB-ID: 2E59)[31] superimposed onto the MD-2 subunit of the agonist full complex (PDB-ID: 3FXI chain C) through PyMOL. Then, coordinates from the TLR4 chain of the 3FXI adjacent to the superimposed MD-2 (PDB-ID: 3FXI chain A) and the superimposed MD-2 in antagonist conformation were retained, forming the TLR4/MD-2 monomer in antagonist conformation. Finally, both agonist and antagonist the structures were subjected to 10000 cycles of steepest descent energy minimization under the Amber force field via Maestro

(see Chapter 3). Also PDB-ID: 2E56 were used to consider different antagonist conformation of MD-2.

Receptor Grid Preparation

Glide. For preparing the receptor grids for the two protein conformations, Glide software (Schrodinger, version 6.9) was used.[49] All the parameters from the software were kept at their default values. We only determined where the scoring grids will be positioned and their sizes. The coordinates of the box were set up to fully contain *E. coli* LPS. Glide software uses two "boxes" that can be parametrized to organize the calculation: the inner box, which can be monitored in the advanced panel, and where the ligand center is allowed to move within that box during the site point search; and the outer box, which is the box within all the ligand atoms must be contained. Its size is function of the inner box, and the inner box has to be included within the outer box. For the inner box, the center was set up at residue serine 120 and the lengths of the boxes for both protein conformations were the following ones: 33 Å in X, 40 Å in Y and 35 Å in Z. For the outer box, 10 Å has been chosen, that is to say 10 Å bigger than the inner box (43 Å in X, 50 Å in Y and 45 Å in Z).

AutoDock, FLAP and VINA. As the receptor grids were already set up with Glide, the same grids have been chosen for the softwares. Glide coordinates were kept for VINA, but were converted in AutoDock coordinates using scaling calculation tool. In the case of FLAP, the pockets of MD-2 were identified and defined by FLAP's pocket search algorithm.

Docking

Structure Based Virtual Screening (SBVS) with FLAP

The FLAP software explicitly distinguishes between the so-called *SBVS* method and docking.[67] While in FLAP docking is primarily used for pose prediction and a more precise quantification of binding energies, SBVS is a tool for large-scale virtual screenings. Even though docking is often used as a structure based virtual screening technique,[68] the term *SBVS* will hereafter refer only to FLAP's correspondent screening program.

The SBVS program first creates MIFs of the receptor's binding site. During screening, the MIFs of the ligand are compared with those of the binding site. Time consuming calculations describing each atom-atom interaction are not needed here. One downside of this method is that there is no energetic penalty for atom clashing with the target. In some scenarios however, this might even be an advantage, since it overcomes the rigidity of the target to some extent.

The SBVS in FLAP was performed on the 3D structure of the human co-receptor MD-2. The structure was obtained from the PDB (PDB-ID: 2E56)[31] and the MOE software was used to prepare the protein by removing water molecules, adding hydrogens and missing atoms and side chains.[40] The optimized structure was loaded into FLAP and the *Search for pockets* function was used to define the binding area. The results are then treated in analogy to the LBVS approach.

SBVS with Glide

The molecules were subjected to a grid-based ligand docking with energetics (Glide, Schrodinger, version 6.9)[49] using the Virtual Screening Workflow protocol. It is designed to run an entire sequence of jobs for screening large collections of compounds against one or more targets. However, as the compounds and the grids

had already been prepared, in this case, only the docking steps of the program have been used. The compound files and the receptor grid files were imported into the Virtual Screening Workflow program. Regarding the docking step parameters, Epik state penalties for docking were used, and the non-polar part of the ligand potential were soften by scaling the Van der Waals radii of ligand atoms with small partial charges. To do so, the scaling factor was 0.80, and the partial charge cutoff was 0.15. The full workflow includes three docking stages, each step differing from the preceding step in the amount of time taken to dock each molecule and the scoring system used to evaluate each pose. The first stage performs HTVS (High Throughput Virtual Screening) docking. The ligands that are retained are then passed to the next stage, which performs SP (Standard Precision) docking.

The survivors of this stage are passed onto the third stage, which performs XP (eXtra Precision) docking, a more powerful and discriminating procedure.

The Dock flexibility method was used for HTVS, SP and XP dockings allowing us to penalize non-planar conformation for amide bonds. A post-docking minimization was also performed, as well as constraints for the docking stages. One pose per compound state was generated and 100% of the best compounds that passed the HTVS, SP and XP docking have been kept. For HTVS and SP docking, all states have been retained, but only the best scoring state for the XP docking.

Ligand re-docking using Glide

The shortlisted molecules were submitted to a re-docking procedure using Glide. All the parameters were kept as mentioned in the docking paragraph using Glide, except for the docking poses which were set to 50 per molecule.

Molecular re-docking using FLAP

The FLAP software implements a fragmentation-based docking algorithm, called FLAPdock, which works as follows. MIFs are calculated for the target binding site, in a similar manner to the SBVS approach but with more points to describe the site in more detail (Reference manual for FLAP 2.0, © 2014 Molecular Discovery Ltd). A set of ligand conformations is generated using a stochastic search and a customized implementation of the MM3 force field[69] with a cutoff of 30 kcal mol^{-1} to remove high energy and duplicate conformations respectively. The ligands are then split into fragments with only 1-3 rotatable bonds. For each fragment conformation, GRID MIFs are calculated. The first fragment is docked into the binding site and the best scoring solutions, according to the global S-Score, are retained for the next iteration. In the next step, the next fragment, is attached to the first one and scored in the same way. The S-Score is a scoring function that includes terms from the GRID MIF similarities (Hydrogen-bonding and hydrophobic interactions as well as shape matching), Lennard-Jones and electrostatic interactions.

It was validated, amongst other targets, on those of the Astex and DUD datasets.[67b, 70] In each iteration, the best scoring solutions are kept and filtered by RMS clustering. Once the reconstruction of the ligand has finished, the final pose can be optionally optimized by minimization and the final score is recalculated. The benefit of FLAPdock towards the SBVS method lies in the more detailed chemical interactions that are considered for docking and the respect of steric clashes that are not regarded in the SBVS method. In order to obtain score reference values, a set of known MD-2 inhibitors was docked, followed by the docking of compounds from the Log P 1000 and the SPECS dataset.

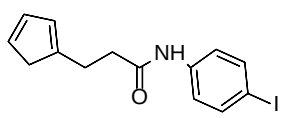
Biological characterization

HEK-Blue TLR4 assay. HEK-Blue TLR4 cells (InvivoGen, Toulouse, France) and parental cell line HEK-Blue Null 2 (InvivoGen) were HEK-Blue cells were used to test the agonist or antagonist effect of different compounds. This cell line expresses TLR4, MD-2 and CD14 and do not express any other TLR. The activation of TLR4 leads to the expression of SEAP, a protease that enzymatically hydrolyze a molecule present in the media. The amount of hydrolyzed molecule can be measured using colorimetric methods. These cells were cultured according to manufacturer's instructions. Briefly, cells were cultured in DMEM high glucose medium supplemented with 10% fetal bovine serum (FBS), 1% glutamine, 1% penicillin/ streptomycin, 1X Normocin (InvivoGen). Experiments were performed when 70-80% of confluence was reached. Cells were detached by the use of PBS, tapping the flask and the cell concentration was estimated using. Four different compound concentrations were used: 0.1, 1, 5 and 10 μg/mL. 20 μL of compound dilution were added in a 96-well plate, in triplicate (3 wells for each concentration), seeded in multiwall plate at a density of 2 x 10^4 cells/well in 200 μl. LPS was used as positive control (20ng/mL final concentration) and PBS 1x was used as negative control. Cells were detached using 4mL of PBS and 140000 cell/mL solution was prepared using Detections Media. 180 μL of this solution were added into each well (25000 cells/well). After a 30 min incubation, 20 μL of LPS solution were added in each well (final LPS concentration: 20 ng/mL) (LPS was diluted in PBS as well). Plates were incubated for 16 h in the dark at 37 °C, 95% of humidity and 5% of CO_2 and then, the plate reading was assessed using a spectrophotometer at 620 nm. The results were normalized with positive control (LPS alone) and expressed as the mean of percentage ± SD of at least three independent experiments.

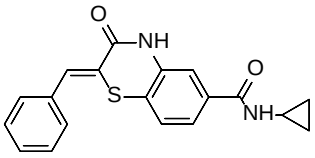
TNF-α detection. An adherent murine macrophage cell line J774.2 was grown in 75 cm2 cell culture flasks in Dulbecco's Modified Eagle Medium supplemented

with 5% Fetal Bovine Serum and 1% Penicillin/Streptomycin. Approximately $2x10^6$ cells were plated in individual wells of a 12-well plate. Cells were stimulated with LPS from Escherichia coli (Sigma Aldrich) at a final concentration of 2ng/mL. 5 μg/mL and 5 μg/mL of the respective compound were added with 1ng/mL of LPS. As positive control, LPS (1ng/mL) with DMSO was used. Two wells per compound and control were used. Plates were incubated for 24h at 37ºC and 5% of CO_2. Cell supernatants collected after 24 h stimulation assays were analyzed for TNF-α. Commercial enzyme-linked immunosorbent assay kits were used (Mouse TNF-α DuoSet ELISA, rndsystems). ELISA was performed in 96 well plates, and plates were read at 450 nm in a microplate reader.

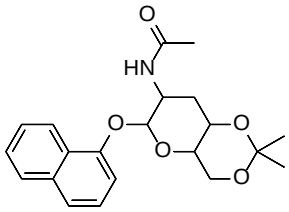
6.5 Annex V



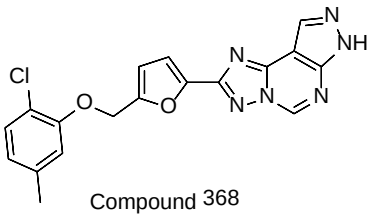
Compound 152



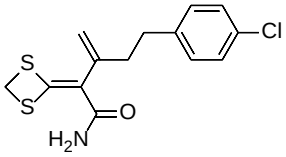
Compound 568



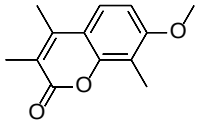
Compound 383



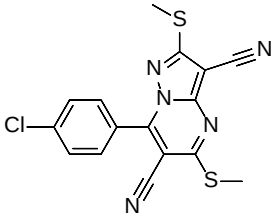
Compound 368



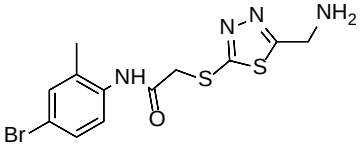
Compound 492



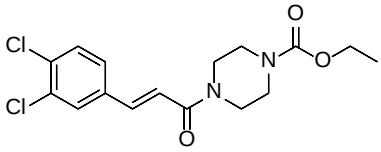
Compound 42



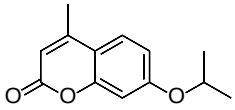
Compound 138



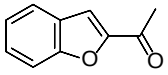
Compound 575



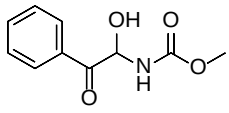
Compound 548



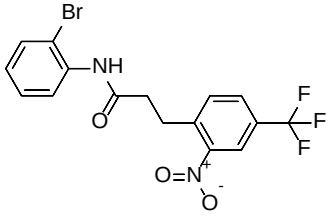
Compound 46



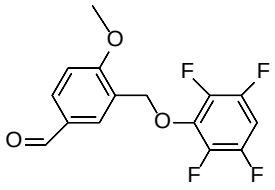
Compound 40



Compound 35

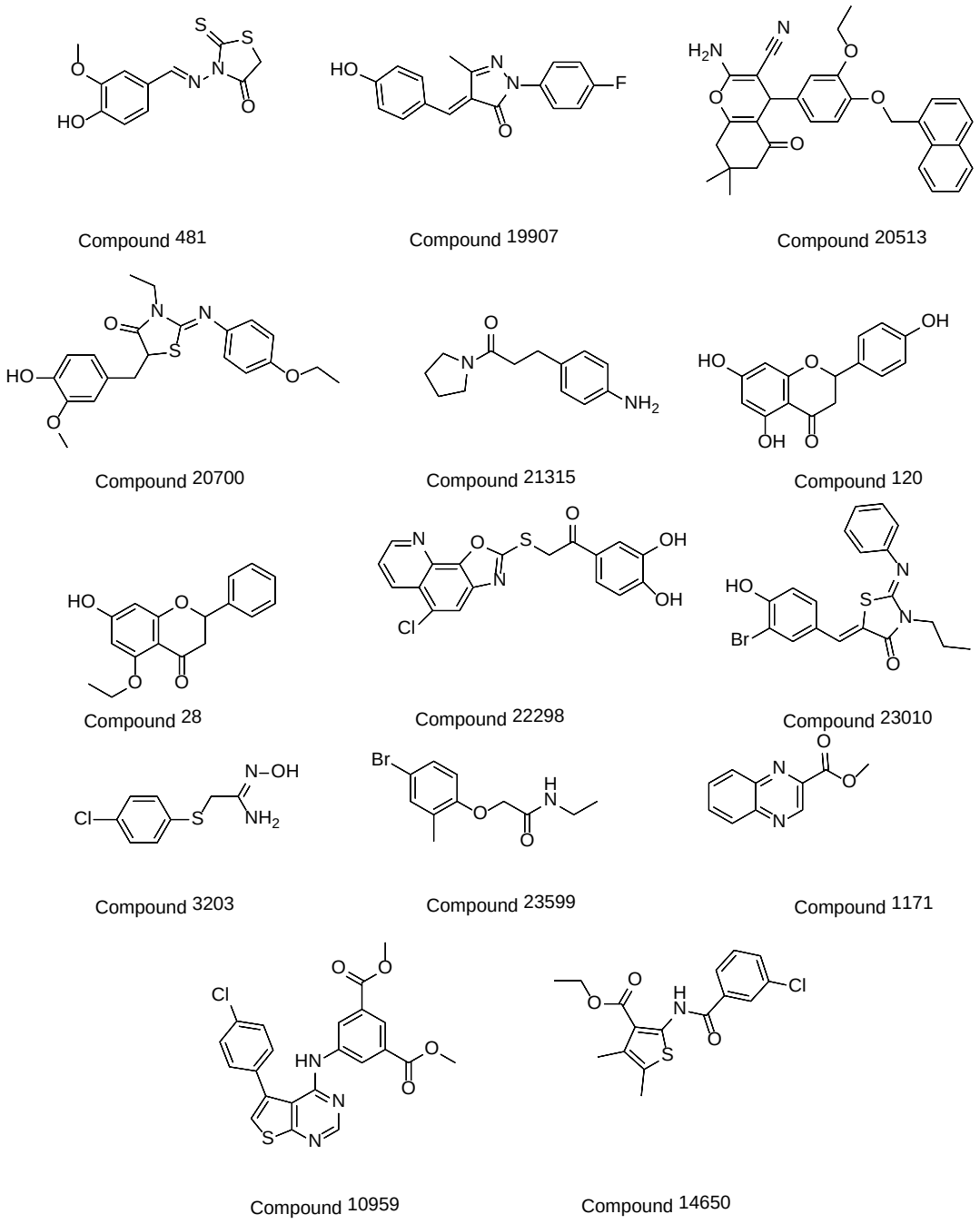


Compound 187

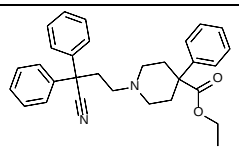
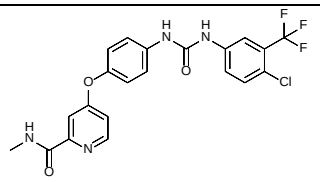
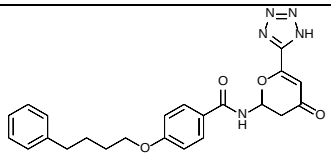
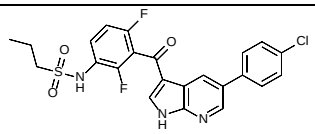
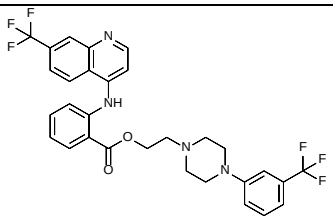
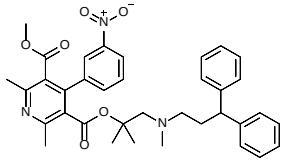


Compound 439

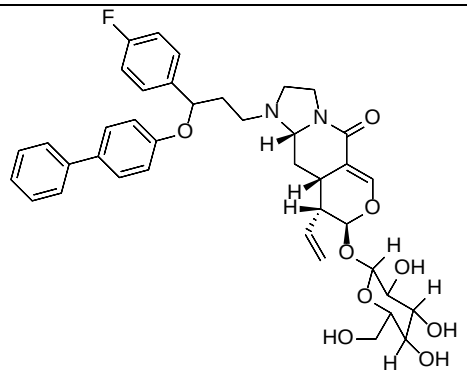
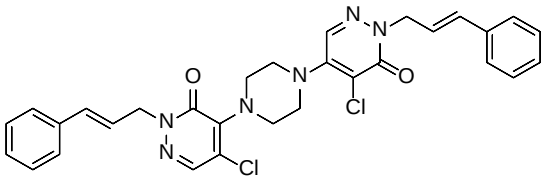
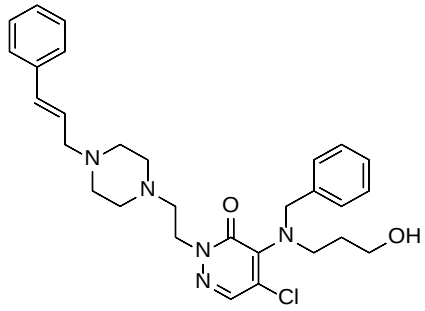
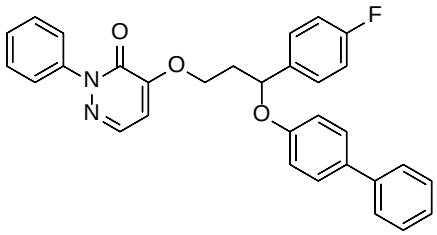
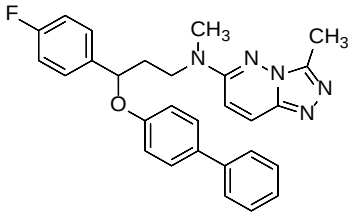
Annex Figure 1: Top scoring compounds obtained by LBVS on the Log P 1000 database.

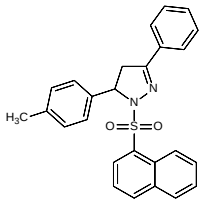
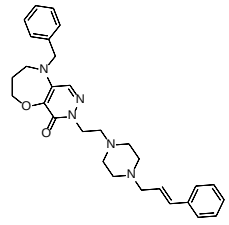
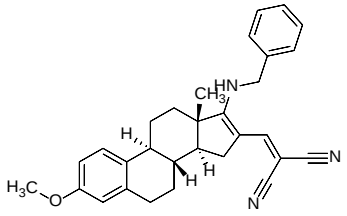


Annex Figure 2: Top scoring compounds obtained by LBVS on the SPECS database.

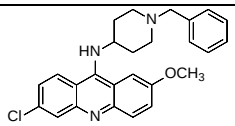
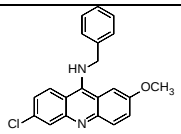
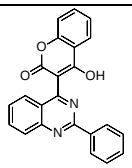
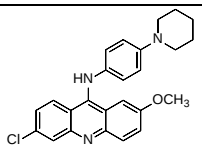
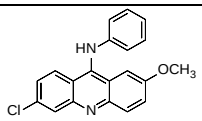
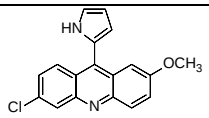
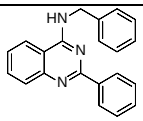
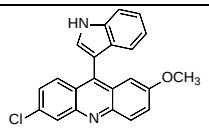
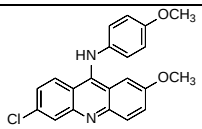
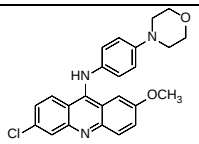
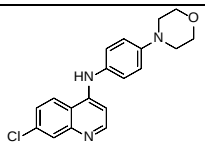
Compound number	ZINC ID	Usual/Commercial name	2D structure
146	3830716	Diphenoxylate	
157	1493878	Sorafenib	
177	15919406	Ono-Rs 411	
179	52509366	Zelboraf	
208	53073961	Antrafenine	
212	19685790	Lercanidipine	

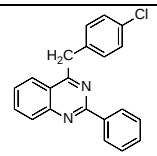
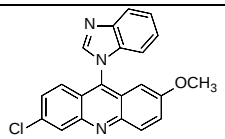
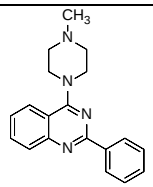
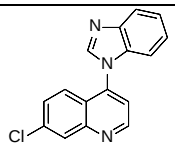
Annex Table 1: 2D Chemical structure of predicted TLR4 modulators identified by computational drug repurposing, and kept for future structure similarity search.

Number	Compound	Structure
1	PM1097_p_R/1097	
2	PM1811	
3	PM1779	
4	PM567S ó R	
5	PM1090	

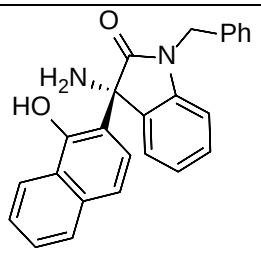
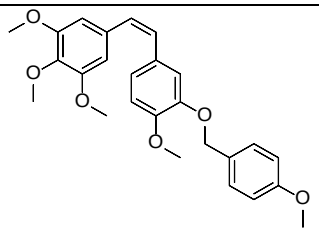
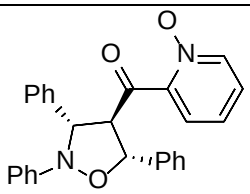
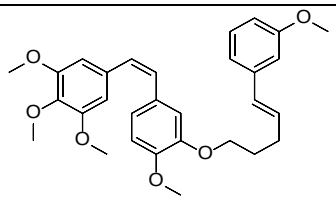
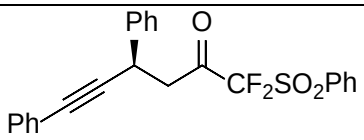
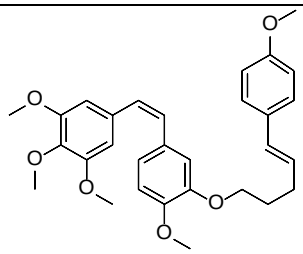
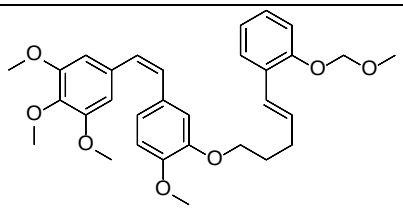
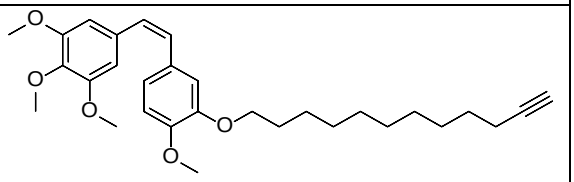
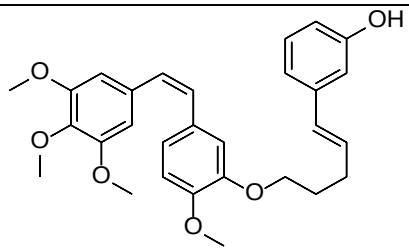
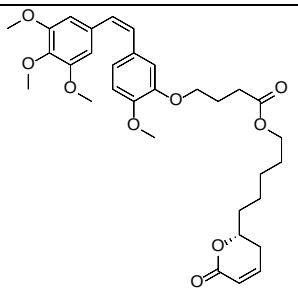
6	PM810	
7	PM1758	
8	PM1200	

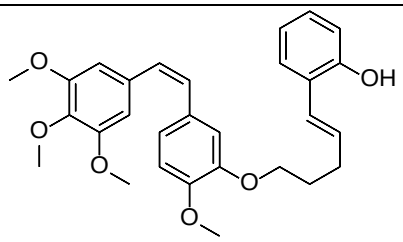
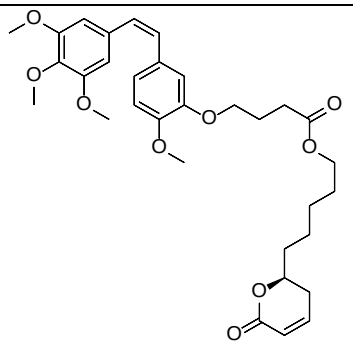
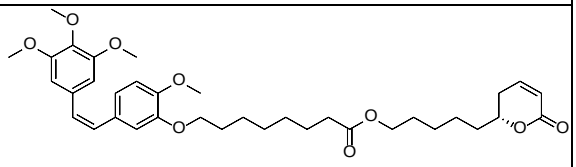
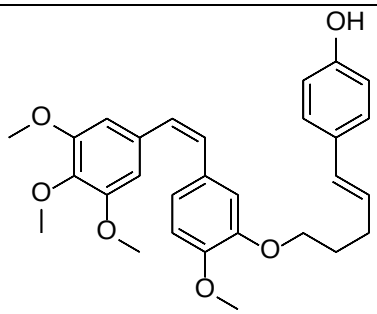
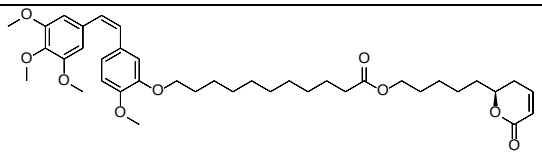
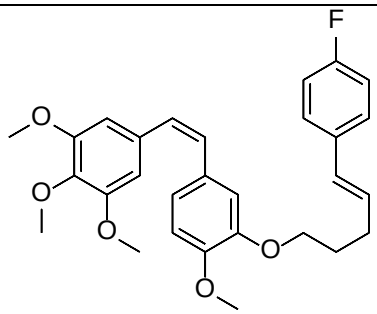
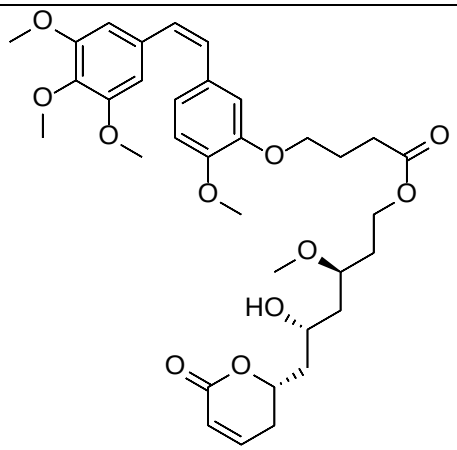
Annex Table 2: 2D Chemical structure from PM databases obtained from SBVS.

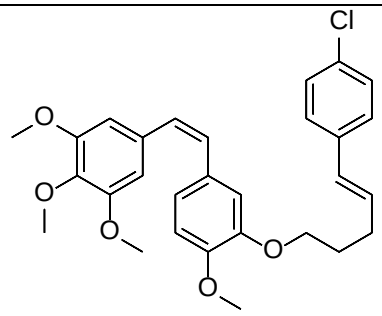
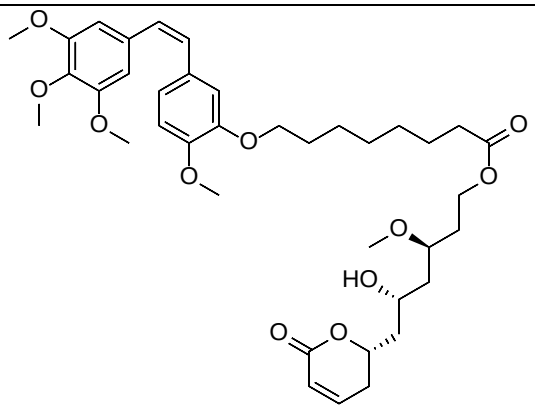
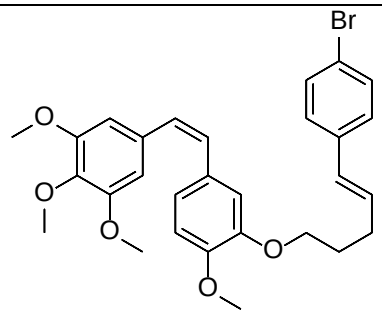
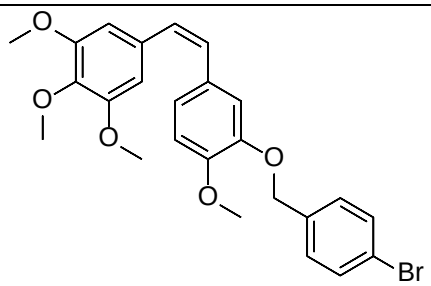
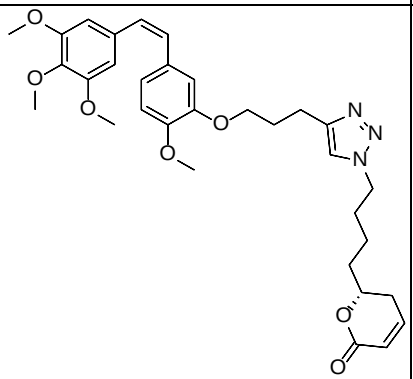
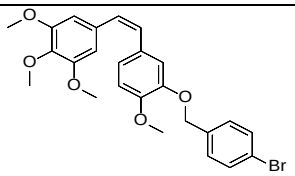
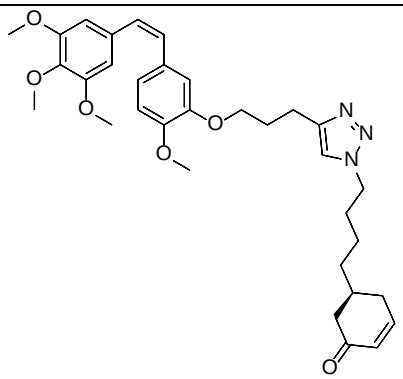
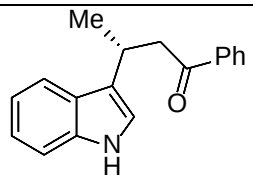
Number	Compound	Structure
1	MS_35/35p	
2	MS_29	
3	MS_40	
4	MS_34	
5	MS_31	
6	MS_22	
7	MS_45	
8	MS_21	
9	MS_32	
10	MS_26	
11	MS_14	

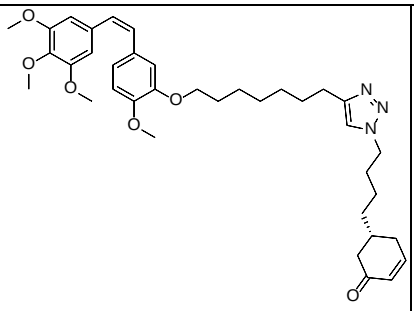
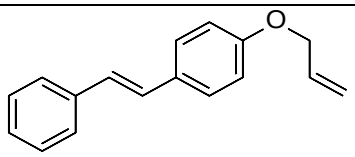
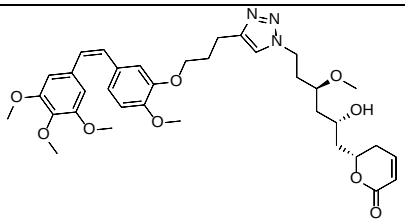
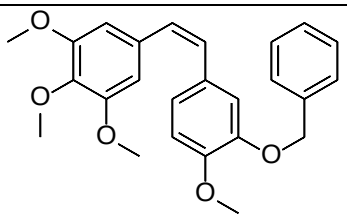
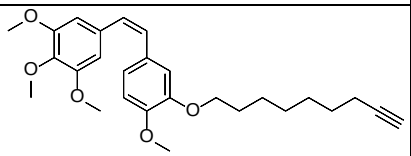
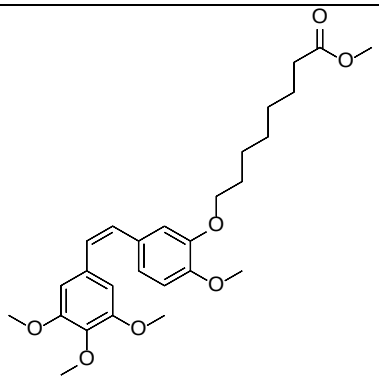
12	MS_49	
13	MS_37	
14	MS_46	
15	MS_20	

Annex Table 3: 2D Chemical structure from JCM databases obtained from SBVS.

Name	Structure	Name	Structure
JRP07		**AM15**	
JRP10		**AM18**	
JRP18		**AM19**	
AM20 momo		**AM54**	
AM21		**AM57**	

AM20		AM58	
		AM59	
AM22		AM62	
AM23		AM65	

AM24		**AM66**	
AM25		**AM72**	
AM40		**AM16**	
AM41		**JRP01**	

AM42		**AM08**	
AM48		**AM14**	
AM53		**AM71**	

Annex Table 4: 2D Chemical structure from JRP and AM databases obtained from SBVS.

Bibliography:

1. Mifsud, E. J.; Tan, A. C.; Jackson, D. C., TLR Agonists as Modulators of the Innate Immune Response and Their Potential as Agents Against Infectious Disease. *Front. Immunol.* **2014,** *5*, 79.
2. Joosten, L. A.; Abdollahi-Roodsaz, S.; Dinarello, C. A.; O'Neill, L.; Netea, M. G., Toll-like receptors and chronic inflammation in rheumatic diseases: new developments. *Nat. Rev. Rheumatol.* **2016,** *12* (6), 344-57.
3. Gooshe, M.; Aleyasin, A. R.; Abdolghaffari, A. H.; Rezaei, N., Toll like receptors: a new hope on the horizon to treat multiple sclerosis. *Expert Rev. Clin. Immunol.* **2014,** *10* (10), 1277-9.
4. (a) Nelson, M. H.; Diven, M. A.; Huff, L. W.; Paulos, C. M., Harnessing the Microbiome to Enhance Cancer Immunotherapy. *J. Immunol. Res.* **2015,** *2015,* 368736; (b) Rakoff-Nahoum, S.; Medzhitov, R., Toll-like receptors and cancer. *Nat. Rev. Cancer* **2009,** *9* (1), 57-63.
5. Gambuzza, M. E.; Sofo, V.; Salmeri, F. M.; Soraci, L.; Marino, S.; Bramanti, P., Toll-like receptors in Alzheimer's disease: a therapeutic perspective. *CNS Neurol. Disord. Drug Targets* **2014,** *13* (9), 1542-58.
6. 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.; Jauregui, L.; Krell, K.; Pachl, J.; Takahashi, T.; Peckelsen, C.; Cordasco, E.; Chang, C. S.; Oeyen, S.; Aikawa, N.; Maruyama, T.; Schein, R.; Kalil, A. C.; Van Nuffelen, M.; Lynn, M.; Rossignol, D. P.; Gogate, J.; Roberts, M. B.; Wheeler, J. L.; Vincent, J. L., Effect of eritoran, an antagonist of MD2-TLR4, on mortality in patients with severe sepsis: the ACCESS randomized trial. *Jama* **2013,** *309* (11), 1154-62.
7. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002,** *45* (12), 2615-23.
8. Chan, M.; Hayashi, T.; Mathewson, R. D.; Nour, A.; Hayashi, Y.; Yao, S.; Tawatao, R. I.; Crain, B.; Tsigelny, I. F.; Kouznetsova, V. L.; Messer, K.; Pu, M.; Corr, M.; Carson, D. A.; Cottam, H. B., Identification of substituted pyrimido[5,4-b]indoles as selective Toll-like receptor 4 ligands. *J. Med. Chem.* **2013,** *56* (11), 4206-23.
9. Neve, J. E.; Wijesekera, H. P.; Duffy, S.; Jenkins, I. D.; Ripper, J. A.; Teague, S. J.; Campitelli, M.; Garavelas, A.; Nikolakopoulos, G.; Le, P. V.; de, A. L. P.; Pham, N. B.; Shelton, P.; Fraser, N.; Carroll, A. R.; Avery, V. M.; McCrae, C.; Williams, N.; Quinn, R. J., Euodenine A: a small-molecule agonist of human TLR4. *J. Med. Chem.* **2014,** *57* (4), 1252-75.
10. Shanmugam, A.; Rajoria, S.; George, A. L.; Mittelman, A.; Suriano, R.; Tiwari, R. K., Synthetic Toll like receptor-4 (TLR-4) agonist peptides as a novel class of adjuvants. *PloS one* **2012,** *7* (2), e30839.
11. Park, S. J.; Kang, S. H.; Kang, Y. K.; Eom, Y. B.; Koh, K. O.; Kim, D. Y.; Youn, H. S., Inhibition of homodimerization of toll-like receptor 4 by 4-oxo-4-(2-oxo-oxazolidin-3-yl)-but-2-enoic acid ethyl ester. *Int. Immunopharmacol.* **2011,** *11* (1), 19-22.
12. Jin, G. H.; Li, H.; An, S.; Ryu, J. H.; Jeon, R., Design, synthesis and activity of benzothiazole-based inhibitors of NO production in LPS-activated macrophages. *Bioorg. Med. Chem. Lett.* **2010,** *20* (21), 6199-202.
13. Kawamoto, T.; Ii, M.; Kitazaki, T.; Iizawa, Y.; Kimura, H., TAK-242 selectively suppresses Toll-like receptor 4-signaling mediated by the intracellular domain. *Eur. J. Pharmacol.* **2008,** *584* (1), 40-48.

14. Chavez, S. A.; Martinko, A. J.; Lau, C.; Pham, M. N.; Cheng, K.; Bevan, D. E.; Mollnes, T. E.; Yin, H., Development of beta-amino alcohol derivatives that inhibit Toll-like receptor 4 mediated inflammatory response as potential antiseptics. *J. Med. Chem.* **2011,** *54* (13), 4659-69.
15. Lyne, P. D., Structure-based virtual screening: an overview. *Drug Discov. Today* **2002,** *7* (20), 1047-1055.
16. (a) Schneider, G., Virtual screening: an endless staircase? *Nat. Rev. Drug Discov.* **2010,** *9* (4), 273-6; (b) Haga, J. H.; Ichikawa, K.; Date, S., Virtual Screening Techniques and Current Computational Infrastructures. *Curr. Pharm. Des.* **2016,** *22* (23), 3576-84; (c) Zhu, T.; Cao, S.; Su, P. C.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M. E.; Hevener, K. E., Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *J. Med. Chem.* **2013,** *56* (17), 6560-72; (d) Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z., Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr. Top. Med. Chem.* **2014,** *14* (16), 1923-38; (e) Yan, X.; Liao, C.; Liu, Z.; Hagler, A. T.; Gu, Q.; Xu, J., Chemical Structure Similarity Search for Ligand-based Virtual Screening: Methods and Computational Resources. *Curr. Drug Targets* **2016,** *17* (14), 1580-1585; (f) Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W., State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. *Chem. Soc. Rev.* **2015,** *44* (22), 8238-59.
17. Perez-Regidor, L.; Zarioh, M.; Ortega, L.; Martin-Santamaria, S., Virtual Screening Approaches towards the Discovery of Toll-Like Receptor Modulators. *Int. J. Mol. Sci.* **2016,** *17* (9).
18. Joce, C.; Stahl, J. A.; Shridhar, M.; Hutchinson, M. R.; Watkins, L. R.; Fedichev, P. O.; Yin, H., Application of a novel in silico high-throughput screen to identify selective inhibitors for protein-protein interactions. *Bioorg. Med. Chem. Lett.* **2010,** *20* (18), 5411-3.
19. http://www.enamine.net, **(accessed on 6th June 2017)**.
20. Svajger, U.; Brus, B.; Turk, S.; Sova, M.; Hodnik, V.; Anderluh, G.; Gobec, S., Novel toll-like receptor 4 (TLR4) antagonists identified by structure- and ligand-based virtual screening. *Eur. J. Med. Chem.* **2013,** *70*, 393-9.
21. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G., ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012,** *52* (7), 1757-68.
22. Chong, C. R.; Sullivan, D. J., Jr., New uses for old drugs. *Nature* **2007,** *448* (7154), 645-6.
23. Shibayama, S.; Tanikawa, K.; Fujimoto, R.; Kimura, H., Effect of mergers and acquisitions on drug discovery: perspective from a case study of a Japanese pharmaceutical company. *Drug Discov. Today* **2008,** *13* (1-2), 86-93.
24. Langedijk, J.; Mantel-Teeuwisse, A. K.; Slijkerman, D. S.; Schutjens, M. H., Drug repositioning and repurposing: terminology and definitions in literature. *Drug Discov. Today* **2015,** *20* (8), 1027-34.
25. (a) Allarakhia, M., Open-source approaches for the repurposing of existing or failed candidate drugs: learning from and applying the lessons across diseases. *Drug Des. Dev. Ther.* **2013,** *7*, 753-766; (b) Allison, M., NCATS launches drug repurposing program. *Nat. Biotechnol.* **2012,** *30* (7), 571-572; (c) Marusina, K.; Welsch, D. J.; Rose, L.; Brock, D.; Bahr, N., The CTSA Pharmaceutical Assets Portal–a public–private partnership model for drug repositioning. *Drug Discov. Today* **2012,** *8* (3), 77-83; (d) Murteira, S.; Ghezaiel, Z.; Karray, S.; Lamure, M., Drug reformulations and repositioning in pharmaceutical industry and its impact on market access: reassessment of nomenclature. *J. Mark Access Health Policy* **2013,** *1*; (e) Novac, N., Challenges and opportunities of drug repositioning. *Trends Pharmacol. Sci.* **2013,** *34* (5), 267-272; (f) Ashburn, T. T.; Thor, K. B., Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev.* **2004,** *3* (8), 673-683.
26. (a) Ekins, S.; Williams, A. J.; Krasowski, M. D.; Freundlich, J. S., In silico repositioning of approved drugs for rare and neglected diseases. *Drug Discov. Today* **2011,** *16* (7), 298-310; (b)

Muthyala, R., Orphan/rare drug discovery through drug repositioning. *Drug Discov.Today* **2012,** *8* (3), 71-76; (c) Sardana, D.; Zhu, C.; Zhang, M.; Gudivada, R. C.; Yang, L.; Jegga, A. G., Drug repositioning for orphan diseases. *Brief. Bioinform.* **2011,** *12* (4), 346-356; (d) Aubé, J., ACS Med. Chem. Lett. *ACS medicinal chemistry letters* **2012,** *3* (6), 442-444.

27. Fagan, S. C., Drug Repurposing for Drug Development in Stroke. *Pharmacother.* **2010,** *30* (7P2), 51S-54S.

28. (a) Singhal , S.; Mehta , J.; Desikan , R.; Ayers , D.; Roberson , P.; Eddlemon , P.; Munshi , N.; Anaissie , E.; Wilson , C.; Dhodapkar , M.; Zeldis , J.; Siegel , D.; Crowley , J.; Barlogie , B., Antitumor Activity of Thalidomide in Refractory Multiple Myeloma. *N. Engl. J. Med.* **1999,** *341* (21), 1565-1571; (b) Yang, T.-J.; Yang, T.-S.; Liang, H.-M., Thalidomide and congenital abnormalities. *The Lancet* **1963,** *281* (7280), 552-553.

29. 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), 2672-83.

30. 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), 1191-1195.

31. 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.

32. Staderini, M.; Cabezas, N.; Bolognesi, M. L.; Menendez, J. C.; Carlos, J., *Solvent- and chromatography-free amination of π-deficient nitrogen heterocycles under microwave irradiation. A fast, efficient and green route to 9-aminoacridines, 4-aminoquinolines and 4-aminoquinazolines and its application to the synthesis of the drugs amsacrine and bistacrine*. Elsevier: Kidlington, ROYAUME-UNI, 2013; Vol. 69.

33. (a) Sanz-Marco, A.; Garcia-Ortiz, A.; Blay, G.; Pedro, J. R., Catalytic asymmetric conjugate addition of terminal alkynes to beta-trifluoromethyl alpha,beta-enones. *Chem. Commun.* **2014,** *50* (18), 2275-8; (b) Blay, G.; Hernandez-Olmos, V.; Pedro, J. R., The construction of quaternary stereocenters by the Henry reaction: circumventing the usual reactivity of substituted glyoxals. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2011,** *17* (13), 3768-73; (c) Barroso, S.; Blay, G.; Al-Midfa, L.; Munoz, M. C.; Pedro, J. R., Copper(II)-bis(oxazoline) catalyzed asymmetric Diels-Alder reaction with alpha'-arylsulfonyl enones as dienophiles. *J. Org. Chem.* **2008,** *73* (16), 6389-92; (d) Holmquist, M.; Blay, G.; Munoz, M. C.; Pedro, J. R., Enantioselective addition of nitromethane to 2-acylpyridine N-oxides. Expanding the generation of quaternary stereocenters with the Henry reaction. *Org. Lett.* **2014,** *16* (4), 1204-7; (e) Blay, G.; Hernandez-Olmos, V.; Pedro, J. R., Enantioselective Henry addition of methyl 4-nitrobutyrate to aldehydes. Chiral building blocks for 2-pyrrolidinones and other derivatives. *Org. Lett.* **2010,** *12* (13), 3058-61; (f) Blay, G.; Incerti, C.; Muñoz, M. C.; Pedro, J. R., Enantioselective LaIII-pyBOX-Catalyzed Nitro-Michael Addition to (E)-2-Azachalcones. *Eur. J. Org. Chem.* **2013,** *2013* (9), 1696-1705; (g) Blay, G.; Fernandez, I.; Munoz, M. C.; Pedro, J. R.; Recuenco, A.; Vila, C., Enantioselective synthesis of tertiary alcohols through a zirconium-catalyzed Friedel-Crafts alkylation of pyrroles with alpha-ketoesters. *J. Org. Chem.* **2011,** *76* (15), 6286-94; (h) Blay, G.; Fernandez, I.; Monleon, A.; Pedro, J. R.; Vila, C., Enantioselective zirconium-catalyzed Friedel-Crafts alkylation of pyrrole with trifluoromethyl ketones. *Org. Lett.* **2009,** *11* (2), 441-4; (i) Barroso, S.; Blay, G.; Muñoz, M. C.; Pedro, J. R., Highly Enantio- and Diastereoselective Inverse Electron Demand Hetero-Diels–Alder Reaction using 2-Alkenoylpyridine N-Oxides as Oxo-Heterodienes. *Adv. Synth. Catal.* **2009,** *351* (1-2), 107-111; (j) Sanz-Marco, A.; Blay, G.; Munoz, M. C.; Pedro, J. R., Highly enantioselective copper(I)-catalyzed conjugate addition of 1,3-diynes to alpha,beta-unsaturated trifluoromethyl ketones. *Chem. Commun.* **2015,** *51* (43), 8958-61; (k) Blay, G.; Fernandez, I.; Pedro, J. R.; Vila, C., Highly enantioselective Friedel-Crafts alkylations of indoles with simple enones catalyzed by zirconium(IV)-BINOL complexes. *Org. Lett.*

2007, *9* (13), 2601-4; (l) Barroso, S.; Blay, G.; Munoz, M. C.; Pedro, J. R., Highly enantioselective nitrone cycloadditions with 2-alkenoyl pyridine N-oxides catalyzed by Cu(II)-BOX complexes. *Org. Lett.* **2011,** *13* (3), 402-5; (m) Blay, G.; Fernandez, I.; Munoz, M. C.; Pedro, J. R.; Vila, C., Synthesis of functionalized indoles with a trifluoromethyl-substituted stereogenic tertiary carbon atom through an enantioselective Friedel-Crafts alkylation with beta-trifluoromethyl-alpha,beta-enones. *Chemistry (Weinheim an der Bergstrasse, Germany)* **2010,** *16* (30), 9117-22; (n) Blay, G.; Fernández, I.; Monleón, A.; Muñoz, M. C.; Pedro, J. R.; Vila, C., Synthesis of Functionalized Indoles with an α-Stereogenic Ketone Moiety Through an Enantioselective Friedel–Crafts Alkylation with (E)-1,4-Diaryl-2-butene-1,4-diones. *Adv. Synth. Catal.* **2009,** *351* (14-15), 2433-2440.

34. (a) Vilanova, C.; Diaz-Oltra, S.; Murga, J.; Falomir, E.; Carda, M.; Redondo-Horcajo, M.; Diaz, J. F.; Barasoain, I.; Marco, J. A., Design and synthesis of pironetin analogue/colchicine hybrids and study of their cytotoxic activity and mechanisms of interaction with tubulin. *J. Med. Chem.* **2014,** *57* (24), 10391-403; (b) Vilanova, C.; Torijano-Gutierrez, S.; Diaz-Oltra, S.; Murga, J.; Falomir, E.; Carda, M.; Alberto Marco, J., Design and synthesis of pironetin analogue/combretastatin A-4 hybrids containing a 1,2,3-triazole ring and evaluation of their cytotoxic activity. *Eur. J. Med. Chem.* **2014,** *87*, 125-30; (c) Marco, J. A.; Garcia-Pla, J.; Carda, M.; Murga, J.; Falomir, E.; Trigili, C.; Notararigo, S.; Diaz, J. F.; Barasoain, I., Design and synthesis of pironetin analogues with simplified structure and study of their interactions with microtubules. *Eur. J. Med. Chem.* **2011,** *46* (5), 1630-7; (d) Marti-Centelles, R.; Cejudo-Marin, R.; Falomir, E.; Murga, J.; Carda, M.; Marco, J. A., Inhibition of VEGF expression in cancer cells and endothelial cell differentiation by synthetic stilbene derivatives. *Bioorg. Med. Chem.* **2013,** *21* (11), 3010-5; (e) Panos, J.; Diaz-Oltra, S.; Sanchez-Peris, M.; Garcia-Pla, J.; Murga, J.; Falomir, E.; Carda, M.; Redondo-Horcajo, M.; Diaz, J. F.; Barasoain, I.; Marco, J. A., Synthesis and biological evaluation of truncated alpha-tubulin-binding pironetin analogues lacking alkyl pendants in the side chain or the dihydropyrone ring. *Org. Biomol. Chem.* **2013,** *11* (35), 5809-26; (f) Torijano-Gutierrez, S.; Diaz-Oltra, S.; Falomir, E.; Murga, J.; Carda, M.; Marco, J. A., Synthesis of combretastatin A-4 O-alkyl derivatives and evaluation of their cytotoxic, antiangiogenic and antitelomerase activity. *Bioorg. Med. Chem.* **2013,** *21* (23), 7267-74.

35. Sterling, T.; Irwin, J. J., ZINC 15-Ligand Discovery for Everyone. **2015**.

36. Martel, S.; Gillerat, F.; Carosati, E.; Maiarelli, D.; Tetko, I. V.; Mannhold, R.; Carrupt, P. A., Large, chemically diverse dataset of logP measurements for benchmarking studies. *Eur. J. Pharm. Sci.* **2013,** *48* (1-2), 21-9.

37. Irwin, J. J.; Shoichet, B. K., ZINC--a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005,** *45* (1), 177-82.

38. Cruciani, G.; Pastor, M.; Guba, W., VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000,** *11, Supplement 2*, S29-S39.

39. (a) Cummins, D. J.; Andrews, C. W.; Bentley, J. A.; Cory, M., Molecular diversity in chemical databases: comparison of medicinal chemistry knowledge bases and databases of commercially available compounds. *J. Chem. Inf. Comput. Sci.* **1996,** *36* (4), 750-63; (b) http://www.specs.net/, **(accessed on 6th June 2017)**.

40. *MOE 2011.10* **Chemical Computing Group Inc.: Montreal, 2011**.

41. Wild, D. J.; Blankley, C. J., Comparison of 2D fingerprint types and hierarchy level selection methods for structural grouping using Ward's clustering. *J. Chem. Inf. Comput. Sci.* **2000,** *40* (1), 155-62.

42. (a) Tanimoto, T. T., *An Elementary Mathematical Theory of Classification and Prediction*. International Business Machines Corporation: 1958; (b) Willett, P.; Barnard, J. M.; Downs, G. M., Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998,** *38* (6), 983-996; (c) Lipkus, A. H., A proof of the triangle inequality for the Tanimoto distance. *J. Math. Chem.* **1999,** *26* (1), 263-265.

43. Scior, T.; Lozano-Aponte, J.; Figueroa-Vazquez, V.; Yunes-Rojas, J. A.; Zahringer, 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*, e201305003.
44. Park, S. H.; Kim, N. D.; Jung, J. K.; Lee, C. K.; Han, S. B.; Kim, Y., Myeloid differentiation 2 as a therapeutic target of inflammatory disorders. *Pharmacol. Ther.* **2012,** *133* (3), 291-8.
45. (a) Resman, N.; Gradisar, H.; Vasl, J.; Keber, M. M.; Pristovsek, P.; Jerala, R., Taxanes inhibit human TLR4 signaling by binding to MD-2. *FEBS Lett.* **2008,** *582* (28), 3929-34; (b) 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), 2251-4.
46. Peri, F.; Piazza, M., Therapeutic targeting of innate immunity with Toll-like receptor 4 (TLR4) antagonists. *Biotechnol. Adv.* **2012,** *30* (1), 251-60.
47. (a) Gradisar, H.; Keber, M. M.; Pristovsek, P.; Jerala, R., MD-2 as the target of curcumin in the inhibition of response to LPS. *J. Leukoc. Biol.* **2007,** *82* (4), 968-74; (b) Roh, E.; Lee, H. S.; Kwak, J. A.; Hong, J. T.; Nam, S. Y.; Jung, S. H.; Lee, J. Y.; Kim, N. D.; Han, S. B.; Kim, Y., MD-2 as the target of nonlipid chalcone in the inhibition of endotoxin LPS-induced TLR4 activity. *J. Infect. Dis.* **2011,** *203* (7), 1012-20; (c) Peluso, M. R.; Miranda, C. L.; Hobbs, D. J.; Proteau, R. R.; Stevens, J. F., Xanthohumol and related prenylated flavonoids inhibit inflammatory cytokine production in LPS-activated THP-1 monocytes: structure-activity relationships and in silico binding to myeloid differentiation protein-2 (MD-2). *Planta Med.* **2010,** *76* (14), 1536-43.
48. (a) Yadav, M. K.; Pandey, S. K.; Swati, D., Drug target prioritization in Plasmodium falciparum through metabolic network analysis, and inhibitor designing using virtual screening and docking approach. *J. Bioinform. Comput. Biol.* **2013,** *11* (4), 1350003; (b) Scior, T.; Bender, A.; Tresadern, G.; Medina-Franco, J. L.; Martinez-Mayorga, K.; Langer, T.; Cuanalo-Contreras, K.; Agrafiotis, D. K., Recognizing pitfalls in virtual screening: a critical review. *J. Chem. Inf. Model.* **2012,** *52* (4), 867-81; (c) Collins, J. C.; Armstrong, A.; Chapman, K. L.; Cordingley, H. C.; Jaxa-Chamiec, A. A.; Judd, K. E.; Mann, D. J.; Scott, K. A.; Tralau-Stewart, C. J.; Low, C. M. R., Prospective use of molecular field points in ligand-based virtual screening: efficient identification of new reversible Cdc25 inhibitors. *MedChemComm* **2013,** *4* (8), 1148-1155; (d) Liu, C.; He, G.; Jiang, Q.; Han, B.; Peng, C., Novel Hybrid Virtual Screening Protocol Based on Molecular Docking and Structure-Based Pharmacophore for Discovery of Methionyl-tRNA Synthetase Inhibitors as Antibacterial Agents. *Int. J. Mol. Sci.* **2013,** *14* (7), 14225-14239.
49. (a) Small-Molecule Drug Discovery Suite 2015-4: Glide, v., Schrödinger, LLC, New York, NY, 2015; (b) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006,** *49* (21), 6177-96; (c) 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.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004,** *47* (7), 1739-49.
50. 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.
51. 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.

52. (a) Baxter, J., Local optima avoidance in depot location. *J. Oper. Res. Soc.* **1981**, 815-819; (b) Blum, C.; Roli, A., Hybrid metaheuristics: an introduction. In *Hybrid Metaheuristics*, Springer: 2008; pp 1-30.
53. Oberhauser, N. Lipophilicity in Computer-Aided Drug Design: New Tools and Applications. University of Geneve (Switzerland), 2015.
54. Sirci, F.; Istyastono, E. P.; Vischer, H. F.; Kooistra, A. J.; Nijmeijer, S.; Kuijer, M.; Wijtmans, M.; Mannhold, R.; Leurs, R.; de Esch, I. J.; de Graaf, C., Virtual fragment screening: discovery of histamine H3 receptor ligands using ligand-based and protein-based molecular fingerprints. *J. Chem. Inf. Model.* **2012,** *52* (12), 3308-24.
55. 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), 600-5.
56. Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M., DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **2008,** *36* (Database issue), D901-6.
57. (a) Lewis, S. S.; Hutchinson, M. R.; Rezvani, N.; Loram, L. C.; Zhang, Y.; Maier, S. F.; Rice, K. C.; Watkins, L. R., Evidence that intrathecal morphine-3-glucuronide may cause pain enhancement via toll-like receptor 4/MD-2 and interleukin-1β. *Neuroscience* **2010,** *165* (2), 569-583; (b) Hutchinson, M. R.; Zhang, Y.; Shridhar, M.; Evans, J. H.; Buchanan, M. M.; Zhao, T. X.; Slivka, P. F.; Coats, B. D.; Rezvani, N.; Wieseler, J.; Hughes, T. S.; Landgraf, K. E.; Chan, S.; Fong, S.; Phipps, S.; Falke, J. J.; Leinwand, L. A.; Maier, S. F.; Yin, H.; Rice, K. C.; Watkins, L. R., Evidence that opioids may have toll-like receptor 4 and MD-2 effects. *Brain Behav. Immun.* **2010,** *24* (1), 83-95; (c) Hutchinson, M. R.; Northcutt, A. L.; Hiranita, T.; Wang, X.; Lewis, S.; Thomas, J.; van Steeg, K.; Kopajtic, T. A.; Loram, L.; Sfregola, C.; Galer, E.; Miles, N. E.; Bland, S. T.; Amat, J.; Rozeske, R. R.; Maslanik, T.; Chapman, T.; Strand, K.; Fleshner, M.; Bachtell, R. K.; Somogyi, A. A.; Yin, H.; Katz, J. L.; Rice, K. C.; Maier, S. F.; Watkins, L. R., Opioid activation of Toll-Like receptor 4 contributes to drug reinforcement. *J. Neurosci.* **2012,** *32* (33), 11187-11200.
58. (a) Ishinaga, H.; Takeuchi, K.; Kishioka, C.; Suzuki, S.; Basbaum, C.; Majima, Y., Pranlukast inhibits NF-kappaB activation and MUC2 gene expression in cultured human epithelial cells. *Pharmacology* **2005,** *73* (2), 89-96; (b) Woszczek, G.; Chen, L.-Y.; Alsaaty, S.; Nagineni, S.; Shelhamer, J. H., Concentration dependent non-CysLT(1) receptor mediated inhibitory activity of leukotriene receptor antagonists. *J. Immunol.* **2010,** *184* (4), 2219-2225.
59. Ishioka, S.; Hozawa, S.; Haruta, Y.; Hiyama, K.; Maeda, A.; Yamakido, M., Pranlukast, a cysteinyl leukotriene antagonist, reduces serum eosinophil cationic protein levels in patients with asthma. *Hiroshima J. Med. Sci.* **1999,** *48* (4), 105-110.
60. Bollag, G.; Tsai, J.; Zhang, J.; Zhang, C.; Ibrahim, P.; Nolop, K.; Hirth, P., Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat. Rev. Drug Discov.* **2012,** *11* (11), 873-886.
61. (a) Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K., Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (8), 3041-3046; (b) Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P. N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G., Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **2010,** *467* (7315), 596-599.
62. (a) Goto, Y.; Arigami, T.; Kitago, M.; Nguyen, S. L.; Narita, N.; Ferrone, S.; Morton, D. L.; Irie, R. F.; Hoon, D. S., Activation of Toll-like receptors 2, 3, and 4 on human melanoma cells induces inflammatory factors. *Molecular cancer therapeutics* **2008,** *7* (11), 3642-3653; (b) Takazawa, Y.; Kiniwa, Y.; Ogawa, E.; Uchiyama, A.; Ashida, A.; Uhara, H.; Goto, Y.; Okuyama, R.,

Toll-like receptor 4 signaling promotes the migration of human melanoma cells. *The Tohoku journal of experimental medicine* **2014,** *234* (1), 57-65.
63. Schrödinger Release 2015-4: Maestro, v., Schrödinger, LLC, New York, NY, 2015.
64. Schrödinger Release 2015-4: LigPrep, v., Schrödinger, LLC, New York, NY, 2015.
65. Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R., Integrated modeling program, applied chemical theory (IMPACT). *J. Med. Chem.* **2005,** *26* (16), 1752-1780.
66. (a) Schrödinger Release 2015-4: Epik, v., Schrödinger, LLC, New York, NY, 2015; (b) Greenwood, J. R.; Calkins, D.; Sullivan, A. P.; Shelley, J. C., Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **2010,** *24* (6-7), 591-604; (c) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M., Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **2007,** *21* (12), 681-691.
67. (a) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S., A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007,** *47* (2), 279-94; (b) Cross, S.; Baroni, M.; Goracci, L.; Cruciani, G., GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *J. Chem. Inf. Model.* **2012,** *52* (10), 2587-98.
68. Lyne, P. D., Structure-based virtual screening: an overview. *Drug Discov. Today* **2002,** *7* (20), 1047-55.
69. Allinger, N. L.; Yuh, Y. H.; Lii, J. H., Molecular mechanics. The MM3 force field for hydrocarbons. 1. *J. Am. Chem. Soc.* **1989,** *111* (23), 8551-8566.
70. (a) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P., Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004,** *44* (3), 793-806; (b) Huang, N.; Shoichet, B. K.; Irwin, J. J., Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006,** *49* (23), 6789-6801.

CHAPTER 7:

Computational studies on the TLR4 activation

7.1 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.[1]

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.[2] A typical plasma membrane is formed by hundreds of amphipathic lipids.[3] Phospholipids, together with glycolipids, are the most abundant and the large majority of non-lipid membrane components are sterols.[4] 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, being 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.[5] 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. In biomembranes Lo is known as lipid rafts and plays a crucial role during protein signaling.[6] Regarding their composition, mammalian membranes contain phosphatedylcholine (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.[7]

The study of the dynamics of the TLR4/MD-2 complex requires the consideration of different membrane environments, since the membrane constituents are directly involved in the dimerization processes and thus governing the activation.[6] 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.[8] 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. A model for the ID/ID dimer 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.

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

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).[9] 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 (Annex Figure 1, 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 (Annex Figure 1, 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 (Annex Figure 1, 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 Annex Figure 1, 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.

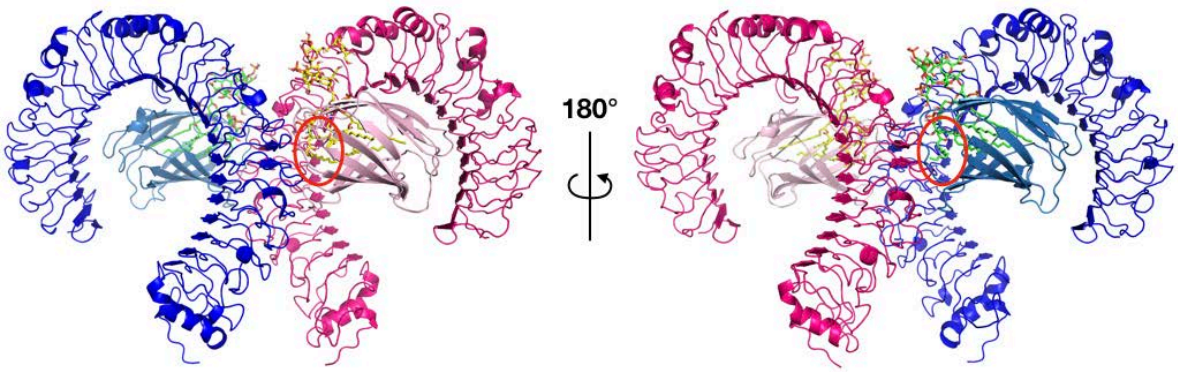


Figure 1: Full view of the dimeric (TLR4/MD-2/E. coli LPS)₂.

MD-2 shows similar amplitude of conformational deviation in both simulations (Annex Figure 2, left panel). Phe126, known as the switch ON/OFF of the receptor complex,[10] remains in an agonist-like conformation in both simulations characterized by an arbitrarily selected angle (Annex Figure 3) plotted in Annex Figure 2, right panel.

7.3 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] (Annex Table 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 Annex Figure 4 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 Annex Figure 4 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.

The importance of membrane domains, known as rafts, for proteins signaling and trafficking has been extensively reported.[11] The recruitment of TLR4 into lipid rafts domain has also been reported.[12] 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 (Annex Figure 5, 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 (Annex Figure 5 left).

7.4 TLR4 transmembrane domain and hydrophobic region

The transmembrane domain of TLR4 (TD) is predicted to span from Thr632 to Tyr652 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.[13] Very recently, Mineev *et al.* reported a NMR study of both the TD and the HR of TLR4.[13a] 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.

We performed a 250ns MD simulation of the TD and the HR structured as a long α-helix, as reported in the NMR study, 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 lipid head groups and the solvent (Figure 2).

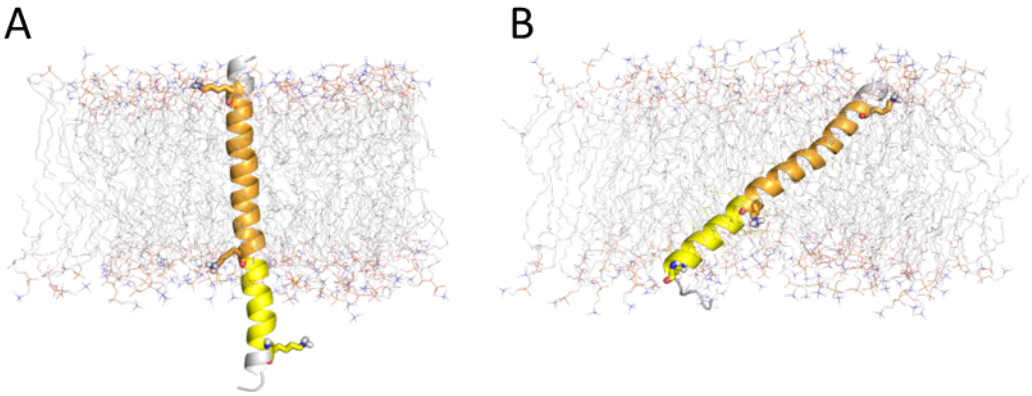


Figure 2: MD simulations of TLR4 TD and HR (uninterrupted α-helix) domain. At t=0ns of the MD simulation (A), the protein is perpendicular to the membrane plan. At t=250ns (B), the protein adopt a titled position in relation with the membrane plan.

In addition, we performed the same simulation but starting with an extended unstructured HR (Figure 3A). 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 3B). Once the HR starts to interact extensively with

the membrane it stabilizes for the rest of the MD simulation (Figure 3C), not experiencing major changes during the final 50ns. To gain further insights into the folding properties of the HR, we performed a MD simulation in water starting from an extended conformation of the HR sequence. Along the simulation, the peptide has a great tendency to adopt an α-helix (Figure 2: D, E and F). In fact, this secondary structure corresponds to the one elucidated by NMR (cf. Figure 1 of the original paper)[13a] and is associated to the insertion in the membrane.

The reported NMR results by Mineev indicate that there is a short portion (around Lys653) inside the full linker sequence TM+HR that would not correspond to a structured alpha-helix. The authors propose a full helix inserted in the membrane as a possible model for the insertion. We wondered about the secondary structure of the linker in water medium, and if it would adopt a preferred conformation. Our MD simulation shows that there is a tendency to this peptide to also adopt an α-helix structure (Figure 3). These results could also suggest the compatibility with the HR region being outside the membrane while maintaining an α-helix conformation. This could point to a relevant role of the HR in the IC-IC dimerization. The dynamics and conformational behavior of the HR peptide could be determinant of the proper IC-IC recognition. Another possibility could be that depending on the composition of the membrane, the HR adopts different conformations. That could explain the necessity of TLR4 receptor to be in a lipid raft to become active.

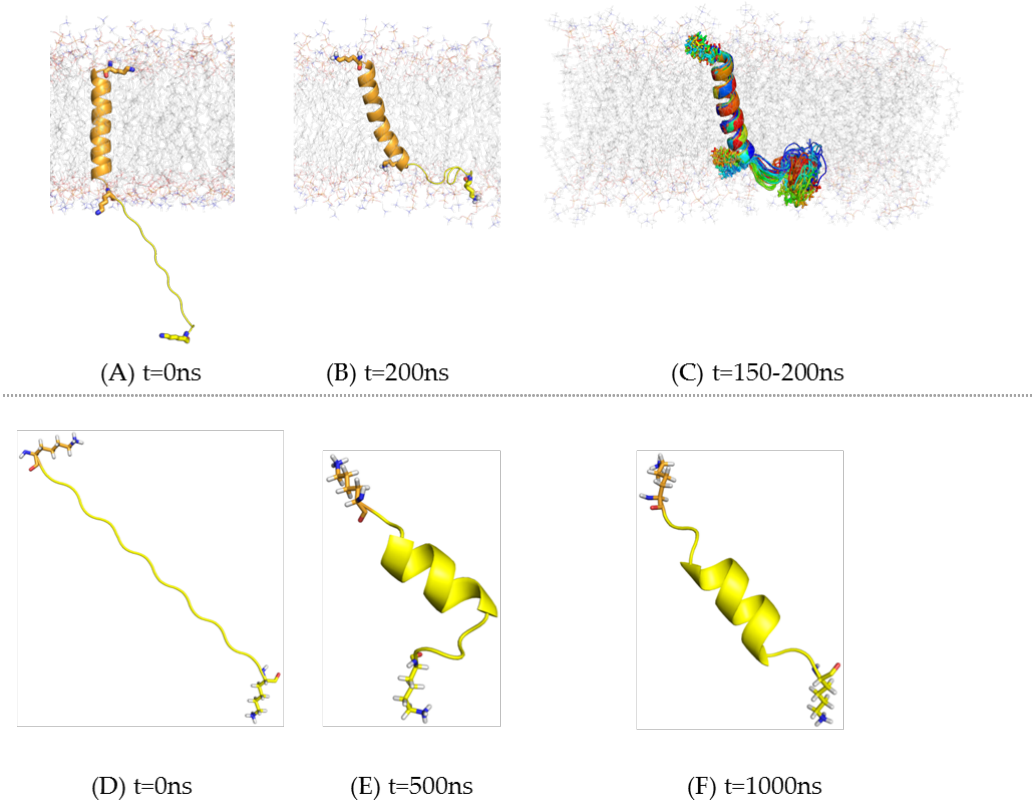


Figure 3: (A, B, and C) MD simulations of TLR4 TD (α-helix, in orange) and HR (extended coil, in yellow) domain. (A) The TD was built as an α-helix and HR, built in an extended conformation, was attached to it. (B) At the end of the simulation the HR is folded against the membrane. (C) Superimposition 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. (D, E and F) Folding study of the HR by MD simulation in water. The 3D structure is shown along the simulation: 0ns (D), 500 ns (E), and 1000 ns (F).

7.5 TLR4 TD-TD dimerization

We used the DAFT approach[14] 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 1024 ns. 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). The model with the highest score, which belongs to largest cluster, was back-mapped to all-atom following the standard backmapping protocol described described by Wassenaar *et al.*[15] and submitted to a 10 ns all-atom MD simulation with AMBER in a DOPC membrane. No major interaction changes were observed.

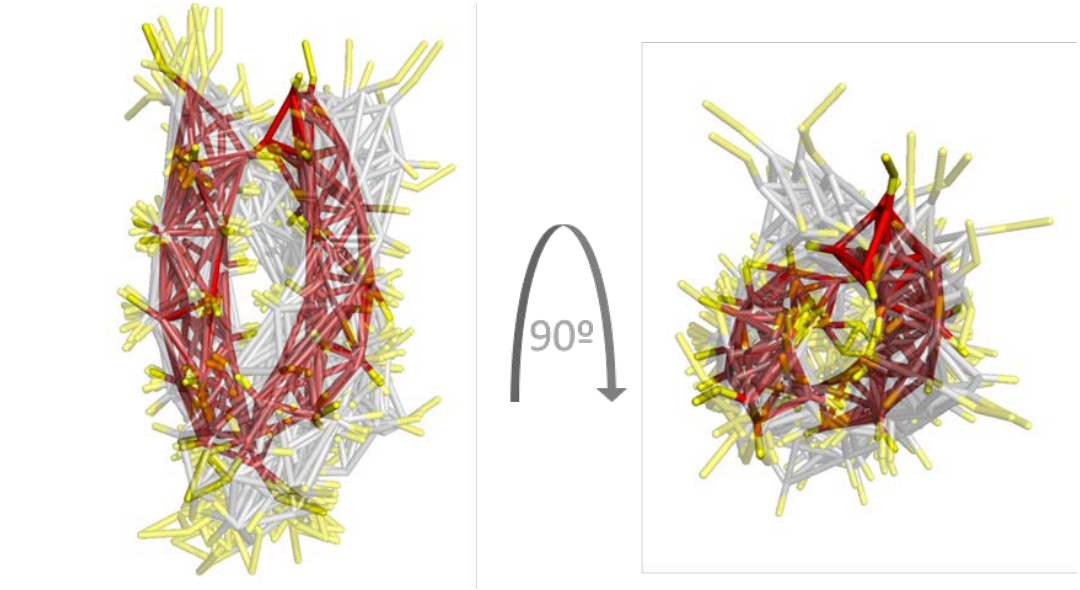


Figure 4: Superimposition, based on one monomer, of the 13 selected dimerized TD-TD poses. Backbone of the pose with the best L-J interactions score is shown in solid red sticks. Backbone (grey) and side chains (yellow) of the other poses are superimposed to the best pose and shown in semi-transparent sticks.

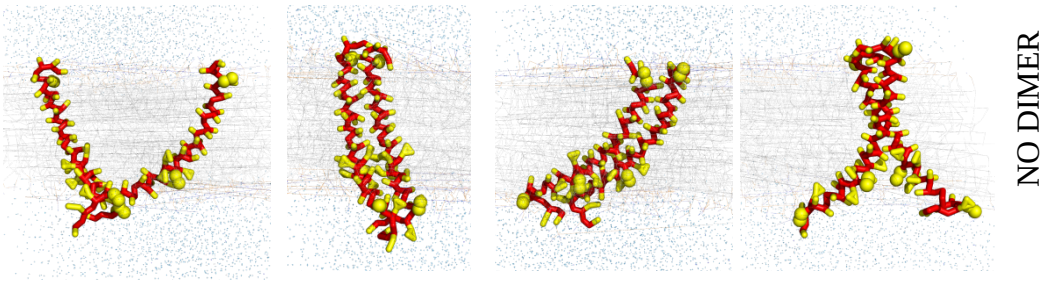
7.6 TLR4 TD HR-TD HR dimerization

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 2) 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 5). 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 3). 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 an 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 5: Different models of dimerization of the TLR4 (TD-HR)$_2$ domain. Resulting dimers at the end of the simulations in which TD and HR were parametrized as α-helix.

In the two other sets of simulations (in Lo and Ld phases) only the HR was imposed to be structure as a α-helix no secondary structure was assigned to the other residues. The results do not point to a preferantial binding mode as it can be seen in Annex Figure 6.

Among the 120 dimerization simulations setted up in the Lo phases, 105 were successfuly carried out for 2.5 μs and 104 featured a dimer in the last frame. Among the dimer the L-J energy of interaction ranges from -1374.0 to -326.9 kcal mol^{-1}. In the Ld phases, 114 achieved 0.512 μs and 91 featured a dimer in the last frame of simulation. Among the dimer the L-J energy of interaction ranges from -1324.3 to -304.2 kcal mol^{-1}.

Unlike in the simulations of the TD alone, a great variety of TD-TD dimer is observed and based on the geometry and the energy of interaction, none seemed to be favored. The TD-TD dimerization, when HR is added to the structure, seems to be mainly driven by HR-HR interactions itself. The introduction of the HR seems to create a dimerization "noise". No major dimerization differences are noted between the simulation and the Lo phase and the one in the Ld phase (Annex Figure 6).

7.7 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.*,[16] three from Guven-Maiorov *et al.*[17] and another from Miguel *et al.*,[18] and found that our model present great similarities with the latter one (Figure 6).

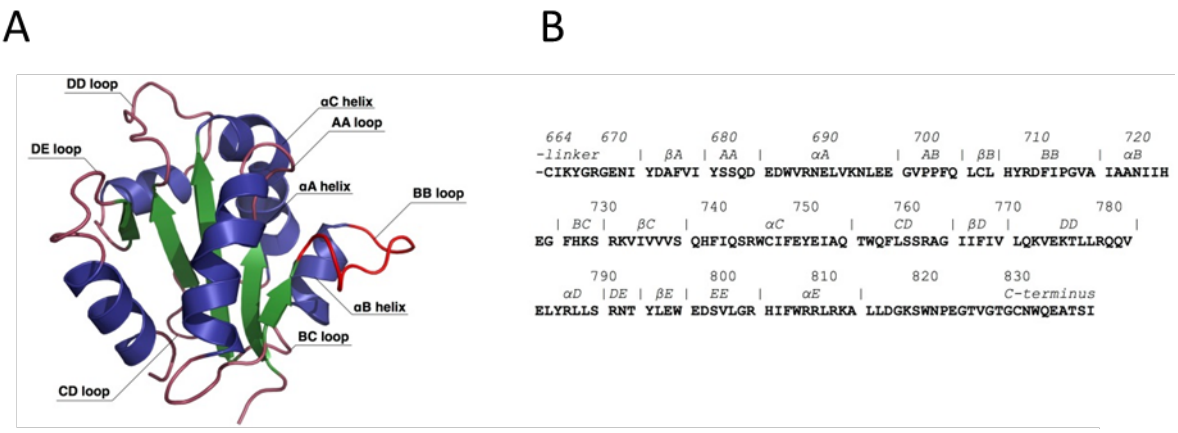


Figure 6: Intracellular TIR domain of TLR4. A) 3D representation of the homology model with details on its structural composition. B) FASTA sequence divided by its secondary structure elements.

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 (Annex Figure 7, 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 S8, left panel, blue line), due to the high motion of the C- and N- terminal linkers that are highly flexible as shown in Figure S8, right panel.

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 (Annex Figure 8). The higher motions correspond to the CD loop, shown in Figure 6 (cf. RMSF plot in Annex Figure 8, right panel). 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 7). 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.

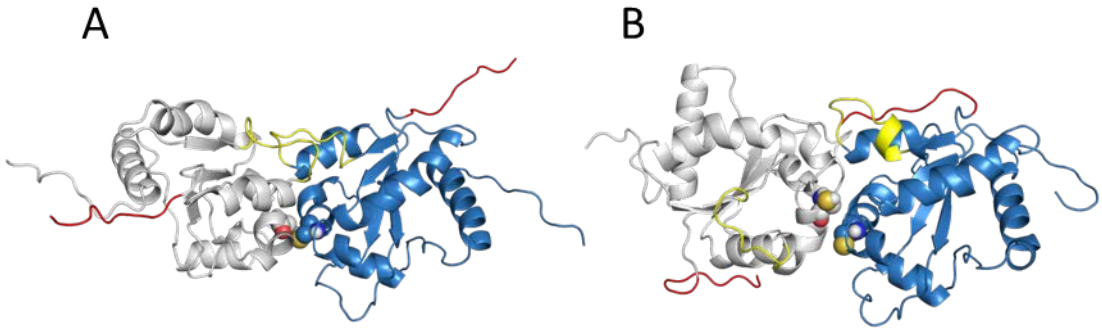


Figure 7: TLR4 ID from 0ns (left) to 200ns (right) of MD simulation in water. The dimer is in cartoon representation, Cys747 from both monomer are represented as spheres, chain A is in blue, chain B in grey, the BB loops in yellow, and the N-terminal linkers in red.

In our model we assessed the binding sites of the downstream adaptors based on the findings of Guven-Maiorov *et al.* (2015).[17] The protein-protein interaction model after 200 ns of MD simulation in 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.[18] Concerning protein-protein interactions and complexation with downstream adaptors, both models seem to be possible. To shed more light into these possibilities, we performed the docking of the MAL protein (PDB ID 3ub2) to our model of ID-ID dimer by using ZDOCK. We obtained ten docked solutions, one of them being placed at the face that is predicted to interact with the membrane according to our MD simulations. The other nine docked solutions are coincident in predicting protein-protein

interactions at the opposed face, the one that would be exposed to the cellular interior.

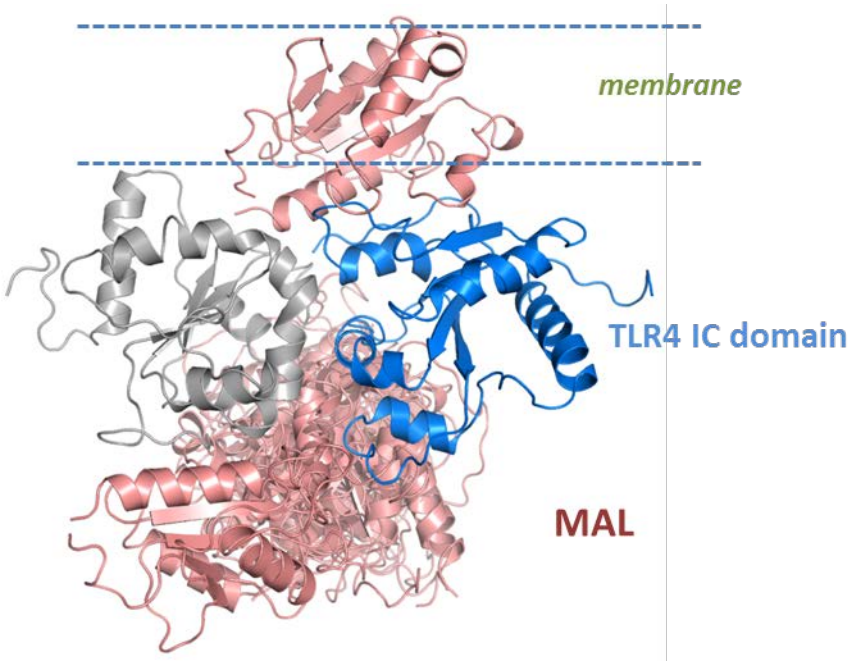


Figure 8: Protein-protein docking of TLR4 TIR domain (ID-ID dimer) and MAL protein. TLR4 ID is shown in grey and blue cartoon and the best 10 docked poses of MAL 8 in salmon.

All poses are binding to the ID-ID dimer at the dimerization interface of the two monomers, as reported by Núñez-Miguel *et al.*[18] Nevertheless, this observation differs from the signalosome model reported by Guven-Maiorov *et al.*,[17] in which the MAL protein seems to interact with one of the monomers of the IC-IC only. Our docking results also revealed that two possible binding interfaces are possible, although one of them is clearly favored, since 9 out of 10 poses are binding in that region (only pose 8 binds on the opposed face). This is in agreement with our TD/ID model because only one face is available for adaptor binding since one of the IC interface would be interacting with the membrane (data not shown). Most of the binding poses predicted by ZDOCK are in agreement as well with Bovijn *et al.*[19] findings about possible binding interfaces with MAL protein of ID domain. They

identified three possible binding surfaces, two of them corresponding to MAL/ID dimerization. One of these interfaces is located at the dimerization interface of TIR domain. In 6 out of 10 poses, we can appreciate interactions of MAL with this binding interface and in 3 of them the interaction takes place with Arg184, Ala185 or Tyr187 of Mal, residues reported as possible binding interface of MAL. All these considerations strengths the consistency of our ID-ID model and makes it a plausible model of the ID-ID dimerization.

Additionally, we performed MD simulation of the best ranked docking pose of MAL with the IC dimer. We observed that the complex remained stable during the whole simulation. MAL protein did not change its binding pose with respect to the intracellular domain of TLR4. Moreover, at binding interface there is a strong web of polar interactions that are maintained during the whole simulation (Figure 9). These residues appear also in the work previously mentioned.[19]

We also studied another way of dimerization based on an asymmetric model following the architecture of MAL protein crystal structure published by Ve *et al*. (PDB ID 5UZB).[20] The crystal structure published in this work revealed that MAL protein assembles by a head-to-tail arrangement, forming filaments. Given the high structure and sequence homology between MAL and IC TLR4, we though that the intracellular domain of TLR4 could also dimerize in a symmetric way.

With this aim, we performed a protein-protein docking using ZDOCK of two IC domain monomers, extracted from our model of the dimer. We introduced restrictions to the docking based on the mutagenesis experiments reported by Bovijn *et al.*[19] We obtained, thus, several poses that are in agreement with this type of arrangement. In this case, the B-B loop are not in contact with each other, but they interact with the following domains of the next monomer, in a head-to-tail

way of dimerization (Figure 10, A). In fact, the superimposition of 5UZB MAL crystal structure with our asymmetric model shows how similar the arrangement between these two proteins is. (Figure 10, B).

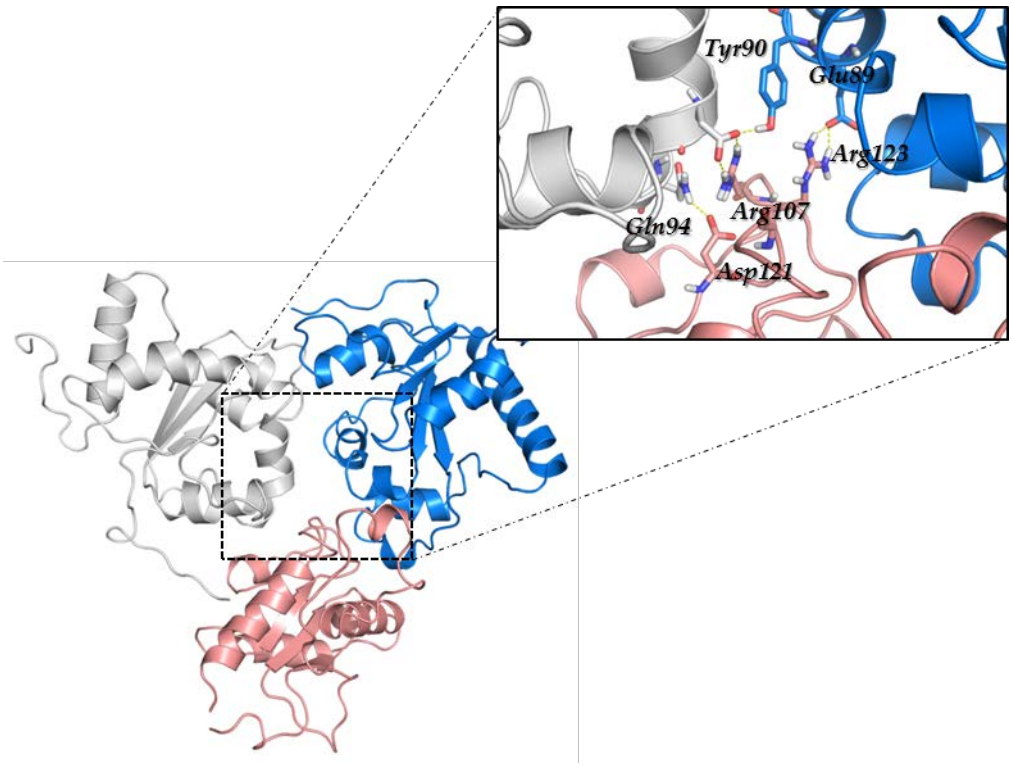


Figure 9: MD simulation of best ranked docked pose of MAL-IC protein-protein docking. IC TLR4 dimer domain is shown in grey and blue cartoons. MAL protein is shown in grey cartoons.

In order to propose a model for the binding of MAL to our asymmetric model or IC domain, we performed another protein-protein. In this case, MAL dimer 5ZUB crystal structure was docked on the asymmetric model of the IC TLR4 dimer domain. The docking was guided following not only mutagenesis studies on the IC domain, but also on MAL protein.[21] Similar to the symmetric model, in this case MAL is also interacting with IC domain at the dimerization interface (Figure 10, C). This type of arrangement is in agreement with previously reported structures and experimental results.[20]

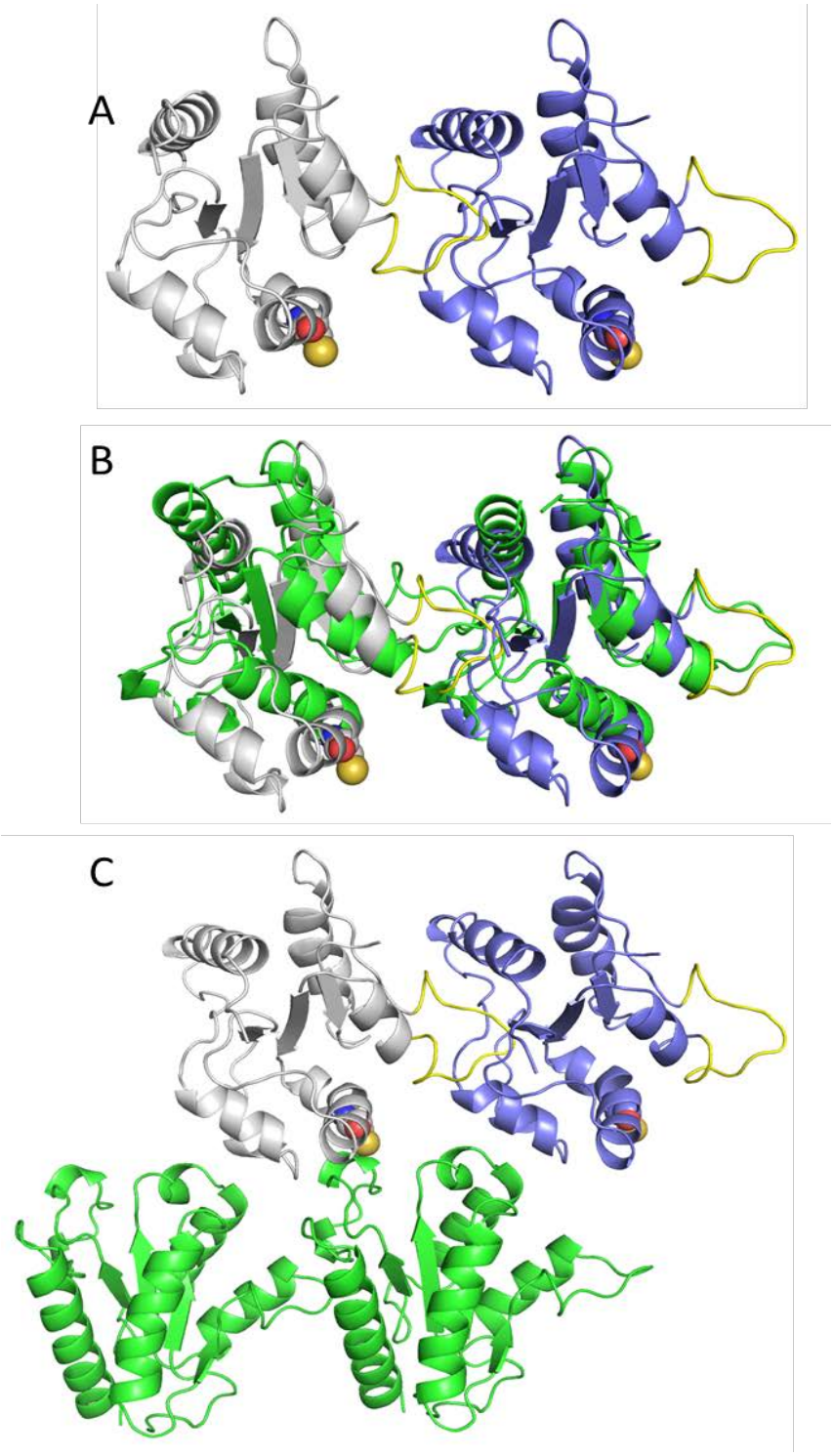


Figure 10: A) Docked pose of protein-protein docking of two monomers of the IC TLR4 domain. B) Superimpostion of docked pose with 5ZUB MAL crystal structure (green). C) Docked pose of protein-protein docking of MAL dimer (green) 5UZB crystal structure on the asymmetric IC domain. B-B loop are shown in yellow; Cys747 is highlighted in spheres.

Thus, we propose here two possible models for dimerization for the IC domain of TLR4: asymmetric and asymmetric ways. Both are in agreement with mutagenesis studies on the IC domain that identify residues and regions that are crucial for the dimerization of the IC domain and the activation of TLR4. Furthermore, both models are also plausible for the binding of MAL protein. This indicates that both dimers could exist inside the cell, mediating the signalization of TLR4 depending on its formation. In fact, this is in agreement with experimental data pointing in that direction.[22]

7.8 Conclusions

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. We have performed several MD simulations of the 3FXI crystal structure, of both dimer and monomer. In the case of the monomer simulation, *E. Coli* LPS maintains its original conformation, with 5 FA inserted in the MD-2 pocket and 1 FA chain protruding from MD-2. However, TLR4 changes its position with respect to MD-2 protein. That behavior is not observed in the dimer simulation, where both TLR4 maintain its position along the simulation, indicating that the dimer TLR4/MD-2/TLR4*/MD-2* is more stable than the monomer TLR4/MD-2. The activated form of TLR4 is dimeric, which is consistent to the results exposed here.

Secondly, we provided here several models for the transmembrane domain. First of all, an MD simulation of the TM with a full α-helix structure inserted in a POPC membrane showed an inclination of the protein. On the contrary, this inclination is not happening when the HR is not folded as α-helix. These results indicate the plasticity of HR, that could be determinant in the dimerization of the IC domain.

Coarse grain simulations of the TM dimer inserted in different membrane models indicated that, in a Ld membrane, the TM tend to partially dimerize. In the case of the simulations in a Lo membrane (mimicking a lipid raft moiety), the dimerization of the TM domain is complete. These results, together with the previous ones, point to a high plasticity of the TM domain, that can adopt different conformations and can change is mode of dimerization depending on the environment.

Finally, in the case of the IC domain, we performed a homology modelling of the dimer IC domain based on several TLR4 IC domains. Although the simulation of one the monomers indicated the good stability of the protein, a 90º rotation can be observed in the MD simulation of the dimer. This rotation exposes the BB loop to the solvent in a symmetric way. Moreover, we constructed another IC dimer performing a protein-protein docking of two IC monomers. In this case, an asymmetric way of dimerization is observed. In the asymmetric model, the BB loops are not in contact with each other, and the IC domain dimerizes is a head-to-tail fashion. This type of arrangement is in agreement with MAL dimer crystal structure (PDB ID 5UZB), which shares high homology with TLR4 IC domain. In any case, in both models, MAL protein could interact. The protein-protein docking and MD simulations performed with the TLR4 IC dimer plus MAL protein showed that MAL can form stable complexes with both of our models and interacts with residues that have been reported to be crucial to its binding. Thus, both ways of dimerization can occur simultaneously and, together with the plasticity of the TM domain, this indicates that the dimerization (and thus the activation) is an intricate combination of ligand binding, membrane composition and protein structure itself.

7.9 Materials and methods

Setting up the all-atom (AA) membranes. The lipid bilayers were built using the online CHARMM-GUI membrane builder.[24] Composition of the different system is given in SI. Systems were converted to Lipid14 compatible PDB format using the charmmlipid2amber.sh script.[25]

Full TLR4/MD-2 model construction. The 3D structures 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. 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.

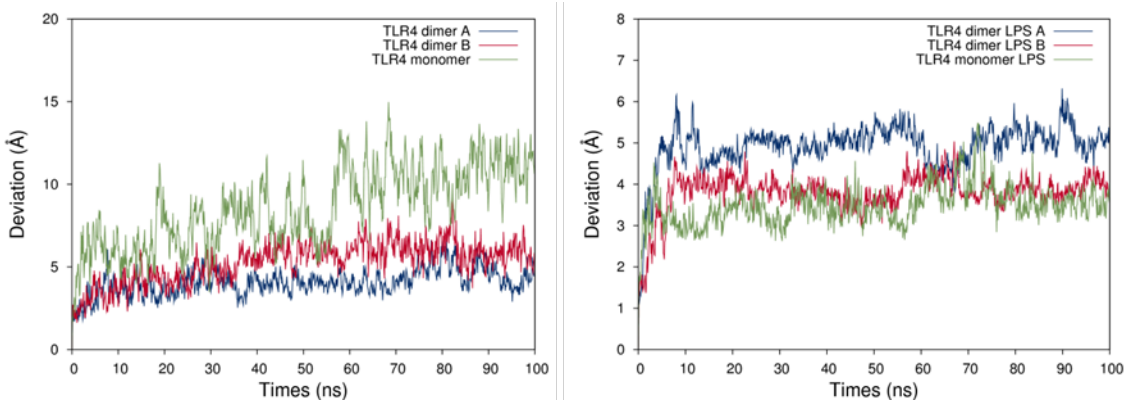
Molecular dynamics simulations. AA MD simulations were performed with either Amber14 or Amber16. The force fields ff14SB,[26] Lipid14,[25] and a combination of GAFF[27] and GLYCAM06[28] were used to described the proteins, the membrane and *E. coli* LPS, respectively. AA simulations of systems containing membranes went through the same simulation protocol. Steepest descent gradient algorithm was iterated for 5000 steps followed by 5000 iterations of conjugate gradient algorithm under no constraint. The system was then heated from 0 to 100 K for 2500 steps in the NVT ensemble while the proteins and the lipids were held by a 10 kcal.mol^{-1} A^{-1} harmonic potential. Subsequently, the system was 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 were performed for a maximum of 500ps. The temperature was controlled by a Langevin thermostat. The warming up phase and the production run were 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.[29]

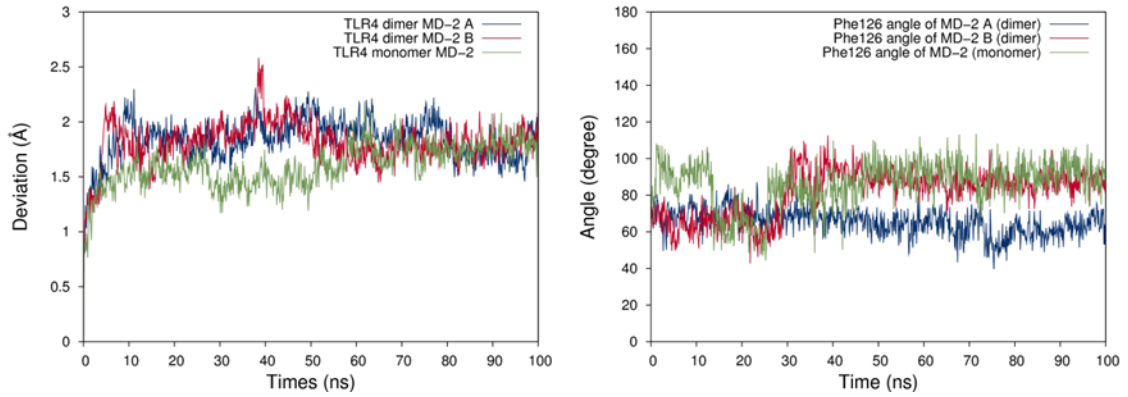
Electron density. If not stated otherwise, the electron-density was calculated over the last nanosecond of simulation.

Coarse-grained MD simulations. Box size was set to 4.5 Å in the directions of the plan of the membrane and to 5 Å in the direction normal to the membrane. Input structure was generated with Pymol based on the sequences given below. Simulations were run for 1024, 1000, 512, 3000 and 2500 ns for simulation of the TD in respectively POPC, DOPC (helix), DOPC (helix and coil), DPPC/CHOL (helix) and DPPC (helix and coil). The sequence used for the simulations of TM TLR4 in POPC is as follows: KTIIGVSVLSVLVVSVVAVLVYK and the secondary structure was set to helical. The sequence used for the four other daft simulations is the following: TCQMNKTIIGVSVLSVLVVSVVAVLVYKFYFHLMLLAGCIKYGRGE, and the secondary structure was set as follows: 7C20H19C and 3C38H5C for the helix and coil model and the helix model, respectively. The coarse-grained simulations were performed under the martini22 force field, the systems were solvated with regular martini water beads and the salt concentration was set to 0.15.

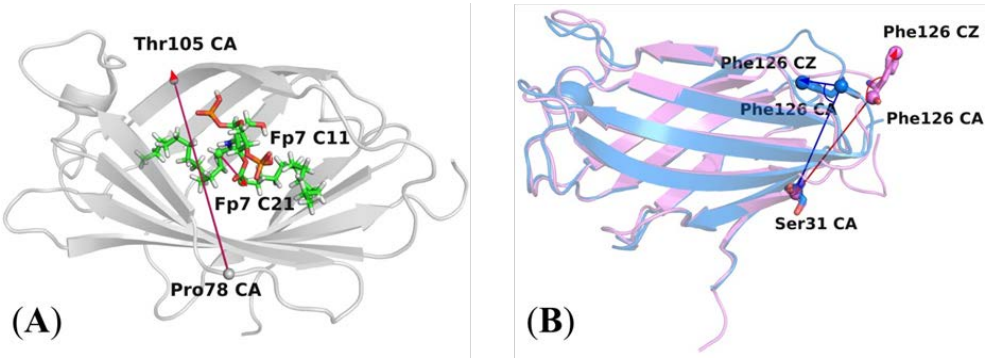
7.10 Annex



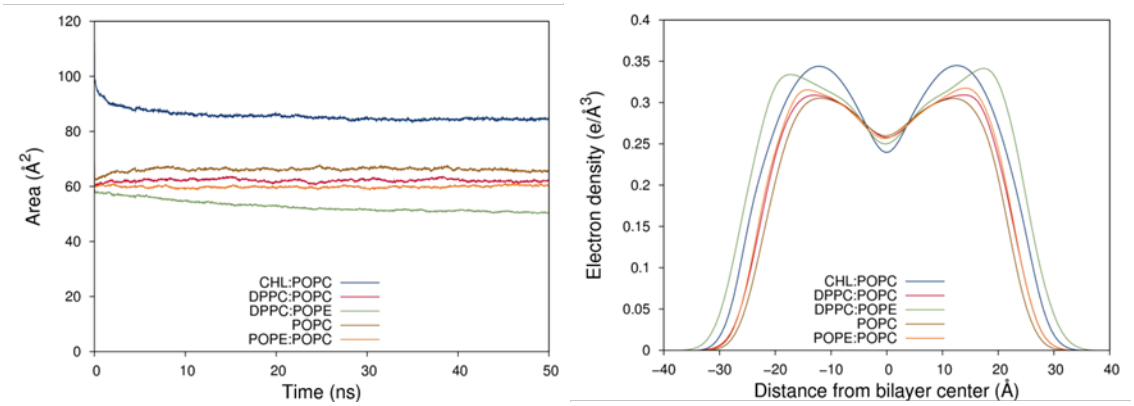
Annex Figure 1: In both plots the minimum fit was performed on the αC of MD-2 in complex with a given TLR4, giving information of how the different components of the simulations deviate from their initial position in relation to MD-2. On the left: RMSD of both TLR4 in the simulation of the dimeric model (in blue and red) and TLR4 in the simulation of the monomer (in green). On the right: RMSD of each of the LPS.



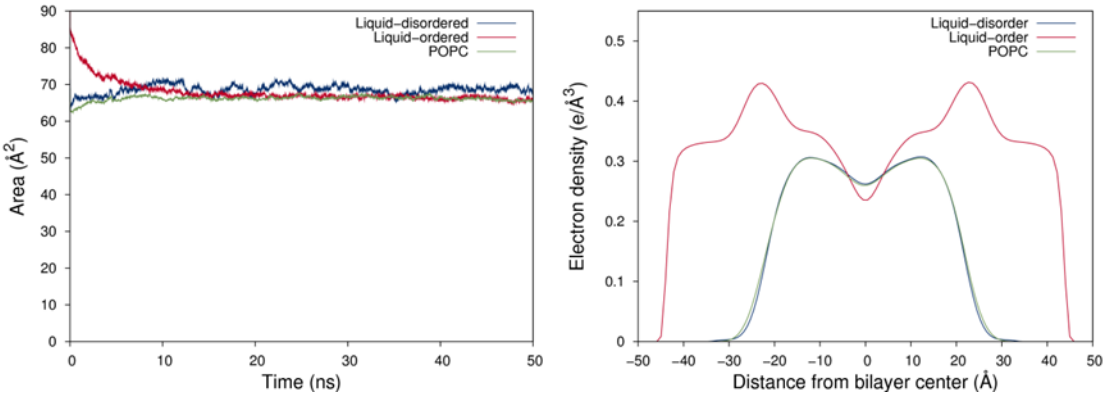
Annex Figure 2: On the left: RMSD of both MD-2 in the simulation of the dimeric model (in blue and red) and MD-2 in the simulation of the monomer (in green). On the right: angle between two vectors defined in Figure S1, providing conformational information about the critical residue Phe126 along the simulation.



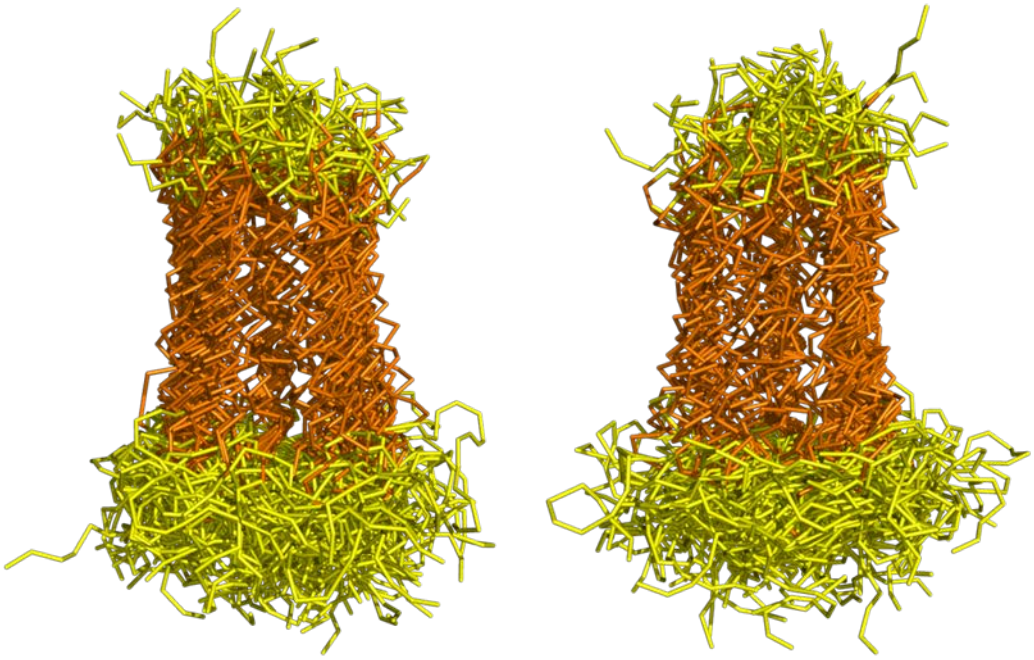
Annex Figure 3: (A) Representation of the vector from the α-carbon (CA) of Pro78 to the α-carbon of Thr105, and the vector from amide α-carbon atom (C11) and the ester α-carbon atom (C21) of FP7, used to follow the orientation of the ligands along the MD simulations (cf. S4). FP7 is used as an example; 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.



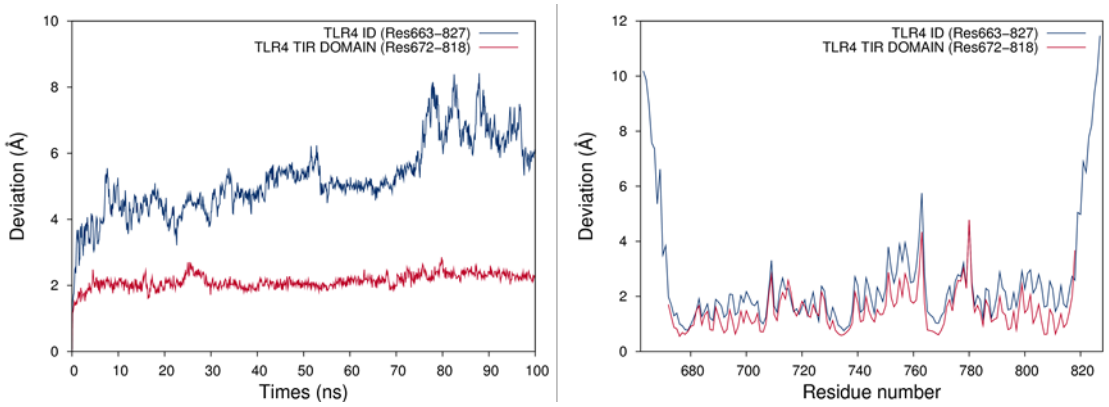
Annex Figure 4: Left: area per lipids. Right: total electron density profiles of the five membrane types (water excluded).



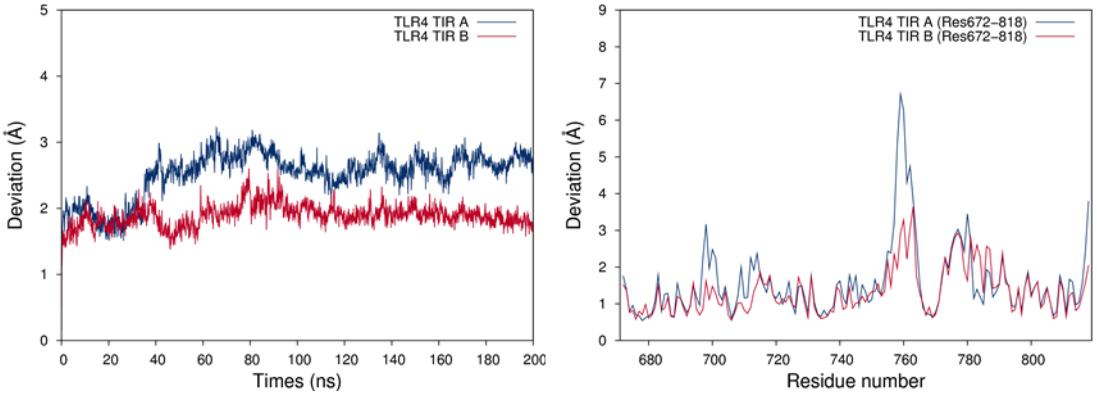
Annex Figure 5: On the left: area (in Å²) occupied by the membrane models Ld, Lo and POPC over time (ns). On the right: total electron density profiles of the five membrane types (water excluded).



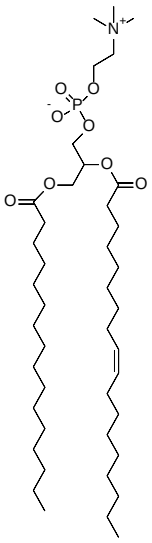
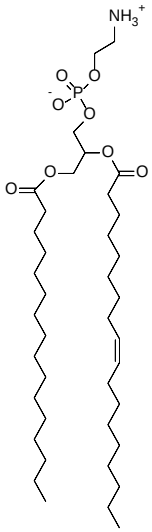
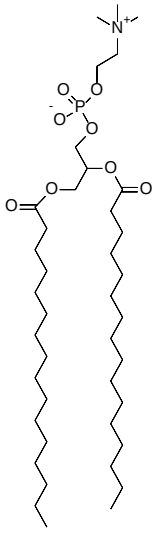
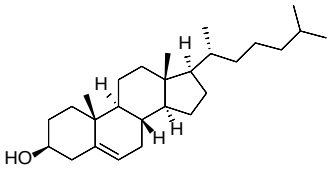
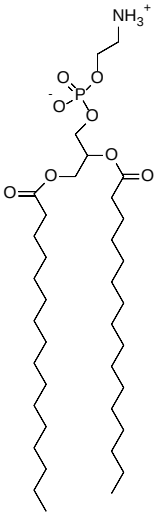
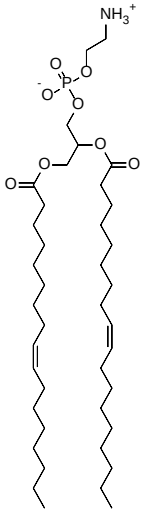
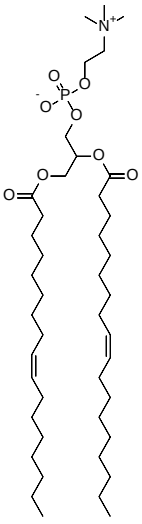
Annex Figure 6: Superimposition, based on one monomer, of all the poses from the CG simulations that presented a dimer in the final step of simulation. Backbone of residues between Lys631 and Lys653, which are the residues spanning through the membrane, are represented in orange sticks. Backbone of other residues are in yellow sticks.



Annex Figure 7: MD simulation of monomeric TLR4 ID in water. Left: RMSD. RMS fluctuation per residues.



Annex Figure 8: MD simulation of dimeric TLR4 ID in water. Left: RMSD. RMS fluctuation per residues.

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	
			

Annex Table 1: Different type of lipids and sterols used in this study.

Membrane Model	Lipid Type	Upper Leaflet Number	Lower Leaflet Number
CHOL-POPC	CHOL	370	370
	POPC	370	370
POPE-POPC	POPE	315	315
	POPC	315	315
POPC	POPC	586	586
DPPC-POPC	DPPC	305	305
	POPC	305	305
DPPC-POPE	DPPC	329	329
	POPE	329	329
ASYMMETRIC MODEL	CHOL DPPC DPPE POPE POPC DOPE DOPC	263	250

Annex Table 2: Composition of the different models of membranes.

Lipid 14 Residues		
	Description	**LIPID14 Residue Name**
Acyl Chain	Lauroyl (12:0)	LA
	Myristoyl (14:0)	MY
	Palmitoyl (16:0)	PA
	Stearoyl (18:0)	ST
Head Group	Oleoyl (18:1 n-9)	OL
	Phosphatidylcholine	PC

Annex Table 3: Lipid14 residues name.

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. (a) Akira, S.; Takeda, K., Toll-like receptor signalling. *Nat. Rev. Immunol.* **2004,** *4*, 499-511; (b) Beutler, á., TLR4 as the mammalian endotoxin sensor. In *Toll-like receptor family members and their ligands*, Springer: 2002; pp 109-120.
3. (a) van Meer, G., Cellular lipidomics. *EMBO J.* **2005,** *24* (18), 3159-65; (b) Holthuis, J. C.; Menon, A. K., Lipid landscapes and pipelines in membrane homeostasis. *Nature* **2014,** *510* (7503), 48-57.
4. 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.
5. Phillips, R.; Ursell, T.; Wiggins, P.; Sens, P., Emerging roles for lipids in shaping membrane-protein function. *Nature* **2009,** *459* (7245), 379-385.
6. 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.
7. Klose, C.; Surma, M. A.; Simons, K., Organellar lipidomics--background and perspectives. *Curr. Opin. Cell. Biol.* **2013,** *25* (4), 406-13.
8. (a) Cuschieri, J.; Bulger, E.; Billgrin, J.; Garcia, I.; Maier, R. V., Acid sphingomyelinase is required for lipid Raft TLR4 complex formation. *Surgical infections* **2007,** *8* (1), 91-106; (b) 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. *Journal of cell science* **2002,** *115* (12), 2603-2611.
9. 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.
10. 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. *Computational and structural biotechnology journal* **2013,** *5* (6), e201302012.
11. (a) Simons, K.; Toomre, D., Lipid rafts and signal transduction. *Nat. Rev. Mol. Cell Biol.* **2000,** *1* (1), 31-9; (b) Simons, K.; Ikonen, E., Functional rafts in cell membranes. *Nature* **1997,** *387* (6633), 569-72.
12. (a) 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; (b) 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; (c) 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.

13. (a) 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; (b) 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.
14. 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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), 4088-98.
20. Ve, T.; Vajjhala, P. R.; Hedger, A.; Croll, T.; DiMaio, F.; Horsefield, S.; Yu, X.; Lavrencic, P.; Hassan, Z.; Morgan, G. P.; Mansell, A.; Mobli, M.; O'Carroll, A.; Chauvin, B.; Gambin, Y.; Sierecki, E.; Landsberg, M. J.; Stacey, K. J.; Egelman, E. H.; Kobe, B., Structural basis of TIR-domain-assembly formation in MAL- and MyD88-dependent TLR4 signaling. *Nat. Struct. Mol. Biol.* **2017,** *24* (9), 743-751.
21. Bovijn, C.; Desmet, A. S.; Uyttendaele, I.; Van Acker, T.; Tavernier, J.; Peelman, F., Identification of binding sites for myeloid differentiation primary response gene 88 (MyD88) and Toll-like receptor 4 in MyD88 adapter-like (Mal). *J. Biol. Chem.* **2013,** *288* (17), 12054-66.
22. Gay, N. J.; Symmons, M. F.; Gangloff, M.; Bryant, C. E., Assembly and localization of Toll-like receptor signalling complexes. *Nat. Rev. Immunol.* **2014,** *14* (8), 546-58.
23. 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.
24. (a) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M.,

CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009,** *30* (10), 1545-1614; (b) 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.
25. Dickson, C. J.; Madej, B. D.; Skjevik, A. 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.
26. 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.
27. 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.
28. 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.
29. 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

Conclusions

This Thesis has been focused in two main receptors related to immunity: lectins and Toll-like receptor 4. On one hand, we have contributed to the advance in lectins modulation by the design of modulators of galectins 1, 3 and 7, and DC-SIGN with improved affinity and selectivity. On the other hand, we have deepen the understanding of the molecular recognition events involving TLR4/MD-2 and some modulators at atomic level, together with the construction of the full model of TLR4/MD-2 heterodimer, and we have identified novel modulators with antagonist activity. This work has been carried out by means of computational methodologies, such as MD and CG simulations, ligand-protein and protein-protein docking, and virtual screening. We have combined the computational work with experimental biological assays, and fruitful collaborations with expert groups in organic synthesis, NMR and ITC. Overall, we have helped to provide new insights for the understanding of the molecular recognition events underlying the biological functions of some lectins and TLR4.

We have developed a fragment-based computational protocol for the rational design of galectin modulators. We have focused in galectins 1, 3, and 7 and performed a virtual screening in the adjacent pocket to the CRD yielding the design of 500 lactose-based possible ligands. These compounds were docked in all the galectins, and the best candidates were selected to be synthesized. Binding and affinity studies by NMR and ITC confirmed the predictions, and allowed us to describe lactose derivatives as galectin ligands with improved affinity and selectivity towards galectins 1, 3, and 7. These selective ligands may contribute to the understanding of the highly relevant biological functions of galectins and their role in several pathological processes, such as cancer and immune diseases.

By applying a similar computational protocol as that for the design of galectin ligands, we approached DC-SIGN: the mannose binding site and the vicinal pockets. The results showed that, while the mannose binding site and the calcium coordinating residues present a stable conformation, the surroundings regions are more dynamic. We observed that Val351, which interacts with the second residue of a dimannoside, can adopt two conformations that may affect the ligand recognition. On the other hand, Phe313 also suffers conformation changes and opens up the so-called "ammonium pocket". This pocket was identified by means of VS, since most of the screened fragments bear an ammonium that was binding in that site. This pocket is adjacent to the mannose binding site and, happily, the oxygen of carbon 2 of the mannose residue points towards it. This finding allowed us to design mannoside-based ligands based on these fragments and taking into account the synthetic availability. Docking and MD simulation of these ligands to DC-SIGN predicted increased affinity compared to reported mannoside derivatives.

Moving along the Thesis, we have studied the binding mode of several TLR4 modulators. Agonist peptides RS01 and RS09 were proposed to bind to the MD-2 pocket. RS09, due to the presence of more hydrophobic residues in its structure, was predicted to be buried inside the MD-2 pocket, while RS01 was predicted to bind at the rim. MD simulations showed that only RS01 was able to maintain the agonist conformation of Phe126 loop of MD-2, in agreement with the reported experimental activity that shows that RS01 is a better agonist than RS09.

The binding mode of the TLR4 antagonist peptide has also been elucidated. This peptide, a mimetic of TLR4-binding motif of MD-2, was predicted to bind to TLR4 mimicking the MD-2 binding mode. In fact, its TLR4-bound conformation was kept in a similar way to that for MD-2. Modifications of the structure of these

peptides could lead to the design of peptide mimetics with useful applications as TLR4 modulators.

The endogenous lipid cardiolipin was also studied. We demonstrated that CL can actually bind to TLR4/MD-2, and our molecular modelling studies revealed that the FA chains of CL are inserted in the MD-2 hydrophobic pocket while the phosphate polar head is located at the rim of MD-2, interacting with residues from both TLR4 and MD-2. Moreover, we have shown that a heterodimer of TLR4/MD-2 with a molecule of CL in one monomer and a molecule of *E. coli* LPS in the other monomer can be formed, and that fact could explain the synergistic effect observed *in vitro.*

Finally, the tetracylated and pentacylated mixture of *Bacteroides vulgatus* LPS was an interesting case of study. Our molecular modelling studies indicate that both LPS_{Bv}-tetra and LPS_{Bv}-penta can bind to the MD-2/TLR4 system in a similar pose to that for *E. coli* LPS. The observed agonist behavior of LPS_{Bv}-penta and the observed antagonist behavior of LPS_{Bv}-tetra could explain the weak agonist behavior of LPS_{Bv}

We have also undertaken a VS protocol on TLR4/MD-2 to identify potential binders to this receptor. After structure-based and ligand-based VS protocols, and the redocking of the best compounds, we selected 50 ligands to carry out *in vitro* studies on HEK BLUE TLR4 cells. This assays identified 7 compounds with antagonist activity on TLR4 and the experiments with J744 macrophage cell line confirmed the activity of 2 of them, without presenting cytotoxicity. Thus, these two molecules came up as new scaffolds with a non-LPS structure with antagonist activity on TLR4

Finally, we studied the dimerization of the different TLR4 domains: extracellular, intracellular, and transmembrane domains. We provided several computational models for the dimerization of the TM domain and our results indicate that the membrane composition affects the dimerization and, thus, possibly, the activation of TLR4. We also provided two possible models for the dimerization of the intracellular domain: symmetric and asymmetric dimers. Both conformations could occur simultaneously in the cell, and our predictions indicate that MAL adaptor protein is able to bind to both models. These results all together contribute to the understanding the dynamism of TLR4 and to highlight the important role of the membrane composition for the ligand recognition and protein assembly, as key players in the TLR4 activation.